

Konstitutiewe voorwaardes vir die ontwerp van 'n toets van akademiese geletterdheid

Tobias Johannes van Dyk

Proefskrif ingelewer om te voldoen aan die vereistes vir die graad *Doctor Philosophiae* in die Fakulteit Geesteswetenskappe, Departement Engels, aan die Universiteit van die Vrystaat.

Promotor: Prof. A.J. Weideman

November 2010

Inhoudsopgawe

Lys van Tabele.....	vii
Lys van Figure	xiv
Erkennings	xvi

Hoofstuk 1 - Taalvermoë en die gereedheid om te studeer: 'n inleiding en uiteensetting

1.1	Inleidend.....	1
1.2	Aanloop tot die studie	1
1.3	Rasionaal vir die studie	10
1.4	Bestek van die studie.....	13
1.4.1	Beskrywing en konstruk van die ELSA-Plus	14
1.4.2	Die geskiktheid van die ELSA-Plus bevraagteken	15
1.4.3	Teoretiese begroning vir die studie.....	16
1.5	Probleemstelling.....	17
1.6	Formulering van hipotese.....	19
1.7	Navorsingsmetodologie.....	20
1.7.1	Navorsingsbenadering	20
1.7.2	Ondersoeksgroep en datagenerering	23
1.7.3	Dataontleding	23
1.8	Begripsverheldering	23
1.9	Uiteensetting van die studie	25

Hoofstuk 2 - Die toegepaste linguistiek: 'n dissipline van ontwerp

2.1	Inleidend.....	28
2.2	Die toegepaste linguistiek gedefinieer	30
2.2.1	'n Historiese oorsig van die toegepaste linguistiek: drie tradisies.....	31

2.2.1.1	<i>Die Noord-Amerikaanse tradisie</i>	32
2.2.1.2	<i>Die Australiese tradisie</i>	32
2.2.1.3	<i>Die Britse tradisie</i>	32
2.2.2	'n Historiese oorsig van die toegepaste linguistiek: 'n geïntegreerde, potensieel meer regverdigbare, beskrywing	33
2.2.2.1	<i>Die eerste generasie</i>	33
2.2.2.2	<i>Die tweede generasie</i>	34
2.2.2.3	<i>Die derde generasie</i>	35
2.2.2.4	<i>Die vierde generasie</i>	35
2.2.2.5	<i>Die vyfde generasie</i>	36
2.2.2.6	<i>Die sesde generasie</i>	37
2.2.2.7	<i>Die sewende generasie</i>	37
2.2.2.8	<i>'n Opsomming van die voorafgaande</i>	38
2.3	Die toegepaste linguistiek as dissipline van ontwerp.....	44
2.4	Samevatting.....	49

Hoofstuk 3 - Die ontwerp van 'n nuwe konstruk: op soek na 'n toepaslike onderbou vir die toetsing van akademiese geletterdheid

3.1	Inleidend.....	50
3.2	Toets en toetsing gedefinieer.....	53
3.3	Benaderings tot taaltoetsing	57
3.3.1	Die voorwetenskaplike/literêr-grammatiese benadering	58
3.3.2	Die psigolinguistiese benadering	58
3.3.2.1	<i>Diskrete-item toetse</i>	60
3.3.3	Die sosiolinguistiese/pragmatiese, oftewel kommunikatiewe benadering	61
3.3.3.1	<i>Geïntegreerde toetse (integratiewe toetse)</i>	62
3.3.3.2	<i>Die kommunikatiewe benadering tot taaltoetsing</i>	63
3.4	Konteks vir die ontwerp van 'n konstruk vir 'n toets van akademiese geletterdheid.....	66
3.4.1	Die voorgestelde nuwe konstruk.....	72

3.4.1.1	<i>Toetskonstrukte</i>	72
3.4.1.2	<i>Konstrukte van akademiese geletterdheid</i>	73
3.4.1.3	<i>Van teorie na praktyk: die ontwerp van 'n konstruk vir akademiese geletterdheid</i>	79
3.5	Samevatting.....	85

Hoofstuk 4 - Op soek na die toepaslike maatstaf: van konstruk na toetsspesifikasie na taaktipe

4.1	Inleidend.....	87
4.2	Die konteks.....	88
4.3	Taaktipes	91
4.4	Taaktipes en ontwerpbeginsels.....	95
4.5	Eksperimentering ter verbetering van taak- en itemtipes.....	99
4.6	Uiteindelijke toetsspesifikasies	112
4.6.1	Die skryf van toetsspesifikasies	112
4.7	Samevatting.....	116

Hoofstuk 5 - Betroubaarheid: gister, vandag en môre dieselfde

5.1	Inleidend.....	119
5.2	Betroubaarheid gedefinieer	121
5.2.1	Die wiskundige onderbou van die konsep betroubaarheid	127
5.2.2	Betroubaarheidskoëffisiënte	129
5.2.2.1	<i>Toets-hertoets-betroubaarheid</i>	131
5.2.2.2	<i>Ekwivalente-vorm-betroubaarheid</i>	142
5.2.2.3	<i>Interne konsistensie / Inhoudsbetroubaarheid</i>	147
5.2.2.3.1	Die halfverdelingsmetode	148
5.2.2.3.2	Die Kuder-Richardson-metode	151
5.2.2.3.3	Cronbach se alpha (α)	153
5.3	Die standaardmetingsfout.....	157
5.4	Faktore wat betroubaarheidskoëffisiënte beïnvloed.....	163

5.5	Die interpretasie van betroubaarheidskoeffisiënte	165
5.6	Samevatting.....	166

Hoofstuk 6 - Geldigheid: drie moontlike perspektiewe

6.1	Inleidend.....	169
6.2	Geldigheid gedefinieer	170
6.2.1	Die eerste paradigma: 'n tradisionele benadering tot geldigheid	172
6.2.2	Die tweede paradigma: die huidige ortodoksie.....	175
6.2.3	Die derde paradigma: 'n alternatief tot die huidige ortodoksie	185
6.3	Samevatting.....	191

Hoofstuk 7 - Validering: 'n eenheid in 'n verskeidenheid van argumente

7.1	Inleidend.....	194
7.2	Toetsvalidering gedefinieer.....	195
7.3	'n Geldigheidsargument vir TAG	199
7.3.1	Interne geldigheid	199
7.3.1.1	<i>Hipotese 1: TAG beskik oor gesigsgeldigheid.....</i>	<i>199</i>
7.3.1.2	<i>Hipotese 2: TAG beskik oor inhoudsgeldigheid.....</i>	<i>202</i>
7.3.1.3	<i>Hipotese 3: TAG beskik oor konstrugeldigheid.....</i>	<i>211</i>
7.3.1.3.1	Sub-hipotese 3a: TAG beskik oor interne konsekwentheid.....	214
7.3.1.3.2	Sub-hipotese 3b: TAG beskik oor intratoetsgeldigheid.....	215
7.3.1.3.3	Sub-hipotese 3c: TAG beskik oor intertoetsgeldigheid.....	219
7.3.1.3.4	Sub-hipotese 3d: TAG beskik oor groep-differensiële geldigheid .	246
7.3.1.3.5	Sub-hipotese 3e: TAG beskik oor domeinspesifieke geldigheid....	261
7.3.2	Eksterne geldigheid.....	261
7.3.2.1	<i>Hipotese 4: TAG beskik oor voorspellingsgeldigheid.....</i>	<i>262</i>
7.3.2.2	<i>Hipotese 5: TAG beskik oor samevallende geldigheid.....</i>	<i>274</i>
7.3.2.2	<i>Hipotese 6: TAG beskik oor konsekwensiële geldigheid.....</i>	<i>282</i>
7.4	Samevatting.....	285

Hoofstuk 8 - 'n Finale kwessie: die verantwoordelike interpretasie van toetsresultate

8.1	Inleidend.....	287
8.2	Die akkommodering van 'n verskeidenheid van perspektiewe.....	288
8.3	'n Versoberende perspektief op die resultate van TAG.....	289
8.3.1	'n Onderbou van tegnies-teoretiese verdedigbaarheid.....	289
8.3.2	Bepaling en hantering van afsnypunte.....	290
8.3.3	Potensiële misklassifikasie van grensgevalle	292
8.3.4	Kategorisering van toetslinge	292
8.3.5	Gevolge vir toetslinge	293
8.3.6	Deursigtigheid.....	294
8.3.7	Koste	294
8.3.8	Bruikbaarheid.....	295
8.4	Samevatting.....	295

Hoofstuk 9 - Slot en samevatting

9.1	Inleidend.....	297
9.2	Opsomming en gevolgtrekkings van die studie	299
9.3	Verdere navorsingsmoontlikhede.....	302
9.4	Ten slotte.....	304

BIBLIOGRAFIE	305
Opsomming en sleuteltermes	330
Bylaag A: Toetsafnemershandleiding vir die Noordwes Universiteit, die Universiteit van Pretoria, en die Universiteit van die Vrystaat.....	334
Bylaag B: Toetsspesifikasies vir voornemende studente aan die Universiteit Stellenbosch.....	343

Lys van Tabelle

Tabel 1.1: Fokusareas van die navorsingsbenadering van hierdie studie met spesifieke toepassing op TAG	22
Tabel 2.1: Oorvleuelende en onderskeidende generasies in die toegepaste linguistiek	39
Tabel 2.2: Terugskouende en antisipatoriese momente van linguistiese ontwerpe.....	47
Tabel 2.3: Konstitutiewe en regulatiewe momente in toegepaste linguistiese ontwerpe	48
Tabel 3.1: Opsomming van die verskille tussen 'n beperkende en 'n oop siening van taal (uit Weideman, 2003b:38)	69
Tabel 4.1: Vergelyking van potensieel bruikbare taaktipes met konstruk	93
Tabel 4.2: Vergelyking van toetsgemiddeld, betroubaarheid, gemiddelde moeilikheidsgraad en gemiddelde diskriminasiewaarde vir 'n eerste en tweede loodsfase van sinsluitingsubtoets.....	103
Tabel 4.3: Vergelyking van toetsgemiddeld, betroubaarheid, gemiddelde moeilikheidsgraad en gemiddelde diskriminasiewaarde vir 'n derde en vierde loodsfase van sinsluitingsubtoets.....	103
Tabel 4.4: Vergelyking van toetsgemiddeld, betroubaarheid, gemiddelde moeilikheidsgraad en gemiddelde diskriminasiewaarde van sinsluitingsubtoets van TAG en TALL 2004	104
Tabel 4.5: Vergelyking van toetsgemiddeld, betroubaarheid, gemiddelde moeilikheidsgraad en gemiddelde diskriminasiewaarde van sinsluitingsub-toets vir TAG 2007.....	106
Tabel 4.6: Vergelyking van toetsgemiddeld, betroubaarheid, gemiddelde moeilikheidsgraad en gemiddelde diskriminasiewaarde van sinsluitingsub-toets vir 'n verdere loodsfase in 2008	110
Tabel 4.7: Toetsspesifikasies van TAG vir toetsontwikkelaars	115
Tabel 5.1: Benaderings tot die bepaling van toetsbetroubaarheid.....	130
Tabel 5.2: Opsomming van die toets-hertoets-betroubaarheidskoëffisiënt.....	132
Tabel 5.3: Beskrywende statistiek vir studente wat taalintervensie in Afrikaans gevolg het (Afrikaans 143 én Professionele Kommunikasie 113 in Afrikaans)	134
Tabel 5.4: Beskrywende statistiek vir studente wat Afrikaans 143 gevolg het.....	135
Tabel 5.5: Beskrywende statistiek vir studente wat Professionele Kommunikasie 113 in Afrikaans gevolg het.....	136

Tabel 5.6: Beskrywende statistiek vir alle studente wat <i>Wetenskapskommunikasievaardigheid 172 (WKV 172)</i> in Afrikaans gevolg het	138
Tabel 5.7: Beskrywende statistiek vir alle studente wat <i>Wetenskapskommunikasievaardigheid 172 (WKV 172)</i> in Engels gevolg het	138
Tabel 5.8: Opsomming van die ekwivalente-vorm-betroubaarheidskoëffisiënt	145
Tabel 5.9: Voorbeeld van opsomming van twee afdelings van TAG 2005 en TAG 2006 vir doeleindes van bepaling van ekwivalente-vorm-betroubaarheid	146
Tabel 5.10: Korrelasiematriks	147
Tabel 5.11: Uiteensetting van alpha-waardes van TAG en TALL sedert 2004	155
Tabel 5.12: Riglyne vir die oorweging van toleransievlakke by betroubaarheids- koëffisiënte.....	166
Tabel 6.1: Messick se benadering tot geldigheid	179
Tabel 6.2: McNamara en Roever se herinterpretasie van Messick se benadering tot geldigheid.....	181
Tabel 6.3: Weideman se herinterpretasie van Messick se benadering tot geldigheid met 'n fokus op die verband tussen fundamentele oorwegings in taaltoetsing.....	189
Tabel 7.1: Matriks vir die bepaling van die konstrukverteenvoerdigende aard van die toets	204
Tabel 7.2: Matriks vir beoordelaars / vakkundiges se kommentaar	205
Tabel 7.3: Matriks vir beoordelaars / vakkundiges se kommentaar oor teks	208
Tabel 7.4: Beskrywende statistiek onderliggend aan die resultate vir Tabel 7.5	216
Tabel 7.5: Betroubaarheidskoëffisiënte van die TAG-subtoetse oor ses administrasies heen by die Universiteit van Pretoria en die Noordwes Universiteit.....	217
Tabel 7.6: Beskrywende statistiek vir die bepaling van inter-subtoetskorrelasies vir die 2007-inname aan die Universiteit Stellenbosch, die Noordwes Universiteit en die Universiteit van Pretoria	218
Tabel 7.7: Korrelasies tussen verskillende subtoetse vir die 2007-inname aan die Universiteit Stellenbosch, die Noordwes Universiteit en die Universiteit van Pretoria	218
Tabel 7.8: Beskrywende statistiek vir die bepaling van kongruente geldigheid vir die 2007-inname aan die Universiteit Stellenbosch.....	221
Tabel 7.9: Korrelasies tussen verskillende datastelle vir die 2007-inname aan die Universiteit Stellenbosch vir die bepaling van kongruente geldigheid	221

Tabel 7.10: Beskrywende statistiek vir die bepaling van kongruente geldigheid vir die 2008-inname aan die Universiteit Stellenbosch.....	222
Tabel 7.11: Korrelasies tussen verskillende datastelle vir die 2008-inname aan die Universiteit Stellenbosch vir die bepaling van kongruente geldigheid	223
Tabel 7.12: Beskrywende statistiek vir die bepaling van kongruente geldigheid vir die 2009-inname aan die Universiteit Stellenbosch.....	224
Tabel 7.13: Korrelasies tussen verskillende datastelle vir die 2009-inname aan die Universiteit Stellenbosch vir die bepaling van kongruente geldigheid	225
Tabel 7.14: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2007-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch	233
Tabel 7.15: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2007-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch	233
Tabel 7.16: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2008-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch	235
Tabel 7.17: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2008-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch	235
Tabel 7.18: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2009-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch	236
Tabel 7.19: Korrelasies tussen TAG en die ander toetse vir die 2009-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch.....	236
Tabel 7.20: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2007-inname van die Humaniora aan die Universiteit Stellenbosch	238
Tabel 7.21: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2007-inname van die Humaniora aan die Universiteit Stellenbosch.....	238
Tabel 7.22: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2008-inname van die Humaniora aan die Universiteit Stellenbosch	239

Tabel 7.23: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2008-inname van die Humaniora aan die Universiteit Stellenbosch.....	239
Tabel 7.24: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2009-inname van die Humaniora aan die Universiteit Stellenbosch.....	239
Tabel 7.25: Korrelasies tussen TAG en die ander toetse vir die 2009-inname van die Humaniora aan die Universiteit Stellenbosch.....	240
Tabel 7.26: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2007-inname van die Scientiae aan die Universiteit Stellenbosch .	241
Tabel 7.27: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2007-inname van die Scientiae aan die Universiteit Stellenbosch.....	241
Tabel 7.28: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2008-inname van die Scientiae aan die Universiteit Stellenbosch .	241
Tabel 7.29: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2008-inname van die Scientiae aan die Universiteit Stellenbosch.....	242
Tabel 7.30: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2009-inname van die Scientiae aan die Universiteit Stellenbosch .	242
Tabel 7.31: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2009-inname van die Scientiae aan die Universiteit Stellenbosch.....	242
Tabel 7.32: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2009-inname vir studente met Afrikaans as moedertaal aan die Universiteit Stellenbosch	243
Tabel 7.33: Korrelasies tussen TAG, TALL, PTEEP en die NBT(AG) vir die 2009-inname vir studente met Afrikaans as moedertaal aan die Universiteit Stellenbosch .	243
Tabel 7.34: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2009-inname vir studente met Engels as moedertaal aan die Universiteit Stellenbosch	244
Tabel 7.35: Korrelasies tussen TAG, TALL, PTEEP en die NBT(AG) vir die 2009-inname vir studente met Engels as moedertaal aan die Universiteit Stellenbosch	244
Tabel 7.36: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per geslag vir die 2007-inname aan die Universiteit Stellenbosch.....	248
Tabel 7.37: Levene se toets vir homogeniteit van variansies ten opsigte van geslag vir die 2007-inname aan die Universiteit Stellenbosch	249

Tabel 7.38: ANOVA vir verskil in gemiddelde TAG-prestasie per geslag vir die 2007-inname aan die Universiteit Stellenbosch.....	249
Tabel 7.39: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per geslag vir die 2008-inname aan die Universiteit Stellenbosch.....	250
Tabel 7.40: Levene se toets vir homogeniteit van variansies ten opsigte van geslag vir die 2008-inname aan die Universiteit Stellenbosch	250
Tabel 7.41: ANOVA vir verskil in gemiddelde TAG-prestasie per geslag vir die 2008-inname aan die Universiteit Stellenbosch.....	250
Tabel 7.42: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per geslag vir die 2009-inname aan die Universiteit Stellenbosch.....	251
Tabel 7.43: Levene se toets vir homogeniteit van variansies ten opsigte van geslag vir die 2009-inname aan die Universiteit Stellenbosch	251
Tabel 7.44: T-toets vir verskil in gemiddelde TAG-prestasie per geslag vir die 2009-inname aan die Universiteit Stellenbosch.....	252
Tabel 7.45: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2007-inname aan die Universiteit Stellenbosch.....	253
Tabel 7.46: Levene se toets vir homogeniteit van variansies ten opsigte van ras vir die 2007-inname aan die Universiteit Stellenbosch.....	253
Tabel 7.47: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2007-inname aan die Universiteit Stellenbosch.....	254
Tabel 7.48: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2008-inname aan die Universiteit Stellenbosch.....	254
Tabel 7.49: Levene se toets vir homogeniteit van variansies ten opsigte van ras vir die 2008-inname aan die Universiteit Stellenbosch.....	255
Tabel 7.50: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2008-inname aan die Universiteit Stellenbosch.....	255
Tabel 7.51: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2009-inname aan die Universiteit Stellenbosch.....	256
Tabel 7.52: Levene se toets vir homogeniteit van variansies ten opsigte van ras vir die 2009-inname aan die Universiteit Stellenbosch.....	256
Tabel 7.53: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2009-inname aan die Universiteit Stellenbosch.....	256
Tabel 7.54: Beskrywende statistiek vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2007-inname aan die Universiteit Stellenbosch.....	258

Tabel 7.55: Levene se toets vir homogeniteit van variansies ten opsigte van Graad 12-groep vir die 2007-inname aan die Universiteit Stellenbosch	258
Tabel 7.56: T-toets vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2007-inname aan die Universiteit Stellenbosch	258
Tabel 7.57: Beskrywende statistiek vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2008-inname aan die Universiteit Stellenbosch	259
Tabel 7.58: Levene se toets vir homogeniteit van variansies ten opsigte van Graad 12-groep vir die 2008-inname aan die Universiteit Stellenbosch	259
Tabel 7.59: T-toets vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2008-inname aan die Universiteit Stellenbosch	259
Tabel 7.60: Beskrywende statistiek vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2009-inname aan die Universiteit Stellenbosch	259
Tabel 7.61: Levene se toets vir homogeniteit van variansies ten opsigte van Graad 12-groep vir die 2009-inname aan die Universiteit Stellenbosch	260
Tabel 7.62: T-toets vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2009-inname aan die Universiteit Stellenbosch	260
Tabel 7.63: Beskrywende statistiek vir die bepaling van voorspellingsgeldigheid vir die 2004-inname aan die Universiteit van Pretoria.....	267
Tabel 7.64: Korrelasies tussen TAG, Pcred, Pprog, Pmod, PTEEP, MACH, MCOM, en M-telling vir die 2004-inname aan die Universiteit van Pretoria	267
Tabel 7.65: Beskrywende statistiek vir die bepaling van voorspellingsgeldigheid vir die 2004-inname aan die Universiteit van Pretoria.....	267
Tabel 7.66: Korrelasies tussen TALL, Pcred, Pprog, Pmod, PTEEP, MACH, MCOM, en M-telling vir die 2004-inname aan die Universiteit van Pretoria	268
Tabel 7.67: Korrelasies tussen TAG en TALL en geweegde gemiddelde eerstejaarsprestasie aan die Universiteit Stellenbosch vir die 2005-, 2006-, en 2007-innames onderskeidelik	269
Tabel 7.68: Korrelasies tussen TAG en TALL en geweegde gemiddelde eerstejaarsprestasie per fakulteit aan die Universiteit Stellenbosch vir die 2005-, 2006-, en 2007-innames saamgevoeg	270
Tabel 7.69: Beskrywende statistiek vir die bepaling van voorspellingsgeldigheid vir die 2007-inname aan die Universiteit Stellenbosch.....	272
Tabel 7.70: Korrelasies tussen TAG met geweegde eerstejaars-, tweedejaars, en derdejaarsgemiddeld vir die 2007-inname aan die Universiteit Stellenbosch.....	272

Tabel 7.71: Gemiddelde prestasie in TAG volgens driejarige graad in minimum tydperk behaal vir die 2007-inname aan die Universiteit Stellenbosch.....	273
Tabel 7.72: Beskrywende statistiek vir die bepaling van voorspellingsgeldigheid vir die 2008-inname aan die Universiteit Stellenbosch	273
Tabel 7.73: Korrelasies tussen TAG met geweegde eerstejaars-, en tweedejaars-gemiddeld vir die 2008-inname aan die Universiteit Stellenbosch	273
Tabel 7.74: Beskrywende statistiek vir die bepaling van voorspellingsgeldigheid vir die 2009-inname aan die Universiteit Stellenbosch	273
Tabel 7.75: Korrelasies tussen TAG met geweegde eerstejaarsgemiddeld vir die 2009-inname aan die Universiteit Stellenbosch.....	274
Tabel 7.76: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per moedertaal vir die 2007-inname aan die Universiteit Stellenbosch	278
Tabel 7.77: Levene se toets vir homogeniteit van variansies ten opsigte van moedertaal vir die 2007-inname aan die Universiteit Stellenbosch	278
Tabel 7.78: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2007-inname aan die Universiteit Stellenbosch.....	279
Tabel 7.79: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per moedertaal vir die 2008-inname aan die Universiteit Stellenbosch	280
Tabel 7.80: Levene se toets vir homogeniteit van variansies ten opsigte van moedertaal vir die 2008-inname aan die Universiteit Stellenbosch	280
Tabel 7.81: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2008-inname aan die Universiteit Stellenbosch.....	280
Tabel 7.82: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per moedertaal vir die 2009-inname aan die Universiteit Stellenbosch	281
Tabel 7.83: Levene se toets vir homogeniteit van variansies ten opsigte van moedertaal vir die 2009-inname aan die Universiteit Stellenbosch	281
Tabel 7.84: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2009-inname aan die Universiteit Stellenbosch.....	281

Lys van Figure

Figuur 1.1: Nie-moedertaal- vs. moedertaalgebruikers van Engels as taal van onderrig en leer aan die Universiteit van Pretoria	8
Figuur 2.1: Vyf fases van toegepaste linguistiese ontwerpe	42
Figuur 2.2: Leidende en basisfunksies van die toegepaste linguistiek	45
Figuur 3.1: 'n Toepassing van die model van die toegepaste linguistiek as dissipline van tegniese ontwerp op taaltoetsing (aangepas uit McNamara, 2004:765)	56
Figuur 3.2: Die Bachman en Palmer-konstruk	77
Figuur 4.1: Matriks vir beoordeling van toetsitems	91
Figuur 4.2: Illustrasie van artikulasie tussen konstitutiewe en regulatiewe elemente	98
Figuur 5.1: Illustrasie van die balans tussen die konstitutiewe en die regulatiewe	120
Figuur 5.2: Illustrasie van die invloed van sistematiese en onsistematiese fluktuasie op betroubaarheid	125
Figuur 5.3: Illustrasie van toets-hertoets-betroubaarheid	131
Figuur 5.4a: Januarie-prestasie	140
Figuur 5.4b: November-prestasie	140
Figuur 5.5: Illustrasie van ekwivalente-vorm-betroubaarheid	143
Figuur 5.6: Illustrasie van teenbalansering	144
Figuur 5.7a: Faktoranalise vir TALL	152
Figuur 5.7b: Faktoranalise vir TAG	153
Figuur 7.1: Knakpuntgrafiek vir 2007 vir die Universiteit van Pretoria en die Noordwes Universiteit gesamentlik	228
Figuur 7.2: Knakpuntgrafiek vir 2008 vir die Universiteit van Pretoria en die Noordwes Universiteit gesamentlik	229
Figuur 7.3: Knakpuntgrafiek vir 2009 vir die Universiteit van Pretoria en die Noordwes Universiteit gesamentlik	230
Figuur 7.4: Grafiese voorstelling van verskil in vir die 2007-inname aan die Universiteit Stellenbosch	249
Figuur 7.5: Grafiese voorstelling van verskil in vir die 2008-inname aan die Universiteit Stellenbosch	251

Figuur 7.6: Grafiese voorstelling van verskil in vir die 2009-inname aan die Universiteit Stellenbosch	252
Figuur 9.1: Illustrasie van artikulasie tussen konstitutiewe en regulatiewe elemente	299

Erkennings

Alhoewel ek die voorreg het dat my naam alleen voor op hierdie dokument staan, sou dit nie tot stand gekom het as dit nie vir 'n hele paar persone was nie; 'n proefskrif het beslis nie nét een “outeur” nie:

Dit is gepas dat prof. Albert Weideman se naam saam met myne op die titelblad verskyn. Sy besondere leiding, insig, wysheid en mentorskap het nie net bygedra tot die afhandeling en dokumentering van hierdie studie nie, maar dit het ook 'n nuwe wêreld vir my laat oopgaan. Dit is hy wat die saadjie natgelei het wat 'n ander mentor van my, prof. Christo van Rensburg, geplant het.

My kollegas van die Universiteit Stellenbosch, veral van die Taalsentrum se Eenheid vir Afrikaans en Engels, verdien ook erkenning vir hul gewilligheid om werk by my oor te neem, of vir my in te staan wanneer nodig. Ek moet ook vir Loumarie Kistner uitsonder vir haar hulp met die statistiese verwerkings en geduldige verduidelikings; so ook prof. Martin Kidd vir die gehaltekontrole op sommige van die statistiese bewerkings.

Ek gee voorts erkenning aan die Universiteit Stellenbosch se Taalsentrum vir studieverlof, asook die *Inter-institutional Centre for Language Development and Assessment* vir finansiële steun in die vorm van 'n beurs.

Net die familie van mense wat reeds 'n proefskrif voltooi het, sal weet hoeveel opoffering, ondersteuning en meelewing dit verg. Dankie dus veral aan my ouers Gerda, Coen en Marie hiervoor. Ek dra hierdie werk op aan my vrou, Liezl, en my kinders, Karla en Tobias – hulle wat my in liefde ondersteun en verduur het, nie ter wille van die proefskrif nie, maar ter wille van my.

SOLI DEO GLORIA

Taalvermoë en die gereedheid om te studeer: 'n inleiding en uiteensetting

1.1 Inleidend

Die huidige belangstelling in die meting van taalvermoë is nie 'n blote historiese toevalligheid nie. Hierdie studie handel eerstens oor die komplekse sosiale, ekonomiese, politieke en historiese redes waarom die kwessie van taaltoetsing in Suid-Afrika so belangrik geword het, en vandag veral in die hoëronderwyssektor van toenemende belang is. As 'n mens soek na redes vir die belangstelling, blyk dit gou dat die klem inderdaad moet val op die begrip 'kompleks'. By nadere ondersoek blyk dit ook gou dat hoe dieper daar gedelf word vir die redes vir die huidige belangwekkendheid van die meting van taalvermoë, hoe duideliker is dit dat die oplossing vir die probleem van die meting van taalvermoë nie anders sal kan as om gekompliseerd en gesofistikeerd te wees nie. Komplekse probleme het selde eenvoudige oplossings. Hierdie studie beoog dus, tweedens, om 'n aanvanklike insig in die kompleksiteit van veral die oplossing te bied.

1.2 Aanloop tot die studie

Uit verskeie studies (Ferreira, 1992; Foxcroft & Stumpf, 2005; Hay & Marais, 2004; Nel, 2007) blyk dit dat Suid-Afrikaanse nuweling eerstejaarstudente nie

voldoende voorbereid is vir universiteitstudie nie en na alle waarskynlikheid nie suksesvol sal wees in die minimum voorgeskrewe studietydperk nie. Op grond van data van die Suid-Afrikaanse Nasionale Departement van Onderwys, wys Mamaila (2001:1) byvoorbeeld daarop dat 25% van eerstejaarstudente wat vir universiteitstudie ingeskryf is, reeds vóór die voltooiing van hulle studie die handdoek ingooi. Die situasie bly kommerwekkend in die daaropvolgende jare van die voorgraadse program, aangesien die uitdagings net vermeerder. Tussen 30% en 40% van eerstejaarstudente voltooi byvoorbeeld nie hul voorgraadse studies in die aanbevole tydperk nie. Mkhabela en Malan (2004:2) skets 'n selfs donkerder prentjie: nie net het die persentasie uitsakkers onder eerstejaars toegeneem nie, maar ongeveer 45% van histories benadeelde studente voltooi glad nie hulle studie nie. Die probleem is trouens volgens Rademeyer (2005a:4) van so 'n omvang dat 50% van alle tersiêre studente tussen 2000 en 2002 gedruip het, terwyl net 22% tussen 2002 en 2003 hulle graad verwerf het. 'n Meer onlangse studie (Scott, Yeld & Hendry, 2007) toon aan dat die situasie sedert 2001 eerder versleg as verbeter het, aangesien ongeveer 30% van alle eerstejaarstudente na die eerste studiejaar die stelsel verlaat. Dieselfde outeurs bevestig dat die situasie kommerwekkend bly in die daaropvolgende jare van die voorgraadse program, aangesien die uitdagings net vermeerder: na vyf jaar van studie het slegs 30% van alle voorgraadse studente in Suid-Afrika byvoorbeeld gegradueer, 14% is steeds geregistreer en 56% het die stelsel verlaat sonder om te gradueer. Die gemiddelde koers vir die afhandeling van studies word daarom op 44% bereken. Deurvloeikoerse aan hoëronderwysinstellings is dus onbevredigend en verteenwoordig 'n finansiële verlies van miljarde rand per jaar. Hierdie blyk egter nie net op nasionale vlak 'n tendens te wees nie: in 'n studie deur Astin & Oseguera (2002) is daar aangetoon dat slegs ongeveer een uit elke drie eerstejaarstudente in die VSA hul grade in die voorgeskrewe tydperk verwerf het. Daar was 'n verhoging in deurvloeikoerse na bykans 60%, maar eers na die minimumtydperk plus vier jaar. Dit beteken dus dat ongeveer 40% van die Amerikaanse nuweling eerstejaarstudente na ses jaar in die sisteem steeds nie 'n graad verwerf nie.

Daar is verskeie redes vir hierdie situasie, maar lae vlakke van vaardigheid in die taal van onderrig en leer word allerweë as een van die hoofredes vir 'n gebrek aan akademiese sukses onder Suid-Afrikaanse voorgraadse studente beskou - selfs onder diegene met groot akademiese potensiaal (Blacquiére, 1989; Leibowitz, 2001; Macdonald, 1990; Perkins, 1991; Pretorius, 1995; Van Rensburg & Weideman, 2002; Vorster & Reagan, 1990). Navorsing deur onafhanklike onderwysspesialiste dui ook op 'n moontlike daling in die (akademiese) taalvaardigheidsvlak van studente wat vir tersiêre studie inskryf (Illiterate matrices, 2004:8; Rademeyer, 2005b:40). Dit word gedeeltelik bevestig deur die uitslae van akademiese geletterdheidstoetse wat deesdae by verskeie universiteite as platings- en selfs toegangsmeganisme gebruik word, alhoewel, soos hierdie studie ook sal aantoon, die aard van die 'daling' en die graad daarvan, veral oor korter periodes, met omsigtigheid en genuanseerd beskryf moet word. Wat wel sonder meer waar is, is dat die vlakke van akademiese geletterdheid (ook soms akademiese taalvaardigheid genoem) van so 'n aard is dat dit ernstige risiko's inhou vir studente wat betref die voltooiing van hul studies in die voorgeskrewe tyd. Laasgenoemde stelling word bevestig deur die werk van Boyuwoye (2002:280) waarin aangedui word dat akademiese taalvaardigheid een van die grootste bydraende faktore tot lae deurvloeikoerse in hoëronderwys is.

'n Vanselfsprekende gevolg van die genoemde lae akademiese geletterdheidsvlakke is dat sodanige studente moeilik met voorgeskrewe akademiese materiaal omgaan. Botha, Du Plessis en Menkveld (2007a:10-11) toon in hierdie verband aan dat 'n beduidende persentasie eerstejaarstudente oor alle rasgroeperinge heen oor gebrekkige leesvaardighede beskik. Frick (2007) het tot soortgelyke bevindings gekom in 'n studie waarin sy aandui dat taalvaardighede soos luister, praat, lees en skryf 'n deurslaggewende rol in akademiese sukses speel. Dit vind weerklank in Van Rensburg en Weideman (2002:153) se bewering dat akademiese geletterdheid 'n *sine qua non* vir die suksesvolle voltooiing van tersiêre studie is.

Die toelating van studente wat ondervoorbereid is ten opsigte van akademiese taalvaardigheid (Zamel & Spack, 1998), blyk egter nie net 'n plaaslike verskynsel te wees nie (Butler & Van Dyk, 2004:1). Pierce (2003) doen byvoorbeeld verslag oor navorsing wat daarop dui dat slegs 6% van eerstejaarstudente aan universiteite in die Verenigde State van Amerika (VSA) gereed is om onafhanklik te studeer. Dié toedrag van sake verswak selfs verder, aangesien diegene wat goed kan skryf slegs 2% van die totaal uitmaak, terwyl minder as 'n derde (31%) oor voldoende skryfvermoë beskik. 'n Studie wat McKenzie en Schweitzer (2001) onder Australiese studente uitgevoer het, bevestig hierdie bevindinge, en beklemtoon ook dat akademiese taalvaardigheid akademiese sukses beïnvloed. Daar kan dus tereg na akademiese geletterdheidsvlakke en die invloed daarvan op akademiese sukses verwys word as 'n internasionale kwessie, soos die studies uit die VSA en Australië aantoon, en nie as een wat slegs tot 'n bepaalde konteks (land, ekonomie, demografie, maatskappy, ensovoorts) beperk is nie.

Verskeie redes word plaaslik vir bogenoemde situasie aangevoer, waarvan drie vir doeleindes van hierdie studie bespreek sal word. Die eerste hiervan, wat ook dikwels as een van die belangrikste redes vir akademiese mislukking beskou word, is die politieke geskiedenis van segregasie en die gevolglike ongelyke verdeling van hulpbronne in die Suid-Afrikaanse onderwysstelsel wat 'n groot groep studente – na wie daar algemeen as histories benadeelde studente verwys word – negatief geraak het. Cooper en Van Dyk (2003:68) beaam hierdie siening wanneer hulle opmerk dat 'n gebrek aan akademiese sukses onder 'n groot hoeveelheid histories benadeelde studente te bespeur is weens, soos hulle dit stel, “[a] number of socio-economic and political factors, rooted in the educational policies implemented during the era of the apartheid government.”

Suid-Afrikaanse hoëronderwysinstellings (HOI's) word tans deur die nasionale plan vir hoëronderwys (Ministerie van Onderwys, 2001) genoop om hierdie wanbalanse reg te stel deur tersiêre onderrig vir alle voornemende studente meer toeganklik te maak (Departement van Onderwys, 2001:4). Ntshoe (2002:7) lê ook klem op die

behoefte vir billikheid en regstelling in HOI's. Die Departement van Onderwys (2001:47) waarsku egter teen 'n onverantwoordelike benadering om summier met sogenaamde hekwagaksies deur HOI's weg te doen en histories benadeelde studente voor die voet toe te laat bloot om die diversiteitsprofiel van die bepaalde instelling reg te stel. Desnieteenstaande is daar 'n beduidende toename in registrasies van histories benadeelde studente aan voorheen bevoordeelde HOI's, veral aan die sogenaamde histories-Afrikaanse universiteite (Nel, 2007:3), instellings wat vandag miskien meer akkuraat as meertalige universiteite beskryf kan word, in teenstelling met eentalige universiteite wat die meerderheid vorm.

Swak voorbereidheid en gepaardgaande gebrekkige studiesukses is egter nie 'n verskynsel wat slegs onder histories benadeelde studente te bespeur is nie. 'n Toenemende aantal tradisioneel wit, sogenaamd bevoorregte studente staar ook akademiese mislukking in die gesig. Trouens, ook moedertaalstudente ervaar toenemend die akademiese diskoers in hul moedertaal as 'n "vreemde taal" en universiteite behoort ook hieraan aandag te skenk (Gee, 1998). Nel (2007:5) wys byvoorbeeld daarop dat "[d]ie gaping tussen skool en universiteit nie net vergroot [word] deur die onvoorbereidheid van leerders in die Suid-Afrikaanse skoolstelsel nie, maar ook deur die onvoorbereidheid van universiteite om leerders – veral uit benadeelde omstandighede – te akkommodeer," wat hieronder geïllustreer sal word. Let egter daarop dat studente uit histories bevoorregte omstandighede myns insiens nie uitgesluit kan word van die kategorie van swak voorbereidheid nie.

'n Verdere rede vir lae akademiese geletterdheidsvlakke kan wees dat die Suid-Afrikaanse onderwysstelsel nie spesifiek daarop ingestel is om leerders vir universiteitstudie voor te berei nie, maar ook op voorbereiding vir die arbeidsmark en ander instellings van hoëronderwys soos die kolleges vir verdere onderwys en opleiding (Van Wijk, 1987:24). Viljoen (2005:42) wys juis daarop dat daar nie net op skole staatgemaak kan word om leerders in terme van taal- en akademiese vaardighede vir universiteitstudie voor te berei nie, maar dat dit 'n gedeelte verantwoordelikheid is. Op hierdie punt is dit belangrik om ook daarop te let dat

skoolleerplanne in die verlede sterk leerplangedrewe en positivisties was. Binne hierdie stelsel:

- is kennis as universeel, onveranderlik, konteks-vry en neutraal beskou;
- was die owerheid (onderwyser) die enigste inligtingsbron, en is sy/haar sienings nooit bevraagteken nie, aangesien hy/sy altyd die beste geweet het;
- was persoonlike ervaring en kultuur, of die bekende werklikheid, die enigste geldige verwysingsraamwerk;
- het die klem op linkerbreinaktiwiteit grootliks tot 'n liniêre denkwysing aanleiding gegee; en
- was begrip ongedifferensieerd, en is alle konsepte as ewe belangrik beskou.

(Bencze, 2005; Blanckenberg, 1999:34)

Tans is onderwysers vasgevang tussen twee strydende paradigmas: die positivistiese patrone van die verlede en die hedendaagse, oorwegend konstruktivistiese benadering tot onderrig en leer (Blanckenberg, 1999:41). Onderwysers het oor die algemeen nog nie die skuif na die konstruktivistiese benadering, wat die grondslag van ontwikkelings soos uitkomsgebaseerde onderwys (UBO) vorm, gemaak nie. Hiervoor bestaan daar etlike ingewikkelde redes wat vir die doeleindes van hierdie studie nie hier bespreek sal word nie, maar waarvan 'n hoofrede moontlik 'n gebrek aan opleiding en hulpbronne is om so 'n skuif moontlik te maak. Op die oomblik weifel onderwysers dus tussen die ou en die nuwe benaderings tot onderrig en leer.

Rademeyer (2005b:41) sluit by Van Wijk hierbo aan en stel die ironiese en ietwat kontensieuse moontlikheid dat studente as gevolg van UBO leerprobleme in die tersiêre leeromgewing kan ondervind, aangesien die doel daarvan is om leerders vir die werklike lewe (outentieke situasies) en nie noodwendig vir universiteitstudievoor te berei nie. Leerders ontwikkel dus in hierdie interpretasie van wat UBO behels nie meer noodwendig die vermoë om as selfstandige, strategiese, krities-besinnende denkers op te tree nie. Dit word geag as 'n basiese, noodsaaklike denkhouding vir enige universiteitstudent. Universiteitstudente (veral eerstejaars) kan gevolglik onder andere daarmee sukkel om taakopdragte, toetsvrae

en akademiese tekste met 'n redelik gesofistikeerde argument en abstrakte woordeskat strategies te lees en te verstaan, en sinvol en korrek daarop te reageer (mondelings of skriftelik). Hetsy hierdie interpretasie egter korrek is of nie, is dit nogtans moontlik dat die taalkurrikulum in sekondêre onderwys leerders nie toereikend voorberei op die hoërorde-taaldenkvaardighede wat hulle hiervoor benodig nie. Hierdie situasie mag veral problematies wees onder histories benadeelde studente, met spesifieke verwysing na swart studente, waar die taalkurrikulum en assesseringspraktyke vir Engels op tweede- en addisionele taalvlak byvoorbeeld nie leerders voorberei of taalmatig bemagtig om hul ander vakke te hanteer nie (Umalusi, 2004:8).

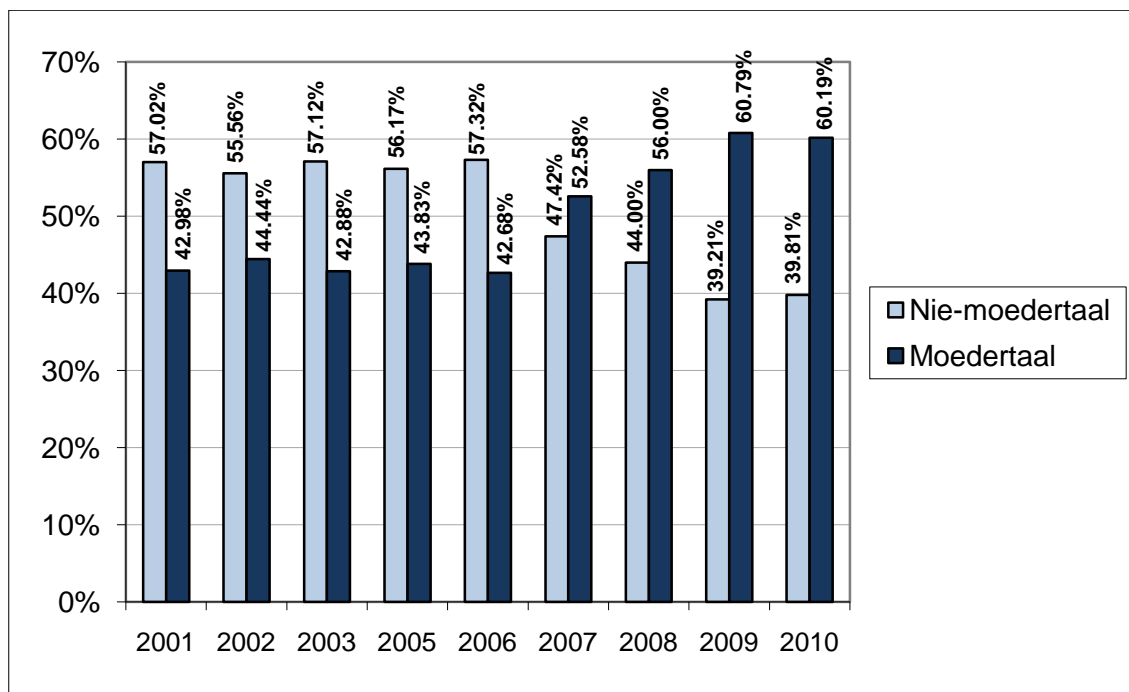
'n Derde moontlike rede is dat al hoe meer universiteitstudente in Engels wil studeer terwyl dit nie noodwendig hulle eerstetaal is nie. Die werk van Squelch (1993:183) maak byvoorbeeld spesifiek vermelding van die invloed van die onderrigtaal op studiesukses. Nel (2007:41) wys ook daarop dat die begrip van komplekse konsepte wat uiteindelik in abstrakte denke moet ontwikkel in die tweede of addisionele taal problematies is. Die situasie aan die Universiteit van Pretoria (UP), een van Suid-Afrika se grootste residensiële universiteite, en slegs ter illustrasie, lyk by nadere ondersoek geensins anders nie. Volgens Webb (2002:50) verkies 'n groot persentasie hedendaagse studente aan die UP byvoorbeeld Engels as taal van onderrig en leer.

Figuur 1.1¹ (bygewerk uit Butler & Van Dyk, 2004:2) illustreer hierdie stelling deur besonderhede te verstrek oor die persentasie nie-moedertaalsprekers teenoor eerstetaalsprekers van Engels as leertaal aan die UP. Die grafiek toon die taalstatus (nie-moedertaal/moedertaal) van diegene wat gekies het om die verpligte akademiese geletterdheidstoets aan die UP in Engels af te lê – studente aan die UP kan kies in watter een van die twee onderrigtaale, Afrikaans of Engels, hulle getoets

1. Die data wat vir figuur 1 gebruik is, kom van die databank (eerstejaarskrywings) van die Eenheid vir Akademiese Geletterdheid aan die Universiteit van Pretoria. Weens 'n verandering in die toets wat vir die evaluering van akademiese geletterdheid gebruik word, was daar geen elektroniese data vir die 2004-toelatings beskikbaar nie.

wil word. Let daarop dat die toets wat vanaf 2001 tot 2003 gebruik is, verskil van die wat vanaf 2004 gebruik is (verwys paragrawe 1.3 en 1.4 hieronder).

Dit is interessant om daarop te let dat die tendens in 2007 verander het en sedertdien min of meer konstant gebly het – hierdie verandering kan moontlik toegeskryf word aan die uiteindelijke afdwing van beleid waar studente vroeër in sekere fakulteite, desnieteenstaande die beleid, tog verplig is om die akademiese geletterdheidstoetse in 'n spesifieke taal af te lê. Nog 'n moontlikheid is dat al minder studente aandui dat hulle moedertaal 'n ander taal as Engels is - op nasionale vlak word daar kommer uitgespreek oor studente wat tweetalig grootword en hul moedertaal dan as slegs Engels aangee; so ook oor studente wat eentalig grootword, maar dan ook Engels as hul moedertaal aangee. Verder is daar ook die moontlikheid dat studente meen hulle benadeel kan word as hulle nie Engels as hul moedertaal aandui nie. Die feit bly egter staan dat daar op nasionale vlak steeds meer en meer studente aandui dat hulle bereid is om in Engels te studeer, terwyl dit inderdaad nie hul moedertaal is nie.



Figuur 1.1: Nie-moedertaal- vs. moedertaalgebruikers van Engels as taal van onderrig en leer aan die Universiteit van Pretoria

'n Ondersoek na die akademiese sukseskoers van onderskeidelik eerste- en addisioneletaal-sprekendes toon egter dat 'n groter aantal eerstetaalsprekers as hulle addisioneletaal-eweknieë aan die einde van hulle eerste en daaropvolgende studiejare sukses behaal (Butler & Van Dyk, 2004:1; Scott, Yeld & Hendry, 2007). Die navorsing van Botha, Du Plessis en Menkveld (2007b:5-6) bevestig dat moedertaalsprekers van 'n onderrigtaal by 'n universiteit byvoorbeeld ongeveer gemiddeld 8% beter presteer as hul addisioneletaal-eweknieë.

Buiten bogenoemde drie moontlike redes vir akademiese mislukking (politieke geskiedenis, benaderings tot onderrig en leer, en die aantal tweede- of addisioneletaallearers wat vir tersiêre studie inskryf), is die Suid-Afrikaanse situasie tans redelik kompleks, aangesien 'n aantal bykomende faktore sukseskoerse kan beïnvloed en gevolglik ook finansiële verliese vir HOI's kan meebring.

Die eerste hiervan is Yeld (2001:iii, 1) se siening dat Suid-Afrikaanse matriekuitslae moontlik nie meer 'n betroubare voorspeller van akademiese sukses in hoërsonderwys is nie, veral binne konteks van die sogenaamde simbole-inflasie (Yeld & Hendry, 2002:4). Bargate (1999), Botha en Cilliers (1999), Paras (2001), en Wood (1998) bevraagteken ook die betroubaarheid van matriekuitslae as die enigste aanduiding van akademiese sukses in hoërsonderwys, aangesien die uitsakkoers aan universiteite jaarliks styg. Die matriekuitslae word egter steeds deur die publiek, politici, beleidsmakers en navorsers as voldoende aanduiding van akademiese sukses beskou en HOI's laat steeds studente op grond hiervan toe (Jawitz, 1995; Umalusi, 2004:2; Reddy, 2006:xii).

Tweedens, in die lig van die verwickelinge waardeur die huidige matrieksertifikaat asook die matriekulasievereistes (universiteitsvrystelling) in 2008 deur 'n nuwe stelsel vervang is, as gevolg van 'n internasionale neiging om skool- en universiteitskurrikula aan te pas om in lyn te wees met die behoeftes van 'n globaliserende wêreld, bestaan daar tot op hede geen aanduiding dat hierdie situasie sal verbeter nie.

Laastens dra akademies swak voorbereide studente daartoe by dat universiteite groot bedrae geld in die vorm van subsidies verloor, aangesien 'n gedeelte van die subsidie eers bekom word wanneer 'n student gradueer. Die subsidiestelsel vir hoëronderwys in Suid-Afrika neem onder andere vier finansieringsgroepe, vier studievlakke en die aantal histories benadeelde studente-inskrywings by die betrokke instellings in ag (Vermeulen & Melck, 2004). In 'n sekere mate is universiteite dus daartoe genoop om (deur middel van verskeie inisiatiewe/meganismes, waarvan een hieronder in meer besonderhede bespreek word) meer swak voorbereide studente as in die verlede toe te laat.

Op hierdie punt is dit egter noodsaaklik om daarop te wys dat daar nie in 'n sogenaamde krisis-narratief verval moet word oor lae geletterdheidsvlakke nie. Ivanič, Edwards, Barton, Martin-Jones, Fowler, Hughes, Mannion, Miller, Satchwell, en Smith (2009:13) stel dit dat

... we are bombarded almost daily with crisis narratives, the notion that something is not working as it should or could. That there is a problem to be addressed, or that there has been a decline in some shape or form is a constant theme of [...] discourse. It seems to be endemic to social change that these changes are in some ways seen as marking the decline or end of something. Behind such narratives are a certain fear and often the notion of some golden age in the past against which the present can and should be judged.

Hoewel die indruk gewek kan word dat hierdie studie van 'n krisis-narratief as vertrekpunt gebruik maak, is dit nie die geval nie. Die redes hiervoor sal duidelik blyk uit die res van hierdie proefskrif (veral die teoretiese raamwerk wat in Hoofstuk 2 uiteengesit word) waar die argument gevoer sal word dat interpretasie altyd gekontekstualiseer is. Die opmerking van Ivanič *et al.* hierbo moet dus verreken word teen 'n raamwerk en konteks wat weliswaar ook verskil van die een waaroor hulle kommentaar lewer.

1.3 Rasionaal vir die studie

Bostaande situasie laat universiteite met geen ander keuse nie as om meganismes in werking te stel om swak voorbereide studente te ondersteun (Botha & Cilliers, 1999:144; Bubb, 1991:18) deur onder andere aandag te skenk aan faktore soos lae

akademiese geletterdheidsvlakke wat akademiese sukses ná inskrywing beïnvloed. Nel (2007:7-8; 221-242), soos reeds genoem, dui aan dat taal 'n invloed op akademiese sukses het, maar lys ook 'n aantal ander faktore wat direk met hoë uitvalsifers verband hou wat vir doeleindes van hierdie studie nie verder bespreek nie, maar volledigheidshalwe slegs hier gelys sal word: die verkeerde keuse van studierigting, swak studiemetodes, probleme met studiefinansiering, selfbeeldprobleme, gebrekkige integrasie by die universiteitslewe, kulturele aanpassingsprobleme, die invloed van 'n alternatiewe of aan studente onbekende onderrigbenadering, klasbywoning, die aanpassing van veral nie-tradisionele eerste generasie-studente, en selfdissipline.

Een sodanige meganisme om swak voorbereide studente te ondersteun, is die instel van taalsentra of taaleenhede, soos die geval by die Universiteite van Stellenbosch en Pretoria en die Noordwes Universiteit. 'n Spesifieke voorbeeld van so 'n taaleenheid, is die instel van die Eenheid vir Akademiese Geletterdheid (EAG) aan die Universiteit van Pretoria in 'n poging om die sukseskoers aan die UP te laat styg deur studente se akademiese geletterdheidsvlakke te verbeter. Hierdie is volgens Weideman (2007c:4) 'n tipiese voorbeeld van derde generasie oplossings vir probleme soos hierbo genoem waar ondersteuningskursusse verpligtend vir alle studente gemaak is. Die eerste generasie oplossing was die instel van generiese akademiese ontwikkelingsprogramme wat meermale in verlengde graadprogramme oorgespoel het. Die tweede generasie oplossing was weer daarop gemik om binne die konteks van die verbreding van deelnamekoerse van histories-benadeelde studente aan hoëronderwys te fokus op dissiplinespesifieke programme soos die "University of Pretoria Foundation Year (UPFY)" wat spesifiek histories-benadeelde studente in die natuurwetenskaplike studierigtings wou ondersteun en oplei.

Die EAG – wat aanvanklik as die Eenheid vir die Ontwikkeling van Taalvaardigheid bekend gestaan het – is in April 1999 in die lewe geroep om studente wat weens 'n te lae vlak van akademiese geletterdheid/taalvaardigheid

gevaar loop om in hulle studie te misluk, te ondersteun deur eerstens die (taal)probleme te identifiseer en dan met ontwerpe vorendag te kom wat die probleme kan aanspreek of oplos, soos tipies van 'n toegepaste taalkundige perspektief op sake verwag kan word. Beleid wat sedert 2000 aan die UP toegepas word, vereis dat alle studente akademies geletterd verklaar moet word alvorens hulle 'n graad aan hierdie universiteit verwerf (Van Dyk & Weideman, 2004a:3).

Die eerste van drie verantwoordelikhedsareas waarmee entiteite soos die EAG aan die UP gewoonlik getaak word, is die evaluering van nuwe eerstejaarstudente se akademiese geletterdheid met behulp van 'n betroubare, geldige en potensieel gestandaardiseerde meetinstrument. Die tweede verantwoordelikhedsarea is om daardie studente wat gevaar loop om nie hulle studie binne die minimum toegelate tyd te voltooi nie, van voldoende (taal)ondersteuning te voorsien deur aan hulle volle kredietdraende taalondersteuningskursusse te bied. Hierdie kursusse word by sommige universiteite by die eerstejaars se leerprogram en rooster ingewerk, en is dus geen bykomende las (toevoeging) vir enige student nie. Laastens kan daar ook van sulke entiteite verwag word om verskeie ander taal- en taalverwante kursusse te bied aan studente wat volgens die meetinstrument vir akademiese geletterdheid nie risikogevalle is nie. Dit mag insluit kursusse soos akademiese lees- en skryfvaardigheidsontwikkeling op 'n hoër vlak as by die genoemde taalondersteuningskursusse, regsdiskoers, kommunikasiekursusse, ensovoorts.

Die sukses van die eenheid aan die UP en haar werksaamhede (sien Weideman 2003c vir 'n oorsig) het daartoe bygedra om ander universiteite te oortuig om soortgelyke modelle te ondersoek en uiteindelik in werking te stel. Beide die Universiteit Stellenbosch en die Noordwes Universiteit het oor tyd heen die genoemde model met klein aanpassings vir hul spesifieke kontekste suksesvol geïmplementeer, terwyl daar ook aan ander universiteite voorbeelde van ingrypings van 'n min of meer soortgelyke aard bestaan. Botha en Cilliers (1999:144) voer aan dat sulke aksies in ooreenstemming is met internasionale en nasionale tendense

waar universiteite poog om hulle eie standarde te identifiseer ingevolge waarvan hulle voornemende studente se voorbereidingsvlak kan vertolk.

1.4 Bestek van die studie

Hierdie studie sal in besonder op die eerste verantwoordelikhedsarea van intervensies soos die van UP se Eenheid vir Akademiese Geletterdheid (weereens as voorbeeld van derde generasie oplossings) fokus, naamlik die vasstelling van studente se vlak van akademiese geletterdheid vir plasingsdoeleindes aan die Universiteit Stellenbosch, die Noordwes Universiteit en die Universiteit van Pretoria deur middel van 'n betroubare, geldige en potensieel gestandaardiseerde meetinstrument.

Soos reeds genoem, het die Eenheid vir akademiese geletterdheid aan die Universiteit van Pretoria jaarliks sedert 2000 alle eerstejaarstudente 'n toets van akademiese geletterdheid/taalvaardigheid laat aflê ten einde (a) hulle risikovlak te bepaal, en (b) diegene te identifiseer en te plaas wat vir bykomende ondersteuning in toepaslike kursusse kwalifiseer. Aangepaste modelle hiervan is reeds by die Universiteit Stellenbosch en die Noordwes Universiteit sigbaar.

Tot en met 2003 het die Eenheid 'n toets gebruik wat kommersieel beskikbaar is en deur die Hough & Horne-geletterdheidskonsultante ontwikkel is. Hierdie toets, die ELSA (English Literacy Skills Assessment), is aanvanklik met die oog op die industriële en kommersiële sektor ontwikkel. Die ontwikkelaars en die EAG het dit egter later vir gebruik in die hoërondwyssektor verfyn en aangepas. Die “nuwe” toets, die ELSA-Plus (aangepas vir hoër onderwys en opleiding), het – in ooreenstemming met die aanvanklike toetsdoelwit – die insetvlakke van bevoegdheid en opleibaarheid van studente wat doeltreffend in 'n Engelse en Afrikaanse taalomgewing moet funksioneer, geëvalueer (Van Dyk & Weideman, 2004a:3).

1.4.1 Beskrywing en konstruksie van die ELSA-Plus

In Van Dyk en Weideman (2004a:3) se bespreking van die ELSA-Plus beklemtoon hulle die volgende aspekte: die ELSA-Plus is 'n vaardigheidsgebaseerde, normgerigte (met as norm die eerstetaal- Engelssprekende vir die Engelse toets, of Afrikaanssprekende vir die Afrikaanse toets), gestandaardiseerde toets wat vir plasingsdoeleindes gebruik word. Dit beskik na bewering oor 'n voorspellingsgeldigheid van 84%, en 'n betroubaarheidsyfer van 0.86 (Horne, 2002). Verder behels dit 'n uurlange skriftelike evaluering wat objektief bepunt word. Dit kwantifiseer voornemende studente se taalvaardigheid in terme van skoolgrade, met ander woorde 'n aantal jaar se formele skoolopleiding: alle eerstejaarstudente wat byvoorbeeld 'n onvoldoende vlak van taalbedrewenheid getoon het (graad 10 of laer – dus gelykstaande aan tien jaar of minder se formele skoling) moes vir verpligte, volle kredietdraende taalkursusse by die EAG inskryf (Van Dyk & Weideman, 2004a:4). Indien 'n student 'n graad 11-vlak of hoër behaal het, het dit daarop gedui, so is daar geoordeel, dat hy/sy nie enige bykomende taalondersteuning nodig het nie. Die ELSA-Plus word ook as kultureel regverdig beskou, aangesien dit metataal, spreektaaluitdrukkings, idiomatiese uitdrukkings en dialektiese vorme vermy (Horne, 2002). Op makrovlak bestaan die ELSA-Plus uit sewe afdelings:

- geluidsleer (die vermoë om Engelse/Afrikaanse klanke te herken en te onderskei);
- diktee (die vermoë om gesproke Engels/Afrikaans neer te skryf en akademiese skryfgebruik te eerbiedig);
- basiese gesyferdheid (waar 'n vertroudheid met syfers by taalgebruik geïntegreer word);
- leesbegrip (op beide elementêre en intermediêre vlak);
- taal en grammatika van ruimtelike verwantskappe;
- 'n sluitingstoets (die vermoë om 'n betekenisvolle en samehangende geheel te skep); en
- woordeskat binne konteks (die vermoë om inligting uit 'n bepaalde konteks te haal om betekenis aan sekere woorde of sinsnedes te heg).

Uit hierdie beskrywing blyk dit duidelik dat die toets van die veronderstelling uitgaan dat (akademiese) taalvaardigheid die beste in terme van die strukturele komponente daarvan omskryf word, naamlik dat taal 'n kombinasie van klank, vorm en betekenis is. Verder word aanvaar dat taalgebruik in terme van 'n aantal "vaardighede" vasgevang kan word- in hierdie gevalluister-, lees- en skryfvaardighede.

1.4.2 Die geskiktheid van die ELSA-Plus bevraagteken

Die gebruik van hierdie toets is om 'n aantal redes bevraagteken, onder andere weens die logistieke beperkinge met betrekking tot die afneem daarvan: die behoefte aan gesofistikeerde klanktoerusting, stophorlosies en 'n uitgerekte nasientydperk vir die groot aantal studente (ongeveer 7 000 aan die Universiteit van Pretoria alleenlik) wat die toets aflê (Van Dyk & Weideman, 2004a:4), gewoonlik binne die bestek van 10 uur.

'n Teoretiese oorweging, en van selfs groter belang, was dat die konstruk van die ELSA-Plus op 'n verouderde en beperkte siening van taalvermoë berus waarvolgens taalvermoë, soos hierbo genoem, gelyk gestel is aan die kennis van klank, vorm, woordeskat en betekenis, of soos Van Dyk en Weideman (2004a:4,5) dit in tegniese-linguistiese terme stel, fonologiese, morfologiese, sintaktiese en semantiese elemente. Shohamy (2004) waarsku ook teen die gevare daaraan verbonde om taaltoetse te gebruik wat 'n beperkte definisie van taalkennis weerspieël en bestaande sienings daarvan weerspreek (meer besonderhede verskyn in Hoofstuk 3). Voorts is die gebruik van die moedertaalspreker as norm vanselfsprekend 'n kontensieuse kwessie waarop daar nie hier verder ingegaan sal word nie.

'n Aantal ander implikasies vir toetsing (wat ook in meer besonderhede in Hoofstukke 3 en 4 aan bod kom) moes ook oorweeg word: die belangrikste hiervan is dat toetse interaktiewe moontlikhede (byvoorbeeld met die self of met die teks) behoort in te sluit ten einde die vermoë van die kandidaat te kan bepaal om binne bepaalde kontekste betekenis te kan voortbring (Bachman & Palmer, 1996:61; Blanton, 1994:221). Gevolglik is 'n werklike toets van akademiese geletterdheid

dus eerder die meting van studente se kommunikatiewe vermoë binne akademiese verband as 'n toets wat slegs vaardigheidsvlakke meet. Bachman en Palmer (1996:75-76) kom tot die gevolgtrekking dat

... language skills [should not be considered] to be part of language ability at all, but to be the contextualized realization of the ability to use language in the performance of specific language use tasks. We would ... argue that it is not useful to think in terms of 'skills', but to think in terms of specific activities or tasks in which language is used purposefully.

Vierdens was die ELSA-Plus duur om te administreer, en daar was geen beheer oor die stygende koste daarvan nie. Om hierdie en ook 'n aantal ander geringer redes was dit vir die Universiteit van Pretoria raadsaam om 'n ander toets van akademiese geletterdheid/akademiese taalbedrewenheid te ontwikkel wat 'n akkurater en meer bevredigende — in die sin van 'n teoreties meer regverdigbare — weerspieëling kon bied van die mate waarin studente die vermoë aangeleer het om akademiese diskoers te genereer en te vertolk. Om 'n alternatiewe toets te ontwikkel, is 'n vennootskap gesluit met die Noordwes Universiteit en die Universiteit Stellenbosch, sodat daar genoeg deskundigheid byeen gebring kon word uit alle belanghebbende instellings, en veral meertalige universiteite. Hierdie vennootskap is vandag geformaliseer in die Inter-institutional Centre for Language Development and Assessment (ICELDA) wat ook die Universiteit van die Vrystaat insluit.

1.4.3 Teoretiese begroning vir die studie

Die bestek van hierdie studie het as grondslag die toegepaste taalkunde (McNamara, 2004:763). McNamara & Roever (2006:255) stel dit dat “[l]anguage testing is not simply applied psychometrics, but, rather, a central area of applied linguistics.” Die kern van hierdie dissipline lê in die ontwerp van linguistiese artefakte ten einde oplossings te bied vir (taal)probleme, waarop die uiteensettings en bespreking in Hoofstukke 2, 3 en 4 meer lig sal werp. Dit kan egter voorlopig gestel word dat die toegepaste taalkunde as 'n dissipline van ontwerp beskryf kan word (Weideman, 2006:2) – die ontwerp van oplossings vir (talige) probleme. Dieselfde outeur bied in 2009 'n uitgebreide raamwerk (Weideman, 2009a) aan wat die konseptuele onderbou van hierdie studie sal wees. Hierin beklemtoon hy dat daar balans in enige tegniese ontwerp moet wees en dat daar nie nét op enkele aspekte van

toetsontwerp gefokus kan word nie, soos wat meermale die geval met taaltoetsontwikkeling is. Die balans waarna hy verwys, dui op 'n sisteem van ontwerpbeginsels met konstitutiewe (grondliggende) beginsels aan die een kant en regulatiewe (begeleidende) beginsels aan die ander kant. Hoewel hierdie studie in hoofsaak op 'n ondersoek na die konstitutiewe beginsels van 'n tegniese ontwerp (in hierdie geval 'n toets) sal berus, kan dit nie die regulatiewe beginsels ignoreer nie, aangesien die twee kante van die raamwerk in wisselwerking met mekaar is, of deur mekaar beïnvloed word.

1.5 Probleemstelling

Nadat drie verskillende, hedendaagse toetskonstrukte ondersoek en ontleed is (meer besonderhede verskyn in Hoofstuk 3), is daar in 2003 'n nuwe konstruk (Weideman, 2003a:xi; Van Dyk & Weideman, 2004a:7-12) vir 'n plasingstoets van akademiese geletterdheid voorgestel. Hierdie toetskonstruk het mettertyd in die Toets van Akademiese Geletterdheidsvlakke (TAG) gemanifesteer – die Engelse weergawe van TAG is TALL (Test of Academic Literacy Levels).

Pluspunte van TAG/TALL, soos dit ook met die ELSA-Plus die geval was, is dat dit slegs 'n uur neem om af te lê, en dit word ook objektief bepunt. Bykomende voordele is dat dit nie slegs taalvaardigheid en -kennis meet nie, maar ook akademiese geletterdheid, en dat dit terselfdertyd koste-effektief is. Dit poog ook om potensieel kontensieuse kwessies so ver moontlik te vermy. Aanduidings is boonop dat dit uiters betroubaar (meer hieroor in Hoofstuk 5) en geldig (verwys Hoofstukke 6 en 7) is.

Volgens Van Dyk en Weideman (2004a:11) wil dit verder voorkom asof die ervaringe en verwagtinge van akademici van oor verskillende dissiplines heen in die voorgestelde nuwe TAG/TALL-konstruk weerklank vind. In Januarie 2004 is TAG/TALL vir die eerste keer met groot sukses aan die Universiteit van Pretoria en die Noordwes Universiteit (Potchefstroom- en Vaaldriehoekcampusse) afgeneem. 'n Tweede weergawe van TAG/TALL, gegrond op dieselfde konstruk, is met selfs

beter resultate in 2005 en 2006 by sowat 14 000 studente aan die Universiteit Stellenbosch, die Noordwes Universiteit (Potchefstroom- en Vaaldriehoekcampusse), en die Universiteit van Pretoria afgeneem.

As gevolg van die toenemende belangstelling in TAG/TALL is die genoemde pluspunte en 'n oënskynlike hoë betroubaarheidsindeks egter nie meer voldoende bewys van hierdie toets se waarde nie. Dit moet, veral binne konteks van 'n toegepaste linguistiese raamwerk waar daar voortdurend na teoretiese fundering vir ontwerpe gesoek word, aan die hand van meer gesofistikeerde analyses voldoende betroubaar verklaar en gevalideer word, sodat dit potensieel standaardiseerbaar is. Davies (1990:5) stel dit onomwonde dat nie net betroubaarheid nie, maar ook geldigheid tot metingsakkuraatheid bydra. Metingsakkuraatheid is weer 'n voorvereiste vir enige toets, veral vir hoëbelangtoetse (sogenaamde 'high-stakes tests'). 'n Gestandaardiseerde weergawe van 'n toets dra op sy beurt weer tot groter geldigheid by. Daar bestaan dus 'n noodwendige verband tussen die betroubaarheid en geldigheid van 'n toets wat versterk word deur die potensieel gestandaardiseerde aard daarvan. Die verband tussen betroubaarheid, geldigheid en ander konsepte is egter 'n ingewikkelde een, en die kompleksiteit daarvan vereis dat daar telkens weer in die verbygaan en soms heel spesifiek hierna verwys sal word, ten einde reg te laat geskied aan die verskillende (en histories relevante) gesigspunte in hierdie opsig.

Die toename in belangstelling in die toets dui volgens Van Dyk en Weideman (2004a:11) daarop dat dit nie noodwendig tot enkele, plaaslike kontekste beperk hoef te word nie, en dat die toets se potensiaal om die lewens van al hoe meer studente te raak ongetwyfeld groter is as wat aanvanklik daarmee beoog is. Die bewys hiervoor lê in die feit dat die Universiteit Stellenbosch hierdie toetse vanaf 2008 as deel van dié universiteit se toegangstoetsbattery implementeer, wat die status daarvan onmiddellik van laebelang- na medium- of hoëbelangtoetse verhoog – ten tyde van hierdie studie is daar nog nie 'n finale besluit oor die gebruik van die nuwe National Benchmark Tests (NBT's) as toelatingsinstrument by die Universiteit Stellenbosch geneem nie.

Die gevolglike verhoging in status is onder andere die probleem waarop hierdie navorsing gerig is. Die verhoogde status lei noodwendig daartoe dat, soos reeds geïmpliseer, daar meer omsigtig met TAG/TALL omgegaan moet word en dit derhalwe so spoedig moontlik aan gesofistikeerde betroubaarheids- en ander ondersoek onderwerp moet word. Die toetse behoort in die verlengde hiervan gevalideer te word ten einde 'n diepgaande teoretiese fundering daar te stel wat die mees billike uitkoms moontlik vir die gebruikers daarvan behoort te verseker. Let egter daarop dat dit wat met hierdie studie beoog word, soos hierbo beskryf, slegs op die Afrikaanse akademiese geletterdheidstoets (TAG) van toepassing sal wees en voortaan sal daar dus oorwegend na TAG verwys word, tensy 'n betrokke punt ook goed geïllustreer kan word met verwysing na TALL, of waar dit ook direk op TALL van toepassing kan wees.

Die doelwitte van die studie is daarom om TAG/TALL aan die bogenoemde ondersoek te onderwerp ten einde vas te stel hoe die teoretiese fundering daarvan in onder andere die tegniese betroubaarheid, geldigheid en potensieel gestandaardiseerde aard daarvan bereik kan word.

1.6 Formulering van hipotese

TAG, die toets van akademiese geletterdheidsvlakke, is dus aanvanklik as 'n sogenaamde plasingstoets met lae belang ontwikkel, maar weens die toenemende vraag daarna kan dit heel moontlik 'n medium- of selfs hoëbelangtoets word, aangesien dit ook vir ander doeleindes aangewend kan word. Soos hierbo genoem, vervang die Universiteit Stellenbosch byvoorbeeld vanaf 2008 die bestaande taalkomponent van dié universiteit se toegangstoetsbattery met TAG én TALL (studente is verplig om beide die Afrikaanse en Engelse akademiese geletterdheidstoetse af te lê). Dit bring mee dat die toetse onmiddellik die status van medium- tot hoëbelangtoetse begin aanneem. In aansluiting by die probleemstelling hierbo, is dit daarom van groot belang dat die toetse (vir doeleindes van hierdie studie beperk tot TAG) se betroubaarheid bepaal, en dat dit vanuit 'n verskeidenheid van gesigspunte gevalideer (bekragtig) word. Die hipotese

van hierdie studie is dus dat TAG 'n linguistiese artefak is wat 'n taalprobleem aanspreek en dat die teoretiese fundering daarvan in sy tegniese betroubaarheid, geldigheid en potensieel gestandaardiseerde aard lê. Dié ondersoek sal derhalwe poog om aan te toon dat die toets robuus genoeg is om die druk (vir groter betroubaarheid en, in besonder, geldigheid) wat die aanwending daarvan as 'n hoërbelangtoets moontlik tot gevolg kan hê, te kan weerstaan.

1.7 Navorsingsmetodologie

1.7.1 Navorsingsbenadering

Hierdie studie sal van 'n sogenaamde 'mixed methods approach' gebruik maak. Dit sal daarom eerstens vanuit 'n sekondêre perspektief benader word, wat relevant vir hierdie studie is, omdat die studie as geheel 'n tegnies-wetenskaplike regverdiging vir die toets behoort te bied. Dit sluit in 'n literatuurstudie vir doeleindes van 'n filosofiese en konseptuele toegepaste linguistiese ontleding op die terrein van toetsontwikkeling en akademiese geletterdheidstoetsing, en in die besonder betroubaarheid, geldigheid, en validering. Hierdie sekondêre benadering tot navorsing sal egter vir doeleindes van hierdie studie deurlopend aangevul word met primêre, empiriese navorsing van veral oorspronklike data aan die hand van kwantitatiewe ondersoeke, wat in wese op verskillende statistiese tegnieke berus (J.D. Brown, 2004:479).

Binne konteks van hierdie twee perspektiewe tot navorsing in die toegepaste linguistiek, sal daar van die tegniek van 'variable operationalization research' gebruik gemaak word, waar "the researcher clearly outlines the variables of interest from the outset ... and explains how it was operationalized, that is, how each one was observed or measured and quantified" (J.D. Brown, 2004:485).

Die genoemde veranderlikes kan egter nie net vanuit 'n statistiese perspektief beoordeel en benut word om 'n teoretiese basis vir regverdiging van 'n bepaalde linguistiese artefak (in hierdie geval 'n toets van akademiese geletterdheid) te skep

nie. Dit moet ook binne die gebruikskonteks geïnterpreteer en beskryf word en daarom sal die kwalitatiewe veranderlikes nie agterweë gelaat kan word nie.

Die metodologiese paradigmas waarteen hierdie studie gesitueer sal word, sluit gevolglik kwantitatiewe én kwalitatiewe ondersoeksmetodes in. In artikulasie met Weideman (2009) se filosofiese raamwerk, wat later in meer besonderhede bespreek sal word, wys J.D. Brown (2004:486-491) daarop dat 'n digotomiese hantering van die kwantitatiewe en die kwalitatiewe egter nie wenslik is nie, aangesien die twee begrippe as die pole van 'n kontinuum gesien behoort te word, omdat dit met mekaar interakteer, en omdat dit mekaar in die reël komplimenteer. Hierdie wedersydse komplimentering is omdat die een (in hierdie geval byvoorbeeld die kwalitatiewe/interpretatiewe) gebruik word om die ander (in hierdie geval die kwantitatiewe/statistiese) te help interpreteer en te verklaar, die een gebruik word om te definieer en die ander om te operasionaliseer, die een induktief en die ander deduktief van aard is, die een hipotese-vormend te werk gaan en die ander hipotese-toetsend, ensovoorts (J.D. Brown, 2004:490).

Brown gee verder te kenne dat navorsing van 'n hoë standaard in die toegepaste linguistiek aan die kwantitatiewe kant van die kontinuum, “[represents] the concepts of reliability, replicability, validity, and generalizability [and that these are] the standards quantitative researchers hold up in judging the soundness of their research” (J.D. Brown, 2004:492). Aan die kwalitatiewe kant van die kontinuum is “concepts of dependability, confirmability, credibility and transferability” weer van belang (J.D. Brown, 2004:494). Daar is natuurlik ook ander soos deursigtigheid, verantwoordbaarheid, sorgsaamheid en bruikbaarheid en daar sal mettertyd uiteengesit word hoe hierdie toegepaste taalkundige grensbegrippe betrekking het op hierdie studie.

Daar sal in hierdie studie, en uit die aard van die saak, meer aandag aan die kwantitatiewe gegee word (in lyn met die hipotese hierbo gestel), maar met die deurlopende inagneming van die kwalitatiewe. Daar sal dus pertinent op kwessies soos betroubaarheid en geldigheid gefokus word, aan die hand van byvoorbeeld

replikasie-ondersoeke (waarvan sommige met behulp van ander tegnieke aangepak sal word), maar daar sal ook gespekuleer word oor die moontlikheid dat die uitkomst van hierdie studie oor die algemeen toepasbaar en bruikbaar sal wees in ander kontekste, oor die bewysbaarheid en geloofwaardigheid van die resultate en die oordraagbaarheid daarvan na ander gebiede.

So 'n benadering is ook in lyn met die werk van Weideman (2003d, 2006, 2007a, 2009a, 2009b) waarin hy aanvoer dat toegepaste linguistiese artefakte aan die een kant tegnieke ontwerpe is wat deur middel van die analitiese dimensie van ons ervaring 'n rasionaal of teoretiese regverdiging vir die spesifieke ontwerp bied (die sogenaamde basisfunksie van die ontwerp), wat onder meer in die kwantitatiewe dimensies van die ontwerp gevind word. Aan die ander kant weer, wys hy daarop dat toegepaste linguistiese artefakte tegnieke ontwerpe is wat ontsluit of verfyn moet word deur onder andere die toepasbaarheid, geloofwaardigheid en legitimiteit daarvan te ondersoek en te artikuleer (die sogenaamde regulatiewe of leidende funksies van die ontwerp). Hierdie interpretasie van die toegepaste linguistiek word in meer besonderhede in Hoofstuk 2 bespreek. Tabel 1.1 is 'n opsomming van die werk van J.D. Brown en Weideman hierbo bespreek, met die inagneming van die spesifieke konteks van hierdie studie:

Tabel 1.1: Fokusareas van die navorsingsbenadering van hierdie studie met spesifieke toepassing op TAG

	Kwalitatief		Kwantitatief
regulatiewe	Afhanklikheid van veranderlikes	↔ konsistensie ↔	Betroubaarheid
	Bevestigende aard	↔ verifieerbaarheid ↔	Repliseerbaarheid
	Geloofwaardigheid	↔ waarborg/sekerheid ↔	Geldigheid
	Oordraagbaarheid	↔ betekenisvolheid ↔	Veralgemeenbaarheid

}

}

regulatiewe

konstitutiewe

1.7.2 Onderzoeksgroep en datagenerering

Die etlike duisende eerstejaarstudente wat TAG jaarliks in 2004, 2005, 2006, 2007, 2008 en 2009 aan die Universiteit Stellenbosch, die Noordwes Universiteit en die Universiteit van Pretoria afgelê het, dien as die populasie van hierdie studie. Al hierdie studente is oor die moontlike nadele (indien enige) en die uitwerking van deelname aan die navorsing ingelig, en almal het daartoe ingestem dat hulle toetsdata ondersoek en daarvoor verslag gelewer kan word sonder om hulle as individue te identifiseer.

1.7.3 Dataontleding

Soos reeds in paragraaf 1.7.1 hierbo genoem, sal 'n aantal statistiese ontledings gedoen word, en sal dit binne verskeie teoretiese raamwerke wat in die literatuur beskikbaar is en vir hierdie doeleinde geïdentifiseer word, binne konteks geplaas en vertolk word. Uit hierdie vertolking sal sekere gevolgtrekkings oor die betroubaarheid en geldigheid van TAG/TALL gemaak word.

1.8 Begripsverheldering

Akademiese geletterdheid

Die begrip *akademiese geletterdheid* is kompleks om te omskryf. 'n Geïntegreerde, deeglike omskrywing verskyn in Hoofstuk 3 van hierdie dokument. In kort verwys *akademiese geletterdheid* egter na die kognitiewe prosesse en aksies wat studente daartoe in staat stel om aan 'n diskoers wat op akademiese denke, leeswerk, skryfwerk, luister- en praatvaardigheid betrekking het, deel te neem. Let daarop dat die terme akademiese geletterdheid en akademiese taalvaardigheid in wisselwerking met mekaar gebruik word en vir doeleindes van hierdie studie as sinonieme gelees moet word.

Betroubaarheid

Betroubaarheid is in terme van die teoretiese raamwerk waarbinne hierdie studie aangepak word, tesame met geldigheid twee van die konstitutiewe/essensiële begrippe of maatstawwe waarvolgens toetse beoordeel word. *Betroubaarheid*

verwys na die werklike ooreenstemming, met ander woorde die interne konsekwentheid, tussen resultate van 'n toets en die toets self, of 'n ander toets (Davies, Brown, Elder, Hill, Lumley, McNamara, 1999:168). Babbie en Mouton (2006:119) beskryf dit egter nog eenvoudiger as “a matter of whether a particular technique, applied repeatedly to the same object, would yield the same result each time.” Dit word in besonderhede in Hoofstuk 5 bespreek.

Geldigheid

In die tradisionele sin van die woord is 'n toets *geldig* indien dit meet wat dit veronderstel is om te meet. Volgens Davies *et al.*, (1999:221) word die geldigheid van 'n toets bepaal deur die mate waartoe dit daarin slaag om 'n akkurate, konkrete uitdrukking te gee aan 'n abstrakte konsep, byvoorbeeld akademiese geletterdheid. Die soorte geldigheid waarna daar meestal verwys word, is inhouds-, konstruk-, samevallende en voorspellingsgeldigheid. Die eerste twee is konseptueel, terwyl laasgenoemde twee statisties van aard is (Davies *et al.*, 1999:221); daar is egter nog ander tipes geldigheid ook. Babbie en Mouton (2006:123) definieer geldigheid as “the extent to which an empirical measure adequately reflects the real meaning of the concept under consideration.” Die idee van konsekwensiële geldigheid (soms ook gevolgsgeldigheid genoem) as 'n verdere soort geldigheid, wat deur Messick (1994) voorgestel is, blyk 'n verdere verruiming van die konstitutiewe begrip van geldigheid te wees. Messick (1988) is egter ook van mening dat daar slegs een konsep van geldigheid kan wees, en dat alle verskillende ‘tipes’ geldigheid wat tradisioneel onderskei is, verenig moet wees. Verder meen hy dat geldigheid nie 'n kenmerk of attribuut van 'n toets kan wees nie: geldigheid berus op die interpretasie van die resultate van 'n toets en is nie 'n kenmerk van 'n toets nie (Messick, 1988; 1989). Derhalwe pleit hy vir die validering van 'n toets as 'n proses, eerder as om bloot 'n toets (die produk) ‘geldig’ te verklaar, al dan nie (Messick, 1988; 1989) – sien ook hieronder: validering. Die probleem met Messick se pleidooi dat 'n toets gevalideer moet word, is dat hy in sy vertolking hiervan nooit die uitkoms van die proses van validering duidelik vertoon nie, naamlik dat 'n toets (en nie die interpretasie daarvan nie) geldig kan wees – dit word in Hoofstukke 6 en 7 in

besonderhede bespreek. In terme van die teoretiese raamwerk waarbinne hierdie studie aangepak word, sal daar egter ook verwys word na die verdere ontplooiing en alternatiewe konseptualisering van Messick se eens invloedryke siening met betrekking tot geldigheid.

Validering

Validering is die proses waardeur die geldigheid van 'n toets bevestig word. Daar word algemeen aanvaar dat geldigheid deur hierdie proses op soveel moontlik maniere bevestig moet word. Vertroue in 'n toets staan regstreeks in verhouding tot die aantal bewyse wat ter ondersteuning van geldigheid ingewin word (Davies *et al.*, 1999:220). Nogtans is dit nie slegs die aantal bewyse op sigself wat belangrik is in die proses van validering nie, maar die rasonale wyse waarop die veelheid van bewyse tot 'n eenheid geïntegreer kan word in die vorm van 'n argument om die geldigmakende proses te regverdig (Weideman, 2009a).

1.9 Uiteensetting van die studie

Hoofstuk 1 bied die agtergrond tot hierdie studie, en verskaf 'n aanvanklike en opsommende uiteensetting van waarom dit nodig was vir sub-institusionele eenhede soos die Eenheid vir Akademiese Geletterdheid aan die Universiteit van Pretoria om 'n verantwoordbare konstruk en toets (TAG / TALL) vir die meting van akademiese geletterdheidsvlakke in samewerking met 'n stel trans-institusionele vennote te ontwikkel. Dit voer verder aan dat TAG grondig ondersoek moet word met betrekking tot die betroubaarheid en geldigheid daarvan, weens veral die toenemende belangstelling wat 'n aantal tersiêre instellings daarin toon as 'n medium- tot hoëbelangtoets. Dit raak ook die metodologiese benaderinge aan wat vir die uitvoering van hierdie studie gebruik sal word.

Hoofstuk 2 dien as teoretiese vertrekpunt vir die res van die studie in die opsig dat dit 'n oorsig oor die toegepaste linguistiek bied en veral op huidige benaderings in hierdie dissipline fokus in soverre dit relevant is vir taaltoetsing as onderdeel van

die toegepaste linguistiek, in die algemeen, sowel as vir die spesifieke soort toetsing wat hier ter sprake is, naamlik die meting van akademiese geletterdheid.

Hoofstuk 3 is 'n bespreking van 'n toegepaste linguistiese artefak wat ontwerp is om 'n (taal)probleem aan te spreek, soos in Hoofstuk 2 aangedui. Dit bied 'n historiese oorsig oor benaderings tot taaltoetsontwikkeling en die begrip akademiese geletterdheid word gedefinieer. Die kern van die hoofstuk bespreek die ontwikkeling van 'n konstruk vir 'n akademiese geletterdheidstoets, soos gemanifesteer in TAG.

Hoofstuk 4 is 'n verruiming van Hoofstuk 3 deurdat dit een van die grootste uitdagings van toetsontwerp aanspreek, naamlik die belyning van die toetskonstruk met die toetsspesifikasies en gepaardgaande toetsitems. 'n Spesifieke afdeling van TAG word as voorbeeld gebruik om aan te toon hoe die toetskonstruk wat in Hoofstuk 3 bespreek is, manifesteer in toetsspesifikasies en toetsitems.

Hoofstuk 5 fokus veral op die kwantitatiewe elemente van TAG deurdat dit ondersoek instel na die betroubaarheid van hierdie. 'n Aantal statistiese tegnieke word gebruik om toetsdata te analiseer op grond waarvan daar uiteindelik argumente aangebied sal word oor die konsistensie van TAG.

Hoofstuk 6 konsentreer op die konsep geldigheid wat, naas betroubaarheid, as een van die sentrale nosies in ondersoeke na toetse geag word. Hierdie hoofstuk bied derhalwe 'n bespreking van bogenoemde begrip wat verband hou met die krag of effek van toetse soos TAG en TALL. Benewens 'n uiteensetting van die tradisionele beskouing tot geldigheid en die huidige ortodoksie, word geldigheid egter ook vanuit 'n meer resente, verruimde standpunt daarvoor bespreek, naamlik dat dit terselfdertyd 'n attribuut van 'n toets kan wees en as konstitutiewe begrip ontsluit kan word deur verdere (regulatiewe) idees met betrekking tot taaltoetsing.

Die vorige hoofstuk is teoreties-inleidend tot Hoofstuk 7 wat konsentreer op toetsbekragtiging of validering in dié sin dat dit verskillende soorte geldigheid en

gelyklopende bekragtigingsprosesse ondersoek ten einde vas te stel of TAG werklik meet wat dit voorgee om te meet, naamlik akademiese geletterdheid. 'n Aantal statistiese tegnieke word gebruik om toetsdata te orden en te vertolk, op grond waarvan sekere gevolgtrekkings oor die geldigheid van TAG gemaak kan word.

In aansluiting by die vorige drie hoofstukke, bied Hoofstuk 8 'n versoerende perspektief op die interpretasie van die resultate van TAG en die bruikbaarheid en toepaslikheid daarvan binne die institusionele konteks(te) waarin dit aangewend word, en wel teen die agtergrond van die teoretiese raamwerk van hierdie studie, wat in Hoofstuk 2 uiteengesit is.

Laastens bied hierdie studie in Hoofstuk 9 'n opsomming van insigte, bevindinge en gevolgtrekkings oor veral die konsistensie, die verifieerbaarheid, die waarborge en die betekenisvolheid van TAG as meetinstrument van akademiese geletterdheidsvlakke. Dit lig verskeie aanbevelings en toepassingsmoontlikhede uit vir die aanpassing en afneem van TAG, asook vir verdere navorsing.

Die toegepaste linguistiek: 'n dissipline van ontwerp

2.1 Inleidend

In Hoofstuk 1 is daar aangetoon dat die swak voorbereide studente waarmee universiteite gekonfronteer word 'n negatiewe invloed op deurvloei koerse het. Daar is aangevoer dat daar rede is om te glo dat lae vlakke van akademiese geletterdheid, as een van 'n aantal veranderlikes, hiertoe bydra. Universiteite is derhalwe genoop om identifiserings- en ondersteuningsmeganismes in plek te stel om probleme soos hierdie aan te spreek. Die gevolglike plaaslike en internasionale neiging is die stig van entiteite wat onder andere met hierdie problematiek gemoeid is. Voorbeelde van sulke entiteite in Suid-Afrika is die Eenheid vir akademiese geletterdheid in die Fakulteit Geesteswetenskappe aan die Universiteit van Pretoria, die Taalsentrum aan die Universiteit van Stellenbosch, en die Sentrum vir akademiese en professionele taalpraktyk aan die Noordwes Universiteit.

Soos ook in Hoofstuk 1 genoem, is hierdie studie veral gefokus op een aspek van die werksaamhede van die genoemde entiteite, naamlik die identifisering van swak voorbereide studente aan die hand van betroubare en geldige meetinstrumente soos TAG (Toets van Akademiese Geletterdheidsvlakke) en TALL (Test of Academic Literacy Levels), waarop Hoofstukke 3 en verder in besonderhede sal uitbrei. Hierdie toetse word hoofsaaklik gebruik om studente op verskillende akademiese

geletterdheidskursusse te plaas na gelang van hul risikvlak. Dit word egter weens 'n aantal redes nie meer nét as plasingsmeganisme gebruik nie, maar ook as deel van toelating, wat dit die status van sogenaamde medium- tot hoërbelangtoetse gee. Die verhoogde status lei noodwendig daartoe dat daar meer omsigtig en verantwoordbaar met die toetse omgegaan moet word, aangesien dit die lewens van verskeie rolspelers (universiteite, studente, ouers, die hoëronderryssektor, beursgewers, die privaatsektor, ensovoorts) kan beïnvloed. Shohamy (2001:16) stel dit juis dat “it is often the performance on a single test, often on one occasion at a single point in time, that can lead to irreversible, far-reaching and high-stakes decisions.” Dit beteken dat daar op 'n deurlopende en deursigtige wyse ondersoek na die toetse gedoen moet word en dat hierdie ondersoeke deeglik geartikuleer moet word. Die doel van hierdie studie is onder andere om artikulasie te gee aan die navorsing wat daar oor die afgelope aantal jaar oor TAG en TALL gedoen is – die fokus is wel op eersgenoemde, maar daar sal by geleentheid ook na TALL verwys word om 'n meer genuanseerde en verhelderende beeld te gee. Dit impliseer ook verder dat hierdie studie nie die laaste woord sal spreek oor die verantwoording van die gebruik van hierdie toetse nie, maar slegs 'n aanvanklike en voorlopige artikulasie daarvan sal kan gee, wat telkens daarna aangevul en verder gevoer sal moet word.

Ten einde 'n verantwoordbare artikulasie van die belangrikste dimensies van TAG te kan gee, is dit egter nodig om 'n teoretiese konteks te skep waarteen die res van die studie gelees behoort te word. Die teoretiese konteks waarop daar besluit is en wat as grondslag vir die res van hierdie studie sal dien, is die van die toegepaste linguistiek, aangesien taaltoetsing onder die toegepaste taalkunde ressorteer. Die rede hiervoor is dat die toegepaste linguistiek veral gemoeid is met die oplos van sosiale probleme wat met taal verband hou (Davies & Elder, 2004:1). Weideman (2006:71) wys egter daarop dat “[a]ppplied linguists seldom appear to take the time off to step back and theoretically consider those ideas that inform their practice”. Dit was hier ook die geval: daar was 'n probleem, naamlik dat studente taalmatig swak voorbereid is vir die akademiese omgewing, maar universiteite het nie geweet

wat die voorbereidingsvlakke was en watter gefokusde ondersteuning om aan watter studente bied nie. As oplossing vir hierdie probleem is TAG en TALL as plasingstoets ontwerp. Hoewel daar al verskeie artikels oor TAG en TALL die lig gesien het, is die sogenaamde storie van die toets (Shohamy, 2001) nog geensins volledig vertel nie. In hierdie verband voer Weideman (2006:71) heel toepaslik aan dat daar gewoonlik 'n dringendheid aan die werk van toegepaste linguïste is en daarom is sulke werk "... often undertaken for the sake of alleviating the language-related struggles of those in need of designed interventions. Perhaps this leaves too little time for reflection, and even less for a deliberate theoretical articulation."

Deur in hierdie hoofstuk op 'n sistematiese wyse 'n teoretiese basis vir die res van die studie uiteen te sit, kan daar op verantwoordbare wyse oor die waarde van TAG (die fokus van die studie) as plasing- en as gedeeltelike toelatingsinstrument besin word. Die onderskeidings wat in hierdie hoofstuk gemaak gaan word, is nogtans slegs breë kategoriserings en konsepte; in die daaropvolgende hoofstukke word van hierdie onderskeidings verder verfyn en gevoer, en spesifiek op taaltoetsing van toepassing gemaak.

2.2 Die toegepaste linguïstiek gedefinieer

Soos reeds hierbo genoem, is die hedendaagse aanvaarding dat die rol van die toegepaste linguïstiek in die breë behels om oplossings te vind vir sosiale probleme wat met taal verband hou. Hierdie opvatting neem die toegepaste linguïstiek veel meer omvattend op as aanvanklike interpretasies daarvan, naamlik dat dit bloot die toepassing van taalteorie is. Brumfit (1997:93) definieer dit byvoorbeeld as "the theoretical and empirical investigation of real-world problems in which language is a central issue." Schmitt en Celce-Murcia (2002:1) bied 'n soortgelyke, maar ietwat meer uitgebreide definisie aan as Brumfit hierbo, naamlik dat "... [it] is 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 gees, maar selfs nog meer genuanseerd is Grabe (2002:9) se definisie dat "the focus of applied linguistics is on trying to resolve language-based problems that people

encounter in the real world, whether they be learners, teachers, supervisors, academics, lawyers, service providers, those who need social services, test takers, policy developers, dictionary makers, translators, or a whole range of business clients.”

Vir doeleindes van ’n breër definisie as dit wat hierbo gebied is, is dit egter nodig om eers die oënskynlike spanning tussen die linguistiek en die toegepaste linguistiek uit te wys. Die vraag wat telkens gevra word, is of die toegepaste linguistiek nie bloot maar ’n onderafdeling van die linguistiek is nie; is dit die blote toepassing van taalteorie, of is die toegepaste taalkunde ’n dissipline in eie reg? Om hierdie vraag te kan beantwoord en om uiteindelik tot ’n gepaste en werkbare definisie vir die toegepaste linguistiek te kom, waarteen die res van hierdie studie gelees kan word, blyk dit nodig te wees om eers kortliks na die geskiedkundige ontwikkeling daarvan te kyk. Weideman (2003d:5) waarsku in hierdie verband egter dat daar gewaak moet word daarteen om een intellektuele tradisie of generasie teen ’n ander af te speel en uiteindelik onvergelykbare eenhede naas mekaar te plaas. Hy voer aan dat “one would need to identify at least one thematic strand that makes each contribution unique, and then seek to discover how, at the same time, there is continuity, when a tradition either anticipates a subsequent one or expands it by taking further a concern of a former tradition.” Vir doeleindes van hierdie studie is die domein van taalonderrig as deurlopende tema gekies, aangesien dit een van die gebiede is waar die meeste navorsing van die toegepaste linguistiek gedoen is, soos dit duidelik hieronder sal blyk. Verder word taaltoetse (hierdie studie se fokus) meestal in die onderrig- en leersituasie gebruik.

2.2.1 ’n Historiese oorsig van die toegepaste linguistiek: drie tradisies

Volgens Davies en Elder (2004:4) word die totstandkoming van die toegepaste linguistiek gekenmerk deur drie tradisies, te wete die Noord-Amerikaanse, die Australiese en die Britse tradisie wat ’n invloed op die vorming van hierdie vakgebied gehad het. Hierdie tradisies word vervolgens kortliks bespreek, waarna

daar na Weideman se meer gefokusde (op die gebied van onderrig en leer) historiese oorsig aandag geskenk sal word.

2.2.1.1 Die Noord-Amerikaanse tradisie

Die Noord-Amerikaanse tradisie van die toegepaste linguistiek word deur Davies en Elder (2004:6) opgesom as een wat aanvanklik begin met 'n sterk fokus op die linguistiek as sulks (taalteorie dus), na een met nie altyd 'n duidelik sigbare verwantskap aan die linguistiek nie, omdat dit sterk oorhel na die toegepaste kant, naamlik om bloot 'n oplossing vir 'n probleem te kry sonder om te veel ag te slaan op die teoretiese verdedigbaarheid van die oplossing. Die skakel tussen die linguistiek en die toegepaste linguistiek bly egter deurlopend sterk en dit is moeilik om in hierdie tradisie 'n duidelike onderskeid te tref tussen die linguistiek en blote toepassings van taalteorie.

2.2.1.2 Die Australiese tradisie

Davies en Elder (2004:6-7) beskryf die Australiese tradisie van die toegepaste linguistiek as een wat aanvanklik as fokus die taal van immigrante en inboorlinge gehad het – by implikasie dus die studie van Engels as 'n tweede- of vreemdetaal. Daar is derhalwe baie sterk op teoretiserings en toepassings van studies van tweedetaalverwerwing (ESL – English as Second Language) en vreemdetaalverwerwing (EFL – English as Foreign Language) teruggeval. Die toegepaste linguistiek in Australië het gevolglik 'n besondere verwantskap met taalonderrig (spesifiek ESL en EFL) getoon en groot name op die gebied van taalverwerwing opgelewer, veral na die totstandkoming van die Britse tradisie van die toegepaste linguistiek.

2.2.1.3 Die Britse tradisie

Die vertrekpunt van die Britse tradisie was volgens Davies en Elder (2004:7) om deur middel van daadwerklike pogings (die stig van organisasies en die publikasie van artikels en boeke) die toegepaste linguistiek as dissipline in eie reg te vestig en dit duidelik te onderskei van die linguistiek as sulks. Kenmerkend van hierdie tradisie (veral in die sestigs en sewentigs van die vorige eeu) is die groot invloed

wat tendense in taalonderrig daarop gehad het. Weideman (2007a:590-591) wys in hierdie verband juis op die invloed wat historiese ontwikkelings in taalonderrig op die toegepaste linguistiek gehad het. Hy merk op dat “[t]o understand the ongoing debate about the nature of applied linguistics, one has to begin with an understanding of the historical beginnings of such work in the realm of language teaching and learning ...” (Weideman, 2007a:591). Hy voer verder aan dat die waardevolle bydraes van taalpraktisyns op ander gebiede van die toegepaste linguistiek egter ook van kennis geneem moet word, maar volstaan daarby dat die meerderheid van materiaal wat hedendaags as voorbeelde van toegepaste linguistiese artefakte gebruik word, uit die taalonderrig, of sub-terreine daarvan, soos taaltoetsing, afkomstig is (Weideman, 2007a:592). Om hierdie rede word daar vervolgens na ’n geïntegreerde en meer gefokusde, hedendaags meer aanvaarbare, beskrywing van die toegepaste taalkunde gekyk, met spesifieke verwysing na taalonderrig.

2.2.2 ’n Historiese oorsig van die toegepaste linguistiek: ’n geïntegreerde, potensieel meer regverdigbare, beskrywing

Om egter by ’n geïntegreerde, hedendaags meer aanvaarbare en regverdigbare beskrywing van die toegepaste linguistiek te kan uitkom, is dit nodig om oorvleuelende en onderskeidende elemente in bostaande tradisies te identifiseer. Die werk van Weideman (1999, 2003d, 2006, 2007a, 2007b, 2009a, 2009b) is hiervoor gebruik, waarin hy die toegepaste linguistiek in sewe generasies of paradigmas opdeel met spesifieke verwysing na taalonderrig.

2.2.2.1 Die eerste generasie

Die eerste generasie van die toegepaste taalkunde is volgens Weideman (2003d:5) gekenmerk deur ’n sterk teoretiese onderbou met die premis dat oplossings vir taalprobleme sogenaamd wetenskaplik van aard moes wees. Dit stem ooreen met Davies en Elder (2004) se identifisering van ’n “Noord-Amerikaanse” tradisie. Die wyse waarop hierdie eerste paradigma se uitgangspunte geformuleer word, laat duidelik blyk tot watter mate die behaviorisme in die tyd hoogty gevier het, waar daar aanvaar is dat wetenskaplike analise sal lei tot die waarheid of ten minste

gesaghebbende resultate, en op die basis daarvan klinkklare intervensies om die verlangde gedrag te bewerkstellig. Dit het sy beslag eerstens gekry in linguistiese analises van die moedertaal en die doeltaal. Dit het ’n invloed op die onderrigbenadering wat bekendstaan as die ‘audio-lingual’-metode in taalklaskamers gehad in die opsig dat daar aanvaar is dat indien ’n komponent van taal herhalend inge oefen word, dit bemeester sal word. Daar is dus aanvaar dat taal makliker bemeester sal word as dit inkrementeel gedoen word (Weideman, 1999:79, 80).

Hierdie eerste generasie van toegepaste taalkundige werk se invloed op taaltoetsing is veral duidelik in die sogenaamde psigolinguistiese benadering tot taaltoetsing wat in die volgende hoofstuk in 3.3.2 in meer besonderhede bespreek word. Die “wetenskaplikheid” van hierdie generasie en gepaardgaande onderrigbenadering het egter met verloop van tyd en vanuit verskeie oorde onder kritiek gekom en latere werk in taalonderrig het die effektiwiteit van so ’n benadering betwyfel (Weideman, 2003d:5). As teenreaksie hierop is daar na ’n volgende generasie van die toegepaste linguistiek beweeg, naamlik ’n linguisties meer uitgebreide paradigma.

2.2.2.2 Die tweede generasie

Hierdie generasie van die toegepaste taalkunde is ’n voortsetting van die eerste generasie in die opsig dat die basis daarvan steeds die linguistiek is, waar analises van die onderdele van taal, byvoorbeeld die fonologie, die morfologie en die sintaksis, steeds sterk gefigureer het (Weideman, 1999:82). Hierdie generasie het egter doelbewuste pogings aangewend om die linguistiese basis daarvan te verruim deur ook op ander aspekte van taal soos teks, diskoers, register en die semantiek te fokus om aan te toon dat taal altyd binne ’n bepaalde konteks gebruik en as ’n instrument van kommunikasie aangewend word (Weideman, 1999:82; 2003d:6). Dit word dan ook tereg die ‘extended paradigm model’ van die toegepaste linguistiek genoem (Sridhar, 1993:5) en kan ook geassosieer word met wat Davies en Elder (2004 en hierbo in 2.2.1.3) die “Britse” tradisie genoem word.

Daar was volgens Weideman (1999:82) egter 'n ongemak met hierdie paradigma, aangesien die toegepaste linguistiek steeds daardeur as onderdeel van die linguistiek beskou is in die opsig dat dit as voorwaarde gehad het dat taal eers geanaliseer moes word alvorens dit onderrig kon word. Hoewel daar ooreenstemming was dat taalfunksies ook belangrik is en nie net taalvorme nie, soos in die werk van die vorige generasie geglo is, was dit nie duidelik presies hoe hierdie taalfunksies aangeleer moes word nie. Juis dit het aanleiding gegee tot die derde generasie wat dan ook heel gepas 'n multi-dissiplinêre model genoem word.

2.2.2.3 Die derde generasie

Die onduidelikheid wat hierbo genoem is, het tot die derde generasie van toegepaste taalkundige werk aanleiding gegee waar daar op die reeds verworwe kennis en insigte van ander dissiplines, byvoorbeeld die opvoedkunde en die sielkunde, gesteun is om die probleem van hoe 'n taal en sy funksies aangeleer behoort te word, aan te spreek. Daar is dus in hierdie styl van toegepaste taalkundige werk nie van die basis wat die eerste twee generasies van die toegepaste taalkunde gelê het, afgewyk nie, maar eerder daarop voortgebou; trouens dit was myns insiens selfs grensverskuiwend. Weideman (2003d:6, 7) voer in hierdie verband aan dat

[a]ppplied linguistics, as a discipline that had to do with the design of language teaching, was thus emancipated from relying solely on input from linguistic theory. [...] One of the most important continuities of this tradition with current, post-modern orientations in applied linguistics lies not so much in the heritage of this trend in its broadening of the base of applied linguistics research, but in paving the way for the rise of a multi-disciplinary tradition in applied linguistics. Though it did not so much generate a research agenda of its own, it consolidated and laid the foundation for technically more sophisticated investigations.

Die uitstaande kenmerk van hierdie paradigma in die toegepaste taalkunde was ongetwyfeld die prominensie wat 'n inter- en multidissiplinêre aanpak verkry het.

2.2.2.4 Die vierde generasie

Die vorige generasie het die basis vir 'n reflektiewe ingesteldheid gelê en juis dit het tot een van die prominente vraagstellings van die vierde generasie gelei. Die vraag wat beantwoord moes word, het gelê in die onsekerheid van die volgorde waarin die eenhede van taal binne bepaalde kontekste aangeleer behoort te word.

Op hierdie punt in die geskiedenis van die toegepaste taalkunde was daar reeds sterker aanduidings van die volgorde waarin die eenhede van die moedertaal aangeleer word en daar is gehipotetiseer dat daar 'n soortgelyke volgorde vir verwerwing van 'n tweedetaal moes bestaan (Weideman, 2003d:7). Navorsing in die verwerwing van 'n tweede taal het dus tot so 'n mate aandag begin geniet dat dit die uitstaande kenmerk van hierdie vierde generasie van die toegepaste linguïstiek geword het. Die navorsing wat hierdie paradigma gestimuleer het, het onder meer daartoe gelei dat daar uiteindelik aanvaar is dat taal die beste in interaksie aangeleer word, met ander woorde op 'n natuurlike wyse. Taalklaskamers en taalonderrigprogramme is derhalwe so ingerig en beplan dat daar in 'n gesimuleerde werklikheid geëksperimenteer kon word.

Hierdie generasie bou nie net voort op die vorige generasies nie, maar is ook alreeds 'n aanduiding van dit wat na verwagting in die volgende generasie moes volg, naamlik 'n sterk fokus op die sosiale aspek van onderrig en leer, wat nie net meer deur eksperimentele navorsing bevestig kon word nie.

2.2.2.5 Die vyfde generasie

In noue aansluiting by die invloedryke en kenmerkende werk oor taalverwerwing in die vierde generasie van die toegepaste taalkunde, bou die vyfde generasie daarop voort. Soos reeds genoem, is hierdie generasie se fokus gesetel in die interaktiwiteit wat die sosiale dimensie van taal kenmerk en word gekarakteriseer deur die konstruktivisme (Weideman, 1999:84). Dit hou in dat wanneer betekenis of begrip gekonstrueer word deur middel van taal, interaksie met ander nodig is (dus binne 'n bepaalde sosiale konteks). Daar was geen duidelike aanduiding dat taal of eenhede van taal in 'n sekere volgorde aangeleer behoort te word nie. Die uitgangspunt was eerder dat taal as't ware ontwikkel word in interaksie met ander. Die fokus is dus nie meer op die taalonderwyser/taaldosent nie, maar op die leerder wat self verantwoordelikheid neem, binne 'n gesimuleerde werklikheid, vir sy/haar ontwikkeling. Weideman (2003d:9) wys daarop dat die waarde van hierdie paradigma veral daarin lê dat dit by implikasie uitwys wat nie in die vorige

generasies of benaderings gewerk het nie, en taalonderrigontwerp in 'n nuwe rigting gestuur het. Die grootste waarde van hierdie generasie is dalk juis dat dit uiteindelik teoretiese regverdiging vir die kommunikatiewe benadering tot taalonderrig kon bied.

2.2.2.6 Die sesde generasie

Teen die einde van die vorige eeu verskyn daar 'n sesde, postmodernistiese, generasie van die toegepaste taalkunde op die horison. Hierdie benadering beklemtoon veral die sosiaal-politiese en etiese dimensies van taalonderrig, en spesifiek ter wille van groter verantwoordbaarheid. Die akkommodering van 'n verskeidenheid van perspektiewe in taalonderrig kom ook in hierdie generasie na vore, wat reeds in die derde generasie deurgeskemer het. Die uitdaging in hierdie wyse van werk lê egter volgens Weideman (2003d:10) nie soseer in die erkenning van ander perspektiewe nie, maar eerder in die identifisering van die konflikterende en oorvleuelende aspekte daarvan ten einde met 'n sosiaal verantwoordbare en teoreties regverdigbare ontwerp vir die probleem vorendag te kom. Weideman (2003d:15-17) wys daarop dat die kern van 'n postmodernistiese tradisie juis is om 'n benadering te volg wat krities dog ook verantwoordbaar is en wat met integriteit gevolg kan word.

2.2.2.7 Die sewende generasie

Uit 'n projekvoorstel wat deur Larsen-Freeman en Weideman (2008) voorberei is, blyk dit dat daar 'n sewende generasie van die toegepaste taalkunde besig is om tot stand te kom, naamlik die van 'n komplekse sisteem-benadering. Dit word bevestig deur Davies (2008:297) en die meer onlangse werk van Weideman (2010:95). Die projekaansoekers hierbo (2008:2) dui egter aan dat dit op hierdie stadium nog nie duidelik is "... precisely how a complex systems approach coheres with previous styles of working within the discipline, or exactly what redirection of applied linguistics its adoption will entail", maar dat ondersoek daarna nodig is ten einde ook 'n teoreties-filosofiese raamwerk vir die toegepaste linguistiek daar te stel, aangesien die beskouings van die dissipline tot op hede meer beskrywend,

sinchronies en metodologies van aard is as teoreties (McNamara, 2008:303; Weideman, 2010:14). Die ontwikkeling van so 'n teorie met spesifieke verwysing na die totstandkoming van 'n sewende paradigma behoort volgens Larsen-Freeman en Weideman (2008:6, 7, 12) by te dra tot 'n beter begrip vir hoe paradigmaskuiwe die dissipline kan affekteer en beïnvloed ook wanneer die fokus, soos dit tans blyk, verskuif na die van 'n komplekse sisteem-benadering. 'n Komplekse sisteembenadering tot die toegepaste linguistiek word daardeur gekenmerk dat taal beskou word as iets wat geleidelik, maar nie op liniêre wyse nie, te voorskyn tree (Weideman, 2010:99). Hierdie generasie van die toegepaste taalkunde is egter nog nie sonder probleme nie en daar is heelwat vrae wat reeds oor die seminale werk van Larsen-Freeman en Cameron (2008) gestel kan word, soos in bostaande aanhaling gesuggereer (cf. Weideman, 2010:95-106 vir 'n meer volledige uiteensetting).

2.2.2.8 'n Opsomming van die voorafgaande

Elke generasie het dus op die een of ander wyse by die vorige aangesluit of daarop voortgebou, maar elkeen het ook 'n unieke perspektief gehad. Daar is dus oorvleuelende en onderskeidende kenmerke by elk van die generasies. Weideman (2003d:5) merk in hierdie verband op dat “[i]n spite of the fact that they may be distinct, the research agenda initiated in one generation of applied linguistic work sometimes comes to fruition only in a subsequent tradition.” Die sewe generasies van die toegepaste linguistiek, met spesifieke verwysing na hul verwantskap met taalonderrig, vul mekaar dus aan en word in Tabel 2.1 hieronder saamgevat.

Tabel 2.1: Oorvleuelende en onderskeidende generasies in die toegepaste linguistiek

Generasie	Karakteriserende eienskappe
Generasie 1 Linguisties/ behavioristies	<ul style="list-style-type: none"> • Sogenaamd wetenskaplik; streng teoreties
Generasie 2 Linguisties uitgebreide paradigma	<ul style="list-style-type: none"> • Taal as sosiale verskynsel: funksionele, samehangende en gepaste aard daarvan
Generasie 3 Multidissiplinêre model ('n verbreding van die basis wat generasies 1 en 2 gelê het)	<ul style="list-style-type: none"> • Verband met sielkunde en pedagogiek: fokus nie net op taal nie, maar ook op leerteorie
Generasie 4 Tweedetaalverwerwingstudies dien as basis	<ul style="list-style-type: none"> • Die kognitiewe vermoë om 'n taal aan te leer: eksperimentele navorsing oor hoe tale aangeleer word
Generasie 5 Konstruktivisme	<ul style="list-style-type: none"> • Interaktiewe aanleer van taal: kennis van die nuwe taal word onderhandel en interaktief gekonstrueer
Generasie 6 Postmodernisme (krities)	<ul style="list-style-type: none"> • Politiese en etiese oorwegings in onderrig en leer • Verskeidenheid van perspektiewe • Groter verantwoordbaarheid
Generasie 7 Komplekse sisteem-benadering	<ul style="list-style-type: none"> • Die aanleer van taal is nie-liniêr; die proses is een van dinamiese aanpassing

In afdeling 2.2 is die vraag gestel of die toegepaste linguistiek blote toepassing van taalteorie ('n onderafdeling van die linguistiek) is en of dit 'n dissipline in eie reg is. Die antwoord hierop is eenvoudig: die toegepaste linguistiek is deur die geskiedenis as dissipline in eie reg gevestig, aangesien dit volgens Weideman (2006:73) ontwikkel het vanaf 'n eerste generasie perspektief met 'n fokus op die empiriese na 'n derde en uiteindelik sesde generasie waar die belang van 'n meervoudigheid van perspektiewe uitgelig word, sonder om die waarde van 'n wetenskaplike onderbou te ignoreer.

Widdowson (2000:5) verduidelik die verskil tussen linguistiek en die toegepaste taalkunde as volg, wat aansluit by die verduideliking hierbo van die beweging vanaf 'n eerste generasie perspektief na 'n sesde generasie perspektief:

The differences between these two modes of intervention is that in the case of linguistics applied the assumption is that the problem can be reformulated by the direct and unilateral application of concepts and terms deriving from linguistic

enquiry itself. That is to say, language problems are amenable to linguistic solutions. In the case of applied linguistics, intervention is crucially a matter of mediation ... [it has to] relate and reconcile different representations of reality, including that of linguistics, without excluding others.

Uit die projekvoorstel van Larsen-Freeman en Weideman (2008) blyk dit egter ook duidelik dat ondersoek na die ontluikende sewende generasie nodig is om in die eerste dekade van die 21^e eeu 'n nuwe teoretiese onderbou vir die toegepaste taalkunde te vestig. Die sesde generasie het dus nog nie die finale woord gespreek nie.

Die onderskeid tussen die twee dissiplines, linguistiek en toegepaste linguistiek, word selfs duideliker verwoord deur Davies en Elder (2004:11-12) wanneer hulle sê

[w]e distinguish linguistics and applied linguistics in terms of difference of orientation. While linguistics is primarily concerned with language in itself and with language problems in so far as they provide evidence for better language description or for teaching a linguistic theory, applied linguistics is interested in language problems for what they reveal about the role of language in people's daily lives and whether intervention is either possible or desirable. What this means is that applied linguistics is as much concerned with context as with language and will therefore be likely to draw on disciplines other than linguistics, for example, anthropology, education, psychology. It also means that the language problems with which applied linguistics concerns itself are often concerned with institutions, for example the school, the work-place, the law-court, the clinic.

Verder, en voortvloeiend uit die sesde generasie, is die feit dat die toegepaste linguistiek deurlopend op soek is na geldigheid en integriteit (en miskien na die geskikte onderbou van 'n dissipline in eie reg). Weideman (2006:74) wys daarop dat

[t]o many applied linguists, the validity and integrity of their work will be undermined if the rational, scientific basis on which they purport to be working falls away or is shown to be false. In this respect the postmodern critique of applied linguistics is entirely appropriate, since such beliefs are based on a naïve view of applied linguistics as the mere application of linguistic theory to the solution of a language problem.

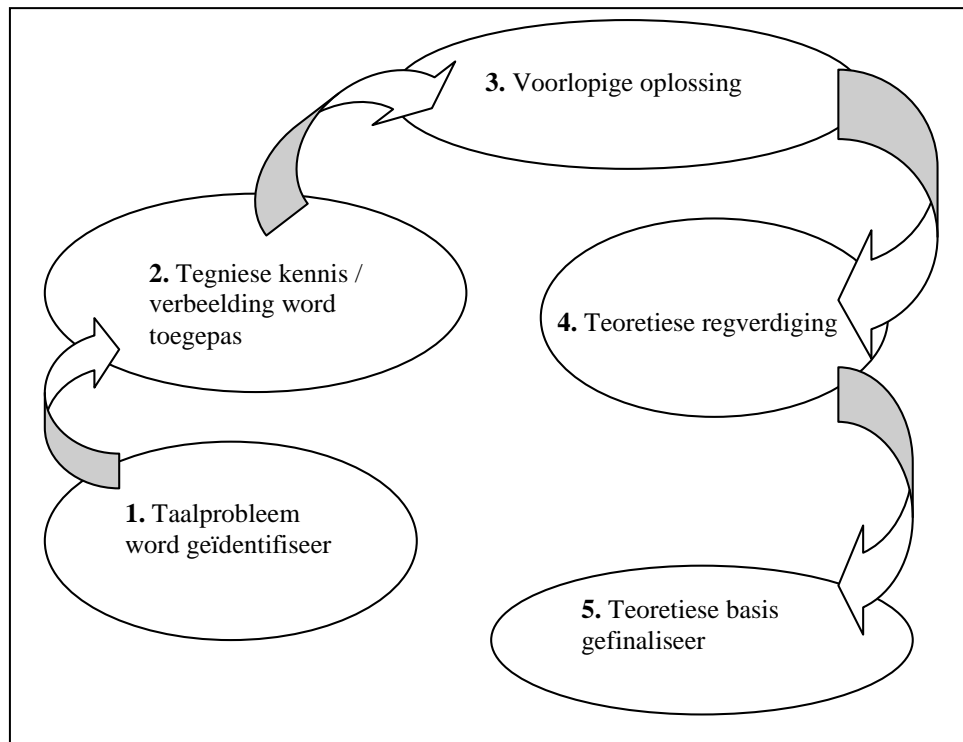
Dieselfde outeur voer verder aan dat 'n insig in die historiese ontwikkeling van die toegepaste linguistiek (soos hierbo) 'n mens in staat stel om te besef dat die ontwerp van oplossings vir taalprobleme eintlik meermale die teoretiese regverdiging van

sulke ontwerpe voorafgaan en dat die geldigheid en integriteit van die ontwerp nie soseer van die teoretiese afhanklik is of daardeur voorafgegaan word nie, maar eerder 'n rasionaal daarin vind. Hierdie rasionaal word dikwels eers agterna gevind. Dit was dan ook tipies wat, soos reeds hierbo genoem, met die fokus van hierdie studie die geval was, naamlik dat daar 'n probleem was en dat die probleem opgelos moes word. Eers nadat 'n voorlopige ontwerp vir die oplossing van die probleem gemaak is, is daar behoorlike teoretiese regverdiging daarvoor gevind. Dit blyk dus asof die toegepaste linguistiek as dissipline in eie reg finaal as 'n ontwerp-dissipline gevestig is en daar word volstaan by Weideman (2006:72) se definisie dat “rather than viewing applied linguistics as the application of linguistic theory, its historical development ... indicates that we should instead conceive of it as a technology of design ... [which] presents the solution [to the problem] in the form of a design or plan.” Die begrippe ‘technology’, ‘design’ en ‘plan’ in hierdie definisie impliseer wel die wetenskaplike of teoretiese verdedigbare aard van ontwerpe as oplossings, aangesien die leidinggewende ontwerp-funksie van 'n toegepaste linguistiese artefak, soos 'n taaltoets, in wisselwerking staan met die teoretiese basis waarop die ontwerp (uiteindelik) rasioneel begrond kan word.

Daar is dus wederkerigheid tussen tegniese ontwerp en teoretiese rasionaal. Weideman (2009a:244) voer in hierdie verband aan dat

[t]he relation between the leading, technical function ... and its founding, analytical function is reciprocal. That is, in the design of an applied linguistic instrument, the technical imagination of the designer indeed 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.

Dit kan skematies as volg voorgestel word (Weideman, 2009a:244):



Figuur 2.1: Vyf fases van toegepaste linguistiese ontwerpe

Die vyf fases van toegepaste linguistiese ontwerpe, soos in Figuur 2.1 geïllustreer, word vervolgens kortliks aan die hand van Weideman (2009a:15-16) se uiteensetting en met spesifieke verwysing na TAG en TALL verduidelik:

Fase 1 is die fase waarin daar 'n bewuswording van 'n probleem is tot op so 'n punt dat die probleem duidelik geïdentifiseer kan word. Weideman (2009a:15) voer aan dat daar tydens hierdie fase nog nie van 'n wetenskaplike benadering of omskrywing sprake is nie. In die geval van TAG en TALL was dit 'n kwessie van universiteitsowerhede wat toenemend bekommerd begin raak het oor die intreevlak-taalvaardigheid van eerstejaarstudente, soos in Hoofstuk 1 uiteengesit.

Fase 2 behels dat kundiges betrek word by die aanspreek van die probleem deur hul teoretiese kennis oor die genoemde problematiek aan te wend in samehang met hul ervaring van die ontwerp van verbeeldingryke toegepaste linguistiese oplossings. Vir TAG en TALL is taaltoetskundigheid en –belangstelling van die destydse Eenheid vir die Ontwikkeling van Taalvaardigheid (deesdae bekend as die Eenheid

vir Akademiese Geletterdheid) sowel as taalkursusontwerp ingeroep om die probleem aan te spreek. Vir die Universiteit van Pretoria (waar hierdie Eenheid gesetel is) was dit belangrik dat kundiges op laasgenoemde gebied ook betrek moes word, omdat die gepaardgaande intervensie op daardie stadium selfs belangriker as die toets geag is (Weideman, 2009a:16). Daar is met verloop van tyd egter ook kundigheid van ander universiteite (die Universiteit Stellenbosch en die Noordwes Universiteit) ingetrek deur die stig van 'n vennootskap tussen die drie instellings.

Na samesprekings tussen die drie vennote is 'n voorlopige oplossing vir die probleem, soos tipies is vir die derde fase van toegepaste linguistiese ontwerpe, aan die onderskeie universiteitsowerhede voorgelê aan die hand van 'n voorgestelde konstruk vir 'n toets van akademiese geletterdheid wat moes artikeleer met die uitkomst van 'n akademiese geletterdheidskursus vir daardie studente wat risiko in terme van akademiese taalvaardigheid toon (Weideman, 2009a:16).

Dit het aanleiding tot die vierde fase gegee, waartydens 'n teoretiese begroning of regverdiging vir die ontwerp bedink moes word – dit word in besonderhede in Van Dyk en Weideman (2004a) en in Hoofstukke 3 en 4 uiteengesit en sal nie hier verder bespreek word nie. Wat wel hier noemenswaardig is, is die feit dat daar 'n nog sterker klem geplaas is op die belyning tussen toetskonstruk en kursusuitkomst, wat op 'n teoreties regverdigbare, deursigtige en verantwoordbare wyse bewerkstellig moes word (Weideman, 2009a:16).

Vir die vyfde fase is daar teruggeval op verdere empiriese ondersoeke ten einde swakhede in die ontwerp aan te spreek. Sekere aspekte van die aanvanklike, voorlopige oplossing is derhalwe herontwerp vir 'n finale, deeglik gefundeerde oplossing. Vir TAG en TALL het dit ingehou dat daar belyning tussen konstruk en taaktipes / itemtipes, tussen konstruk en toetsspesifikasies, en uiteindelik tussen konstruk en kursusuitkomst bewerkstellig moes word wat in meer besonderhede in Van Dyk en Weideman (2004b) en Hoofstuk 4 van hierdie studie bespreek word. Weideman (2009a:16) wys egter daarop dat dit belangrik is om daarop te let dat hierdie nie die finale fase van enige ontwerp behoort te wees nie. Dit moet eerder

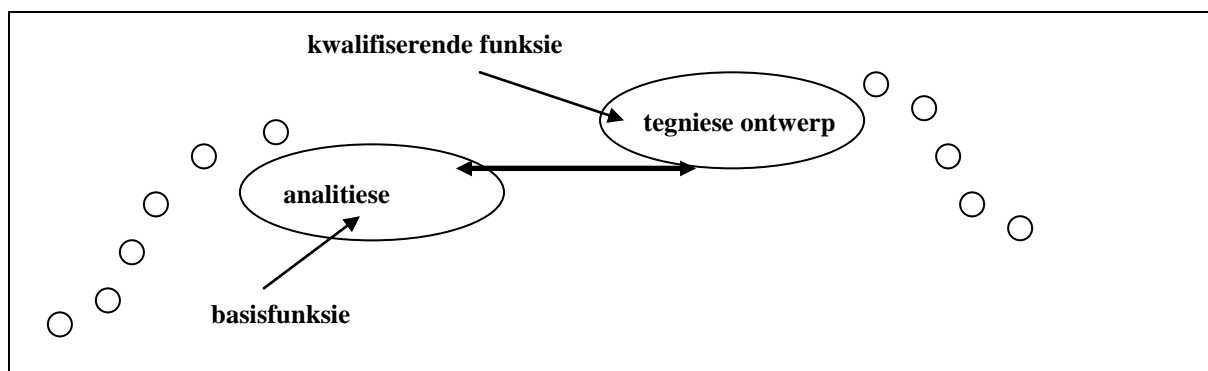
gesien word as een waar die “technical solution may again be modified in order to align it more closely with an adequate and appropriate theoretical justification, and there may be further piloting and trial runs of both the test and the intervention”.

2.3 Die toegepaste linguistiek as dissipline van ontwerp

In die voorafgaande paragrawe is aangevoer dat die toegepaste taalkunde veral met die ontwerp en implementering van oplossings vir taalprobleme te make het. As wyses of tradisies van ontwerp is dit veral die derde en sesde generasies van die toegepaste taalkunde (soos hierbo bespreek) wat aanleiding daartoe gegee het dat ’n verskeidenheid van perspektiewe nodig is om ’n oplossing vir ’n taalprobleem te ontwerp. Hierdie oplossings kan egter nie bloot ontwerp en geïmplementeer word nie sonder ’n deeglike regverdiging daarvan in terme van wetenskaplike teorie en analise, en ook nie sonder om dit tegnies verantwoordbaar en met integriteit te doen nie. Daar moet derhalwe ’n filosofiese raamwerk ontwikkel word wat, soos Weideman (2006:71) dit stel, as ’n teoretiese en eties regverdigbare begroning vir die ontwerp kan dien. Weideman verwys in die verband veral na die werk van Schuurman (2005:26) wat hierdie punt bevestig wanneer hy aanvoer dat “technology takes place in a historical, cultural, social and political context and that various groups within these contexts actively pursue various interests and goals.” Dit is hieruit duidelik dat daar verskeie rolspelers is wat belange in tegniese ontwerpe (soos taaltoetse) het en dat hierdie rolspelers en die kontekste waarin hulle funksioneer dikwels in spanning met mekaar is, veral waar hoërbelangaangeleenthede (soos toelating tot verdere studie) ter sprake kom. Dit is daarom van die uiterste belang dat daar te alle tye verantwoordbaar en met integriteit opgetree moet word wanneer oplossings vir probleme ontwerp word.

Weideman (2006:72) voer aan dat die toegepaste taalkunde oplossings vir taalprobleme bied in die vorm van tegniese ontwerpe en dat hierdie oplossings, soos enige ander tegniese ontwerp, twee funksies het, te wete ’n leidende of kwalifiserende funksie en ’n basisfunksie wat die teoretiese of empiriese fondasie vir die ontwerp bied. Hy oordeel verder dat die leidende of kwalifiserende funksie

van 'n ontwerp, as 'n toegepaste linguistiese oplossing, te vinde is in die tegniese aspek daarvan, maar dat die ontwerp telkens bestudeer of ondersoek, verklaar of verander moet word vanuit die analitiese of teoretiese perspektief wat die analitiese basis van die ontwerp bied. Let dus daarop dat hierdie interpretasie of raamwerk nie veronderstel dat die teoretiese die ontwerp lei nie, maar dat dit eerder as rasionaal daarvoor aangebied word. Figuur 2.2 is 'n skematiese voorstelling van hierdie raamwerk en is oorgeneem uit Weideman (2007a:15). Vir 'n meer volledige grafiese voorstelling word daar egter verwys na Figuur 4.2 in Hoofstuk 4.



Figuur 2.2: Leidende en basisfunksies van die toegepaste linguistiek

Dit moet ook nie uit die oog verloor word dat ontwerpe as oplossings vir probleme telkens in bepaalde kontekste voorkom en dat daar 'n aantal belanghebbendes by elke ontwerp betrokke is nie. Die etiese implikasies en gevolge van die ontwerp verdien dus ook aandag – Rambiritch (s.a.) is tans besig met 'n ondersoek vir haar doktorsale proefskrif na die etiese implikasies en die kwessie van verantwoordbaarheid in toetsing met spesifieke verwysing na TALPS (Test of Academic Literacy for Postgraduate Students) wat op basis van TAG en TALL ontwerp is. Daar word dus vir doeleindes van hierdie studie nie in besonderhede op die etiese en verantwoordbaarheidsaspekte van toetse gefokus nie, hoewel dit deurlopend sal voorkom en bespreek sal word daar waar dit van toepassing is (verwys byvoorbeeld Hoofstukke 6, 7 en veral 8). Waar Rambiritch se werk egter gebaseer is op die sogenaamde regulatiewe dimensies van die opvatting dat die toegepaste taalkunde 'n dissipline is wat gekenmerk word deur 'n tegniese ontwerpsfunksie, is hierdie studie 'n toepassing wat grootliks gefokus sal wees op

die sogenaamde konstitutiewe dimensies daarvan. Die begrippe “regulatiewe” en “konstitutiewe” word hieronder verder uiteengesit en gekontekstualiseer.

Weideman (2006:73) beweer ook dat die raamwerk in Figuur 2.2 hierbo as definisie vir die toegepaste linguistiek kan dien, aangesien dit al die kernaspekte van die vorming van die toegepaste taalkunde, soos in die uiteensetting van die drie tradisies en sewe generasies hierbo, konseptueel kan byeenbring en veral ook ruimte laat vir ’n postmoderne en opvolgende beskouings daarvan. Hy waarsku egter ook dat daar nie, soos in die vroeëre tradisies en generasies van die toegepaste taalkunde, in die huidige tydgleuf bloot weer oorgehel word na die anderkant van die spektrum nie, maar dat ’n sobere inslag tot die ontwerp van oplossings gevolg moet word. Daar moet dus ’n balans tussen die konstitutiewe (wat die basisfunksie van die ontwerp insluit) en die regulatiewe (die ontslote leidende funksie van die ontwerp) wees. In hierdie verband voer hy byvoorbeeld aan dat “[one] should carefully weigh a variety of potentially conflicting demands, and opt not only for the socially most appropriate, but also for a frugal solution” (Weideman, 2006:83). Toegepaste linguïste moet derhalwe in ag neem dat wanneer hulle besig is om oplossings vir probleme te ontwerp, hulle bereid sal moet wees om sekere toegewings en selfs kompromieë te maak: dit is byvoorbeeld soms nodig om die politieke oorwegings swaarder te laat weeg as die koste van die oplossing wat ontwerp word, en ander kere is dit weer nodig om ter wille van die bruikbaarheid of relevansie van die ontwerp tevrede te wees met ’n laer betroubaarheidswaarde, ensovoorts. Hierdie aspekte kom deurlopend in die res van hierdie studie voor en word dus nie in verdere besonderhede hier bespreek nie.

Wat egter wel van belang is vir hierdie hoofstuk, is die respektiewelik terugskouende en antisipatoriese aard van die raamwerk wat Weideman (sien hierbo) as definisie vir die toegepaste taalkunde voorhou. Hy voer in hierdie verband aan dat wanneer ’n oplossing ontwerp word dit altyd vanuit die twee genoemde perspektiewe gedoen moet word: antisipatories in die opsig dat die ontwerper byvoorbeeld die moontlike gevolge van die gebruik daarvan in ag moet

neem (as deel van die leidende of regulatiewe aard daarvan) en terugskouend in die opsig dat daar retrospektief ’n teoretiese begronding moet wees om op terug te val (as deel van die basis of konstitutiewe aard daarvan). In Tabel 2.2 hieronder is ’n verwerkte weergawe van Weideman (2006:84; 2007a:19) se raamwerk, soos in Figuur 2.1 hierbo (met die inagneming van die antisipatoriese en terugskouende element):

Tabel 2.2: Terugskouende en antisipatoriese momente van linguistiese ontwerpe

Toegepaste linguistiese ontwerp	Tipe funksie	Terugskouende / Antisipatoriese moment
word gebaseer op	konstitutiewe (essensiële / basis / fundamentele) funksies	byvoorbeeld tegniese betroubaarheid, geldigheid
word ontsluit deur	regulatiewe (leidende / kwalifiserende) dimensies	byvoorbeeld tegniese bruikbaarheid, verantwoordbaarheid van die ontwerp

In ’n latere artikel van Weideman (2009b) brei hy uit op hierdie raamwerk en gee ’n meer volledige beskrywing van welke elementêre toegepaste linguistiese grondbegrippe van belang mag wees. Soos reeds gemeld kan sulke elementêre grondbegrippe onderskei word as óf konstitutiewe óf regulatiewe begrippe of idees. Die begrippe word as’t ware gegenerer deur die verbande (analogieë, in die vorm van retrospektiewe of antisiperende refleksies / momente) tussen die tegniese dimensie en ’n veelheid ander aspekte van ons ervaring. So reflekteer die tegniese byvoorbeeld die kinematiese of betroubaarheid, en die ekonomiese die idee van tegniese bruikbaarheid of spaarsamigheid. Die belang van elk van hierdie grondbegrippe is daarin geleë dat elkeen uitdrukking gee aan ’n ontwerpbeginsel of stel tegniese beginsels vir die maak van taaltoetse. So is, om weer van die twee voorbeelde hierbo gebruik te maak, dit ’n ontwerpvoorwaarde vir ’n toets dat dit onder andere betroubaar én bruikbaar moet wees. Tabel 2.3 (uit Weideman, 2009b:66) is ’n illustrasie hiervan – hierdie illustrasie is bloot ’n uitgebreide weergawe van dit wat in Tabel 2.2. hierbo voorgelê is).

Tabel 2.3: Konstitutiewe en regulatiewe momente in toegepaste linguistiese ontwerpe

Toegepaste linguistiese ontwerp	Aspek / funksie / dimensie / ervaringsmodus	Tipe funksie	Terugskouende / antisipatoriese moment
word gebaseer op	numeriese	konstitutiewe	eenheid in 'n veelheid van ontwerpbeginsels / bronne van bewyslewering
	ruimtelike		reikwydte
	kinematiese		interne konsistensie (tegniese betroubaarheid)
	fisiese		interne effek / krag (geldigheid)
	organiese		differensiasie en aanpassing van ontwerp
	psigiese		wilskrag en verbeelding
	analitiese	funderende	rasionaal vir ontwerp
word gekwalifiseer deur	tegniese	kwalifiserende / leidende funksie (van die ontwerp)	
word ontsluit deur	linguale	regulatiewe	artikulasie van ontwerp met bloudruk / plan
	sosiale		implementering / administrasie / interaktiwiteit
	ekonomiese		tegniese nut / bruikbaarheid, matigheid / spaarsamigheid
	estetiese		harmonisering, regstel van belyningsfoute / herstel van balans
	juridiese		deursigtigheid, verdedigbaarheid, regverdigheid, legitimititeit / egtheid
	etiese		verantwoordbaarheid, sorgsaamheid, diens
	oortuigings		verbintenis, vertrou

Bostaande opsomming dui daarop dat toegepaste linguistiese ontwerpe, as oplossings vir probleme, in diens staan van daardie persone of belanghebbendes waarvoor dit ontwerp word en altyd vanuit daardie perspektief benader moet word, maar dat die onderskeie ontwerpbeginsels altyd in wisselwerking met mekaar beskou moet word. So kan die tegniese bruikbaarheid (die ekonomiese analogie) van 'n toets nie behoorlik gepeil word sonder om die implementeerbaarheid (die sosiale eggo) in ag te neem nie. Insgelyks is 'n begrip soos die van teoretiese

regverdigbaarheid (logiese analogie) eintlik ondenkbaar sonder konsepte soos konstantheid (die kinematiese) en geldigheid (fisies-energetiese refleksie). Al hierdie ontwerpbeginsels, wat sistematies onthul en blootgelê word in 'n analise van 'n toegepaste linguistiese instrument (in hierdie geval 'n taaltoets), werk saam en dra by tot die verantwoordelike ontwerp en maak van toetse.

2.4 Samevatting

Hierdie hoofstuk het die teoretiese agtergrond waarteen die res van die studie gelees behoort te word, uiteengesit. Die toegepaste taalkunde as dissipline van ontwerp is gedefinieer en daar is 'n gebalanseerde raamwerk voorgelê waarteen linguistiese artefakte ontwerp behoort te word. Daar sal telkens in die res van hierdie studie teruggeval word op die basis wat hierdie raamwerk vir die ontwerp van oplossings inhou. In hierdie sin is 'n belangrike doelstelling van hierdie proefskrif die poging om die spesifieke perspektief op die toegepaste taalkunde wat in hierdie hoofstuk uiteengesit is, verder uit te bou en te beproef. Is hierdie raamwerk byvoorbeeld robuus genoeg om die vraagstellings in Hoofstuk 1 op 'n sinvolle manier te kan konseptualiseer en te beantwoord? Hoe verbind dit tegniese-konstitutiewe begrippe soos betroubaarheid en geldigheid aan regulatiewe idees, soos verantwoordbaarheid en sorgsaamheid? In die argumente wat volg, sal daar voortdurend verwys word na die potensieel-verhelderende impak wat die gebruik van so 'n filosofiese raamwerk inhou, wat uitgaan van die toegepaste taalkunde as 'n dissipline wat gekenmerk word as een wat te doen het met die ontwerp, en uiteindelik die wetenskaplike regverdiging van oplossings vir taalprobleme.

Die volgende twee hoofstukke begin om die storie van so 'n ontwerp te vertel, naamlik die van TAG en TALL. In Hoofstuk 3 word die ontwerp van 'n konstruk vir TAG en TALL spesifiek bespreek. Hoofstuk 4 bou daarop voort deur aan te toon hoe die genoemde konstruk uiteindelik gebruik is as teoretiese regverdiging vir die samestelling en operasionalisering daarvan.

Die ontwerp van 'n nuwe konstruk: op soek na 'n toepaslike onderbou vir die toetsing van akademiese geletterdheid

3.1 Inleidend

Wanneer daar beslissings oor persone of hul vermoëns gemaak moet word, is daar altyd in die een of ander opsig sprake van onsekerheid. Daar is byvoorbeeld, soos in Hoofstuk 1 aangetoon, geen sekerheid dat 'n persoon wat vir universiteitstudie inskryf, sy/haar studies met sukses sal afhandel nie. Universiteite behoort dus soveel as moontlik meganismes in plek te stel om met redelike sekerheid beslissings oor voornemende en reeds toegelate studente te neem ten einde hul op gepaste wyse te ondersteun, indien nodig. Sluijter (1998:1) wys daarop dat daar nie opvoedkundig verantwoordbare en ingeligte besluite oor studente geneem kan word sonder addisionele hulpmiddele nie. Vir doeleindes van hierdie studie en gegewe die onderskeidings wat in die vorige hoofstuk gemaak is, word “addisionele hulpmiddele” geïnterpreteer as tegniese artefakte, en in hierdie geval spesifiek die soort artefakte in die toegepaste taalkunde wat ons ken as toetse van taalvermoë, instrumente of hulpmiddele om akademiese geletterdheidsvlakke mee te bepaal. Sluijter (1998:1) meen egter ook dat enige besluit per definisie in 'n meerdere of 'n mindere mate onseker is en dat die genoemde hulpmiddele die kans op korrekte

besluitnemingsprosesse slegs tot 'n mate verhoog. Ten einde, in soverre dit moontlik is, op verantwoordelike en verantwoordbare wyse beslissings oor persone of hul vermoëns te maak (dus met inagneming van onsekerhede) en om uiteindelik besluite oor hul toekoms te neem wat in die beste belang is van die besluitnemers (vir doeleindes van hierdie studie: universiteite), sowel as die persoon op wie die besluit betrekking het (vir doeleindes van hierdie studie: studente), kan daar van addisionele hulpmiddele soos toetse gebruik gemaak word. Die keuse van gepaste toetse deur universiteite vir onder andere keuring, toelating, plasing, diagnostiese doeleindes, ensovoorts is amper vanselfsprekend 'n saak wat met die grootste omsigtigheid hanteer moet word, aangesien die besluite wat op grond van die resultate daarvan geneem word, aanvegbaar kan wees, veral waar dit gaan om hoërbelang aangeleenthede.

In Hoofstuk 1 is melding gemaak van so 'n aanvegbare toets, die ELSA-Plus. Die elemente daarvan (geluidsleer, diktee, basiese gesyferdheid, leesbegrip, taal en grammatika van ruimtelike verwantskappe, sinsluiting, en woordeskat binne konteks) is uiteengesit in paragraaf 1.4.1 daarvan. Hierdie toets is wel vir laebelang doeleindes (slegs plasing) aangewend, maar die geskiktheid daarvan het konstant onder kritiek gekom, aangesien dit (i) logistiese en administratiewe beperkinge gehad het, (ii) dit nie koste-effektief was nie, (iii) dit taalvaardigheid in die tradisionele sin van die woord gemeet het en nie die vermoë van die student om in die akademiese omgewing te interakteer en te kommunikeer nie, en (iv) dit op 'n beperkte en verouderde siening van taalvermoë berus het. Dit was veral die laaste twee van hierdie vier redes, wat betrekking het op die geldigheid van die konstruk van die toets, wat deurslaggewend was in die besluit om dit te vervang met 'n toets wat 'n teoreties veel meer verdedigbare konstruk het. Om dit in die terme te stel van die onderskeidings wat in die vorige hoofstuk gemaak is: ELSA Plus se konstruk het vasgesteek in die paradigma van die eerste generasie van toegepaste taalkundige werk. Dit was teoreties regverdigbaar slegs in terme van 'n beperkte siening van taal.

Aangesien daar sedert die draai van die eeu oor die algemeen 'n groter bewustheid, veral op bestuurs- en beleidsvlak, in Suid-Afrika oor die waarde van (akademiese) taal en (akademiese) taalondersteuning begin posvat het, kon 'n toets wat onder verdenking was, vir die redes hierbo genoem, dus nie meer gebruik word nie, veral omdat die behoefte aan taaltoetse wat die hoërbelangfunksies moet vervul, byvoorbeeld as deel van toelating, pertinent deur verskeie rolspelers in die hoëronderwyssektor uitgespreek is. Om hierdie rede moes 'n toets ontwerp word wat maklik administreerbaar is (gegewe die potensiële getalle wat dit moes aflê), wat goedkoop is en waarvan die koste beheerbaar is, en wat hedendaagse tendense en teorieë oor taal en toetsing weerspieël, sodat dit aan die eise van beide teoretiese regverdigbaarheid (konstruktiewe geldigheid) en van sosiale verantwoordbaarheid voldoen.

In 'n post-1994 Suid-Afrika is dit belangrik om uiters verantwoordbaar te werk te gaan met oplossings (ontwerpe) vir probleme. McNamara (2004:763,773) voer juis aan dat “[m]ost recently, language tests are under somewhat of a challenge, as they respond to critiques of individualistic notions of performance and are increasingly being scrutinized for their social accountability ... [since they] control access to opportunities within societies, and are used to control the flow of people between societies.” Die verantwoordbaarheidsaspek is nie deel van die bestek van hierdie studie nie, maar Rambiritch (s.a.) is, soos reeds vermeld, tans besig met 'n ondersoek na hierdie aspek van taaltoetsing vir haar doktorsale studie, getiteld *Accessibility, transparency and accountability as regulative conditions for a postgraduate test of academic literacy*, waarin sy na verwagting die verantwoordbaarheidskwessie ook in TAG en TALL na behore sal analiseer en toelig. Rambiritch se studie beoog dus om 'n verdere komponent uit te bou van die raamwerk vir die toegepaste linguïstiek wat in Hoofstuk 2 geartikuleer is.

Die teoretiese regverdigbaarheid van die ontwerp van 'n taaltoets vereis dat daar eerstens 'n gefokusde literatuurstudie onderneem word om die relevante benaderings tot en standpunte oor toetsing te konsolideer ten einde daaruit tot

gepaste gevolgtrekkings te kan kom vir toepassing op die konstruksie vir 'n toets van akademiese geletterdheid wat vir besluitneming op laer (plasing) én hoër vlakke (toelating) gebruik kan word.

In aansluiting by die onderskeidings wat in Hoofstuk 2 gemaak is, wys McNamara (2004:763) ook daarop dat “language testing represents one of the core areas of applied linguistics” en daarom behoort hierdie hoofstuk binne die konteks van die toegepaste linguïstiek as dissipline van ontwerp gelees te word, waar 'n toetskonstruk en uiteindelik 'n toets ontwerp word om 'n bepaalde probleem aan te spreek.

Die genoemde oorsig van die literatuur wat volg, begin met 'n bespreking van die begrippe *toets* en *toetsing* en hoe dit met konstruksie verband hou, waarna daar 'n oorsig van die historiese benaderings tot taaltoetsing wat vir doeleindes van hierdie studie van belang is, gebied sal word. Daar sal ook deurlopend verwys word na hoe die historiese benaderings waarna in die literatuur verwys word, saamhang met die verskillende paradigmas van toegepaste linguïstiese werk wat in breë trekke in Hoofstuk 2 uitgestippel is, en hoe nou die geskiedenis van dissipline (toegepaste taalkunde) en sub-dissipline (taaltoetsing) saamhang. Ook word in wat volg die sogenaamde diskrete-item toetse ('discrete point testing') en geïntegreerde toetse ('integrative tests') as tipes toetse bespreek. Dit word opgevolg met 'n argument vir die ontwikkeling van 'n nuwe toetskonstruk vir die meting van akademiese geletterdheid.

3.2 Toets en toetsing gedefinieer

Brown (2004:3) definieer 'n toets as “a method of measuring a person's ability, knowledge, or performance in a given domain”. In sy verduideliking van hierdie definisie wys hy daarop dat 'n toets 'n samevoeging van tegnieke, items en prosedures is wat as integrale deel van 'n enkele instrument op 'n bepaalde tydstip in bepaalde omstandighede uitgevoer moet word ('method'). Dit word uitgevoer om vaardighede en / of kennis en / of vermoëns ('competence') op 'n spesifieke

gebied ('domain') te meet en dit in die een of ander vorm uit te druk ('measure') met die inagneming van die kenmerke van die spesifieke populasie wat die toets afgelê het ('person = individual'). McNamara (2004:765) definieer taaltoetsing in amper soortgelyke terme as

a process of gathering information about test-takers from observed performance under test conditions. This is done in order to draw inferences either about the likely quality of performance by the test-taker under non-test conditions, or about the test-taker's standing in relation to a relevant domain of knowledge and abilities.

Sluijter (1998:1-2) se definisie eggo ook die konsepte van Brown en McNamara hierbo, maar hy tref 'n definitiewe onderskeid tussen toets as naamwoord en toets as werkwoord, wat die interpretasie daarvan vergemaklik. Laasgenoemde is volgens hom "een gestandaardiseerde en systematiese meetprocedure die het mogelijk maakt kwantitatiewe uitspraken te doen over eigenschappen van personen, met het doel om beslissingen te nemen" (Sluijter, 1998:1). Hy voer verder aan dat die begrip toets (naamwoord) 'n verbesondering van die begrip toets (werkwoord) is:

De term toets is gereserveerd voor instrumenten die door onderwijs en studie verworven kennis, inzicht of vaardigheden meten. Wordt de term test in niet-generieke zin gebruikt, dan verwijst deze naar instrumenten die eigenschappen van personen meten die niet door intentioneel onderwijs en studie verworven zijn.

Sluijter (1998:2)

Vir doeleindes van hierdie studie is dit egter van belang om nie net oor toets (naamwoord) en toets (werkwoord) oor die algemeen te besin nie, maar in die besonder te verwys na die toets van taal vir spesifieke doeleindes (TSD), in aansluiting by Brown se definisie hierbo waarin hy onder andere na vaardigheid/kennis in 'n spesifieke domein verwys.

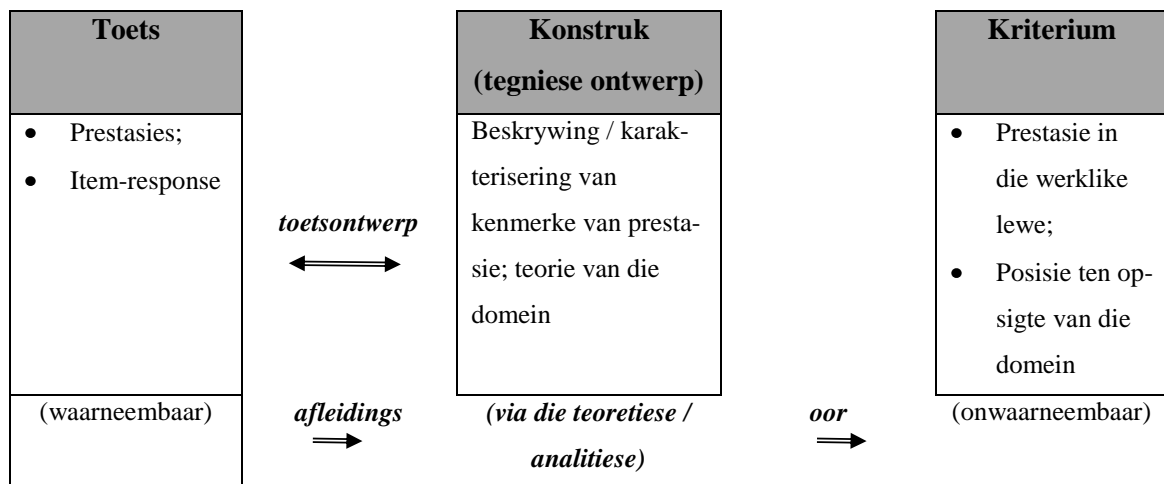
Die toets van taal vir spesifieke doeleindes verwys volgens Douglas (2000:1) na daardie aspek van taaltoetsing waar daar slegs op toetsinhoud en -metodes besluit kan word nadat daar 'n deeglike analise van die spesifieke taalgebruikssituasie (die domein) gedoen is waarin die persone wat die toets moet aflê moet funksioneer. Sy definisie vir die toets van TSD lui:

A specific purpose language test is one in which test content and methods are derived from an analysis of a specific purpose target language use situation, so that test tasks and content are authentically representative of tasks in the target situation, allowing for an interaction between the test taker's language ability and specific purpose content knowledge, on the one hand, and the test tasks on the other. Such a test allows us to make inferences about a test taker's capacity to use language in the specific purpose domain.

Douglas (2001:19)

Die waarskynlike rede vir sy definisie is om te beklemtoon dat daar betroubare, geldige en outentieke toetsinhoud en -taaktipes uit die genoemde analise behoort voort te vloei. Brown (2004:19ff) wys juis daarop dat outentisiteit, geldigheid en betroubaarheid sowel as uitvoerbaarheid as die basiese ontwerpbeginsels van 'n goeie toets beskou behoort te word, wat reeds ook in Hoofstuk 1 (1.6) gesuggereer is. Indien Weideman (2003d, 2006, 2007a, 2009a, 2009b) se siening dat 'n toets 'n toegepaste linguistiese artefak of instrument is (verwys ook Hoofstuk 2), verder korrek is, volg dit vanselfsprekend dat 'n toets aan spesifiseerbare ontwerpsbeginsels of attribute soos hierdie sal moet voldoen. Hierdie beginsels kom telkens in hierdie studie aan bod, maar vir die onmiddellike lyk dit egter asof McNamara (2004:765) se grafiese voorstelling (aangepas in Figuur 3.1) van toetse as tegniese ontwerpe waaruit afleidings oor kandidate en hul (taal)vermoëns gemaak kan word, tot 'n mate 'n bevestiging is van Weideman se siening dat 'n toets attribute kan hê, gesien vanuit die perspektief van die toegepaste linguistiek as 'n dissipline van tegniese ontwerp. In hierdie opsig toon die genoemde figuur duidelik aan dat die interpretasie van afleidings via die teoretiese of analitiese begronding daarvan gedoen moet word, ten einde geldige en verantwoordbare uitsprake te kan maak oor kandidate en hul (taal)vermoëns in ander, outentieke kontekste, of soos McNamara (2004:765) dit stel:

We thus make a distinction between the *test* (the means of drawing inferences) and the *criterion* (the target of test inferences) ...[:] the criterion cannot be 'known' in any direct sense, our inferences are necessarily mediated through test *constructs*, that is, our modelling of the criterion in terms of its essential features or characteristics.



Figuur 3.1: 'n Toepassing van die model van die toegepaste linguistiek as dissipline van tegniese ontwerp op taaltoetsing (aangepas uit McNamara, 2004:765)

So 'n model vind spesifiek neerslag in die situasie waarmee hierdie studie onder andere mee gekonfronteer is (verwys Hoofstuk 1), naamlik dat TAG tans as deel van 'n medium- tot hoëbelangtoetsbattery gebruik word om toegang tot hoëronderwys te reguleer. Dit word dus, tesame met ander toetse in die battery, gebruik om studente se vermoë om in die akademiese omgewing te werk en te oorleef, te bepaal. Om dit op 'n betroubare en geldige wyse te kan doen, moet die eise van die akademiese omgewing derhalwe gemodelleer word, wat 'n deeglike ondersoek vereis na presies wat van studente in die akademiese omgewing (taalmatig) verwag word sodat daar 'n volledige prentjie geskets kan word van hierdie domein. Hierdie hoofstuk spreek juis hierdie deel van die prentjie aan, naamlik die ontwerp van 'n konstruk vir 'n toets van akademiese geletterdheid. Die kwessie is egter veel groter en ingewikkelder as net die teoretiese ontwerp van 'n konstruk; konstrunkte (en veral toetskonstrukte) is nooit onkontroversieel nie.

Daar is inderdaad ook 'n waarskuwing onderliggend aan hierdie model, wat weereens deur Weideman se interpretasie van die beoefening van die toegepaste linguistiek as 'n dissipline van ontwerp gereflekteer word. In die openbare belang moet daar ongetwyfeld op deursigtige wyse met toetskonstrukte te werk gegaan word deur dit deurlopend te ondersoek, te bevraagteken, en te artikuleer - alles as

deel van die proses wat volgens die heersende ortodoksie validering (wat later in hierdie studie ter sprake sal kom) genoem word. So 'n proses is nodig omdat daar gewoonlik op toetsuitslae alleen gefokus word, in byvoorbeeld statistiese analises, en uitsprake daaroor en interpretasies daarvan foutiewelik voorkom en selfs misbruik kan word.

In 'n post-apartheid era in Suid-Afrika spesifiek, en 'n kritiese, post-modernistiese ingesteldheid op globale vlak, is dit uiters noodsaaklik dat toetsuitslae met die nodige omsigtigheid en nuanse geïnterpreteer en beskryf behoort te word, ten einde werklik verantwoordbaar op te tree. McNamara (2004:769) stel dit juis so dat “[t]he need to defend our interpretations of the meanings of test scores is because of the necessarily indirect relationship between the test and what we ultimately want to know about the candidate’s standing in relation to the criterion.” Hy voer verder aan dat “[i]nformation from test scores however is only one part of the picture as it cannot offer explanations for the outcomes so reported; it would be unwise to base decisions on the outcome measures alone; ... [w]e need to consider the reasons for the results” (McNamara, 2004:777). Hierdie ‘reasons’ lê onder andere opgesluit in die konstruk en 'n ontsluiting daarvan is derhalwe nodig. Voordat die storie van die ontwerp van 'n konstruk vir akademiese geletterdheid (soos gemanifesteer in TAG) vertel (geartikuleer/beskryf) kan word, is dit belangrik om te begin by 'n historiese oorsig oor benaderings tot taaltoetsing, aangesien dit die konteks vorm waarbinne konstrunkte ontwikkel word.

3.3 Benaderings tot taaltoetsing

Die afgelope vyftig jaar het verskeie benaderings tot taaltoetsing die lig gesien, wat – alhoewel dit soms nie so aangemerkt word nie - veral aangevoer is deur benaderings tot taalonderrig (Davies, 1968; McNamara, 2004:763). Telkens wanneer daar 'n nuwe benadering tot taalonderrig op die voorgrond gekom het, het voorstanders daarvan alle voorafgaande benaderings krities beoordeel en veral in terme van die geldigheid daarvan afwysende kommentaar gelewer. Die gevolg hiervan was dat instellings wat van taaltoetse gebruik gemaak het soms van die een

uiterste na die ander beweeg het in terme van hul toetsstrategieë, wat taaltoetsing beide kontroversieel en kompleks gemaak het. Daar word in die literatuur hoofsaaklik van drie onderskeibare, dog oorvleuelende, benaderings tot taaltoetsing gewag gemaak: die voorwetenskaplike / literêr-grammatiese benadering, die psigolinguistiese benadering, en die sosiolinguistiese / pragmatiese oftewel kommunikatiewe benadering tot taaltoetsing (Brown, 2004; Carrol, 1980; Madsen, 1983; McNamara, 2004; Oller, 1979; Rivers, 1968; Spolsky, 1978). Hoe nou taaltoetsing verbind is aan die verskillende paradigmas van toegepaste linguistiese werk wat in die vorige hoofstuk bespreek is, sal opval in die bespreking daarvan wat volg. Dit bied 'n verdere bevestiging van McNamara en Roever (2006:1) se beskouing dat taaltoetsing 'n onderdeel is van die toegepaste linguistiek, wat ook een van die grondliggende aannames van hierdie studie is.

3.3.1 Die voorwetenskaplike/literêr-grammatiese benadering

Hierdie benadering vloei voort uit die sogenaamde 'grammar-translation'-metode tot taalonderrig. In hierdie benadering is taal volgens Roets (1990:31) as voertuig vir literêre en kulturele tradisies gesien waarin 'n deeglike studie van onder meer die kultuurgeskiedenis, die letterkunde en die grammatika van 'n taal gedoen is. Latere vorme daarvan steun op strukturaliste soos De Saussure, Bloomfield en Jespersen se formele siening van taal (Rivers, 1968:14ff). Studente se taalvermoë is getoets deur byvoorbeeld tekste of sinne, wat spesiaal gekonstrueer is om bepaalde grammatikale aspekte uit te lig, na of uit die doeltaal te vertaal, of om opstelle te skryf wat subjektief geëvalueer is. Die statistiese kwaliteite van toetsitems is geïgnoreer en die akkuraatheid en regverdigbaarheid van resultate kan derhalwe ernstig bevraagteken word.

3.3.2 Die psigolinguistiese benadering

In 'n teenreaksie op die voorafgaande wat as voorwetenskaplik en in sommige gevalle selfs as onwetenskaplik beskou is, het die sogenaamde 'audio-lingual'-metode tot taalonderrig die lig gesien (Roets, 1990:32). Hierdie metode het sy grondslag in die behavioristiese werk van Skinner en Thorndike. Lado

(1961a:211ff.), wat een van die voorlopers op die gebied van taaltoetsing tydens hierdie era was, het as premis gehad dat taal 'n sisteem van gewoontes is wat op drie vlakke geanaliseer kan word, naamlik die fonologiese, die sintaktiese en die semantiese. Soos wat daar reeds in Hoofstuk 1 opgemerk is, is hierdie 'n atomistiese benadering tot taalonderrig en niks meer as 'n beperkte blik op taal nie. Die samewerking van gedragsielkundiges met linguïste het meegebring dat presiese en sogenaamde objektiewe taaltoetse met redelik aanvaarbare statistiese attribute tot stand gekom het. Hierdie toetse het volgens Roets (1990:32) bestaan uit “stringe atomiese onderdeeljies wat ondubbelsinnig reg of verkeerd gemerk kon word.” Dit het die wêreld van toetsing omgekeer en het later bekend gestaan as diskrete-item toetse (wat hieronder in meer besonderhede bespreek sal word). Volgens Farhady (1979:348) is hierdie benadering so hoog aangeskryf dat dit in sommige lande as die enigste betroubare en wetenskaplike benadering tot die toetsing van tweede- of vreemdetaalleerders gesien is.

Daar is egter ook wyd kritiek (Carrol, 1961; Jakobovits, 1970; Oller, 1975, Spolsky, 1973) teen hierdie benadering uitgespreek wat waarskynlik die beste in die volgende woorde van Oller (1979:212) saamgevat kan word:

Any discourse in any natural language is more than the mere sum of its analysable parts. Crucial properties of the language are lost when it is broken down into discrete phonemic contrasts, words, structures and the like. Nonsense recommended by some experts as the basis for discrete point test items, does not exhibit many of the pragmatic properties of normal sensible utterances in discourse contexts.

Dit is duidelik dat, in toegepaste taalkundige terme, hierdie benadering tuis behoort by wat in die vorige hoofstuk 'n “eerste generasie” aanpak genoem is. Hierdie benadering het egter 'n enorme invloed op die regverdiging van taaltoetsontwerp gehad, maar dit het mank gegaan aan 'n verantwoordelike siening van die rol van die ‘wetenskap’. Dit kom desondanks neer op 'n modernistiese opvatting van wat ‘wetenskaplik’ vermag kan word.

3.3.2.1 Diskrete-item toetse

Die groot invloed van eerste generasie-werk in taaltoetsing word veral duidelik as 'n mens kyk na die debat oor diskrete-item toetse. Die aanname dat taal opgebreek kan word in komponente of dele en dat die dele suksesvol afsonderlik gemeet kan word, is die vertrekpunt vir die ontwikkeling van diskrete-item toetse. Hierdie komponente is gewoonlik die vier sogenaamde taalvaardighede, te wete luister, praat, lees en skryf wat as deeltjies van taal (die fonologiese, die morfologiese, die leksikale, die sintaktiese en die semantiese) en verder in die vorm van diskrete of geïsoleerde items in toetse voorkom. Daar is dus aangevoer dat 'n taalvaardigheidstoets al vier die vaardighede moes meet, gewoonlik in terme van 'n groot aantal diskrete linguistiese deeltjies (Brown, 2004:8). Dit is klaarblyklik ook die teoretiese regverdiging onderliggend aan die ontwerp van die ELSA Plus-toets, waarna vroeër verwys is.

Die grootste nadeel van diskrete-item toetse is volgens Oller (1973:189) dat “they fail (in most cases) to faithfully reflect actual language usage [in that] they require the student to perform highly artificial tasks and on the basis of ...performance attempt to infer ... level of competence for a different sort of task all together.” Dieselfde outeur (1973:189) lys ook enkele ander nadele, naamlik dat dit ten tweede nie gemaklik in taalprogramme pas nie, aangesien dit studente nie die geleentheid bied om taal in te oefen nie. 'n Derde nadeel wat hy lys, is met verwysing na meervoudige keuse items, wat die tipiese voorbeeld van 'n diskrete-item in 'n toets is. Dit vereis 'n besondere vlak van vaardigheid om goeie meervoudige keuse items te skryf. Dit verg byvoorbeeld heelwat tyd en inspanning in die opsig dat dit deeglik geproeflees moet word (veral die afleiers), eers in loodstoetse moes voorgekom het, volgens die statistiese analises moontlik herskryf moet word, en weer uitgetoets moet word voordat dit gereed is om in 'n finale toets opgeneem te kan word.

Oller (1973:190) lys egter ook voordele van diskrete-item toetse: hy dui eerstens aan dat diskrete-item toetse 'n diagnostiese waarde kan vervul indien die aanname

is dat die teikentaal uit afsonderlike dele bestaan en dat hierdie afsonderlike dele getoets moet word om te bepaal of 'n student byvoorbeeld gebrekkige leesvaardigheid het. 'n Verdere voordeel is dat dit maklik administreerbaar is en net so maklik en vinnig, objektief en betroubaar, nagesien kan word, veral in gevalle waar groot getalle studente binne 'n kort tydjie getoets moet word – die tyd en inspanning wat in die skep van meervoudige keuse items gaan, blyk dus geregverdig te wees.

Alhoewel daar voordele aan diskrete-item toetse verbonde is, is daar in 'n era waar die waarde van kommunikasie en onderhandelingsstrategieë, outentisiteit en konteks hoogty begin vier het (die laat sewentiger- en tagtigerjare) egter na ander, minder-atomistiese, benaderings tot itemontwikkeling gesoek. Dit word in meer besonderhede in punt 3.3.3.1 hieronder bespreek.

3.3.3 Die sosiolinguistiese/pragmatiese, oftewel kommunikatiewe benadering

Carrol (1961) se kritiek reeds in die vroeë sestigerjare teen die psigolinguistiese benadering tot taaltoetsing het die begin aangekondig van die derde benadering tot taaltoetsing, naamlik die sosiolinguistiese / pragmatiese benadering. Dit sluit duidelik aan by die werk wat in die tweede generasie van die toegepaste taalkunde gedoen is, met 'n verruimde perspektief op wat taal is (2.2.2.2 hierbo). Dit het 'n fokusverskuiwing voorgehou, weg van diskrete linguistiese komponente na 'n meer sosiale ingesteldheid waar die uitwerking van linguistiese aktiwiteite op kommunikasie bepaal kon word, en het in die laat sewentigerjare en veral in die vroeë tagtigs hoogty gevier. Roets (1990:34) omskryf hierdie benadering as een wat "...'n mens in staat [stel] om die linguistiese, strategiese en pragmatiese veranderlikes wat relevant is vir kommunikasie in die werklike lewe, sistematies op verskeie vlakke te oorweeg."

Die kommunikatiewe teorieë wat in die sewentiger- en veral die tagtigerjare die basis van taalonderrig gevorm het, en wat tans steeds invloedryk is, het hierdie derde benadering tot taaltoetsing finaal gevestig. Die seminale werk oor

kommunikatiewe bekwaamheid in die tweedetaal deur Canale en Swain (1980), wat voortgebou het op die werk van Hymes (1972), is uitstekende voorbeelde hiervan. Dit het egter eers enkele jare later meer genuanseerd tot sy reg gekom (sien punt 3.3.3.2 hieronder). 'n Meer geïntegreerde benadering tot taaltoetsing het derhalwe gevolg waarin daar aangevoer is dat "...the whole of the communicative event was considerably greater than the sum of its linguistic elements" (Clark, 1983:432) wat in lyn is met Carrol se kritiek teen die psigolinguistiese benadering meer as twintig jaar vantevore. Die sosiolinguistiese / pragmatiese oftewel kommunikatiewe benadering staan algemeen ook bekend as 'n meer geïntegreerde benadering tot taaltoetsing.

3.3.3.1 Geïntegreerde toetse (*integratiewe toetse*)

Die oorgang na die kommunikatiewe benadering tot taaltoetsing word ingelei deur die soeke na 'n soort toets wat taal integratief kon toets. Soos in punt 3.3.2.1 hierbo aangetoon is, was daar ernstige kritiek teen die sogenaamde diskrete-item toetse wat selfs vandag nog in debatte aangevoer word wanneer daar oor die waarde van diskrete- of geïntegreerde-item toetse diskussie gevoer word. Uitgaande van die sogenaamde 'unitary trait hypothesis' het Oller (1979) vroeg al deurlopend in sy boek oor taaltoetsing geargumenteer dat taal 'n komplekse en geïntegreerde sisteem is wat nie bloot net in dele soos woordeskat, grammatika, lees, skryf, en ander diskrete elemente afgebreek en geïnterpreteer kan word nie. Hy (1979:37) definieer geïntegreerde toetse derhalwe as volg:

If discrete items take language skill apart, integrative tests put it back together ... [in an] attempt to assess a learner's capacity to use many bits all at the same time, and possibly while exercising several presumed components of a grammatical system, and perhaps more than one of the traditionally recognized skills or aspects of skills.

Geïntegreerde-item toetse, in teenstelling met diskrete-item toetse, poog dus om die totale taalvermoë van 'n student te bepaal deurdat dit die vaardighede betrokke by normale kommunikasie wil meet. Let daarop dat daar na taalvermoë as taalvaardigheid verwys word. Taalvaardigheid het egter in hierdie konteks niks te make met die vier taalvaardighede, naamlik luister, lees, praat en skryf nie. Hier gaan dit eerder oor taalvermoë in die algemeen. Oller (1979:198) wys daarop dat

[i]n spite of the fact that some of the integrative skills tests seem to have little in common, and regardless of the fact that they may seem to be unreliable as far as scoring is concerned, repeated studies show that scores on tests of integrative skills tend to correlate better with teacher judgements, better among themselves, and better with other measures of language skills than do any of the discrete-point types because they more neatly reflect what people actually do when they use language.

Dieselfde outeur wys egter alreeds 'n paar jaar voor bogenoemde uitspraak daarop dat die twee tipes toetse, hoewel teoreties verskillend, eintlik op 'n kontinuum hoort met diskrete-items aan die een kant en geïntegreerde items aan die ander kant (Oller, 1973). Deur 'n bepaalde toets as byvoorbeeld 'n grammatika- of 'n woordeskattoets te etiketteer, beteken nie noodwendig dat dit 'n diskrete-item toets *per se* is nie. Grammatika kan in elk geval net so maklik deur 'n geïntegreerde-item toets in meervoudige keuse formaat gemeet word. Die verskil tussen diskrete-item toetse en geïntegreerde-item toetse blyk dus eerder 'n verskil in graad te wees, en tipes toetse waaraan daar voor- en nadele aan verbonde is.

Ingram (1978:12) sit hierdie ietwat grys area heel verstaanbaar uiteen wanneer sy sê dat

[i]t is in any case, quite unnecessary to suppose that one has to make an either/or choice, that if one approves of integrative tests, one should therefore disapprove of discrete-point ones. This 'disjunctive fallacy' ... stems, it seems to me, from misunderstanding about the nature of language command.

Hierdie stelling is 'n presiese aanduiding van waarom daar miskien eerder 'n totale wending moet wees na 'n kommunikatiewe benadering tot taaltoetsing, wat vervolgens bespreek sal word.

3.3.3.2 Die kommunikatiewe benadering tot taaltoetsing

In die middel tagtigerjare van die vorige eeu is die behoefte aan 'n meer genuanseerde of 'n verruimde beskrywing van persone se taalvermoë uitgespreek en is daar finaal wegbeweeg van die sogenaamde 'unitary trait hypothesis' binne konteks van taaltoetsing wat ingehou het dat taalvermoë nie opbreekbaar of onderskeibaar is nie en dat "... vocabulary, grammar, phonology, the 'four skills,' and other discrete points of language could not be disentangled from each other" (Brown, 2004:9). Die rede hiervoor was, wat reeds in 1979 as volg aangegee is: "...

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” (Farhady, 1979:349). Hoewel kommunikatiewe vermoë steeds as een, enkele vermoë geag is, is daar vir doeleindes van toetsing hiervan wegbeweeg.

Die fokus het dus 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 (Roets, 1990:35). Dit het meegebring dat toetsontwerpers daarmee geworstel het om toetsitems te ontwikkel wat dus nie meer net een aspek van taal meet nie (soos wat ’n diskrete-item toets veronderstel was om te doen nie) of taal net as ’n enkele sisteem te sien nie (soos wat integratiewe toetse gedoen het nie). Daar het ’n behoefte aan ’n verruimde of sosiolinguisties-regverdigbare benadering tot taaltoetsing ontstaan. In hierdie verband is die vraag wat na vore gekom het: wat word getoets - taalkennis, taalvaardigheid, taalgebruik, of taalaanleg? En verder nog: behoort die vier taalvaardighede afsonderlik, of saam, of geïntegreerd getoets te word? Selfs meer kontroversieel geformuleer: bestaan daar in werklikheid vier afsonderlike “vaardighede” (Kumaravadivelu, 2003: 29)? Die oorkoepelende konsep waarmee vorendag gekom is om hierdie vrae te help beantwoord, is kommunikatiewe bekwaamheid (kommunikatiewe bevoegdheid word soms as wisselvorm gebruik).

Die toetsing van kommunikatiewe bekwaamheid is aanvanklik volgens die model van Canale en Swain gedoen, waar taalkennis as onderdeel van kommunikatiewe bekwaamheid gesien is, wat sosiolinguistiese bekwaamheid, linguistiese bekwaamheid, strategiese bekwaamheid en diskoersbekwaamheid insluit (Canale & Swain, 1980; Canale, 1983). Dit het egter ongeveer ’n dekade geneem voordat Skehan (1988:211) die problematiek hiermee akkuraat kon saamvat deur hierdie model van toepassing te maak op taaltoetsing:

[T]he central problem in language testing (is) – what and how to sample with our tests, so that decisions can be made about performance in a wide range of non-test situations. And we can only decide what and how to test by recourse to theory – theory which guides and predicts how an underlying communicative

competence is manifested in actual performance; how situations are related to one another, how competence can be assessed by examples of performance on actual tests; what components communicative competence has; and how these inter-relate. In other words, testers would like as a starting point for their endeavours a completely specified linguistic theory, which would give them a well-defined target to aim at.

Skehan (1988:213) verduidelik verder dat

competence has four components: a linguistic component concerned with the operation of the language system, e.g. syntax and vocabulary; a sociolinguistic component, which implicates our capacity to work out the meanings of actual utterances as well as to know how to use language appropriately in different situations, e.g. to change register and formality; a discourse component, which involves our ability to process language at a scale beyond the level of a sentence, i.e. an ability to participate in conversation as well as to process written texts of some size; and a strategic component, which is implicated when other components are in some way insufficient. The relationship between these competences is not entirely clear, nor is the way in which they are integrated into overall communicative competence... Nor is it clear how communicative competence is translated into communicative performance.

As navorser was Bachman (1990), en later Bachman & Palmer (1996), moontlik egter die eerste – en beslis die mees invloedryke – kenner wat die nosie van taalvermoë as bestaande uit taalkennis en strategiese vaardigheid behoorlik geëkspliseer het. Hy het dit gedoen deur die model van Canale en Swain uit te brei en dit te herorganiseer en het tot die slotsom gekom dat strategiese bekwaamheid 'n kognitiewe vaardigheid is eerder as 'n vaardigheid wat met taal verband hou (verwys ook 3.4 hier onder). Let daarop dat, na die beste van my wete, daar merkwaardig genoeg nie enige verwysing in die werk van Bachman en die van Bachman en Palmer is wat daarop dui dat hy/hulle by Skehan se voorbereidende onderskeidings en toepassing aansluit nie. Die beste afleiding wat gemaak kan word, is dat al drie hul werk gebaseer het op die van Canale en Swain.

Die nosie van taalvermoë wat bestaan uit taalkennis én strategiese bekwaamheid behoort volgens Brown (2004:10) in enige toetskonstruk gereflekteer te word. Bachman & Palmer (1996:9) wys in hierdie verband juis op die ooreenkoms wat daar tussen die prestasie op taaltoetse en die werklike taalgebruiksituasie vir daardie spesifieke konteks behoort te wees wanneer hulle sê dat “[i]n order for a particular language test to be useful for its intended purposes, test performance must

correspond in demonstrable ways to language use in non-test situations.” Dit sluit aan by Weir (1990:11) se siening dat “to measure language proficiency ... account must now be taken of: where, when, how, with whom, and why language is to be used, and on what topics, and with what effect.”

Uit die bostaande blyk dit duidelik dat toetse wat op ’n verruimde definisie van taal geskoei is, “will incorporate the impact of factors other than simple knowledge of the language... [They will] present candidates with test tasks which seem to simulate the communicative demands of the target context, and evaluate performance in terms of criteria operating in that context” (McNamara, 2004:767).

Die bespreking hierbo dui aan hoe groot die uitdagings is vir hedendaagse taaltoetsontwerpers. Nie net moet hulle ’n verruimde, kommunikatiewe siening van taal inkorporeer in hul toetsontwerpe nie; hulle moet ook ’n belyning, ’n tegniese harmonie, probeer bewerkstellig tussen hoe die taalsituasie wat hulle teiken daar uitsien, en die items / take wat in hul toets voorkom. Hierdie is nie eenvoudige uitdagings nie. Die moontlikheid om die genoemde harmonie te bewerkstellig is juis wat die ontwerpers van TAG en TALL aangespoor het om ’n nuwe toetskonstruk, en uiteindelik ’n toets, te ontwikkel. ’n Verruimde definisie van taal, wat as basis gedien het vir die ontwerp van ’n konstruk vir TAG, word vervolgens as voorbeeld bespreek.

3.4 Konteks vir die ontwerp van ’n konstruk vir ’n toets van akademiese geletterdheid

Teen die bostaande agtergrond, en in aansluiting by die breër onderskeidings wat in Hoofstuk 2 gemaak is, sal die res van hierdie hoofstuk fokus op die ontwerp van ’n toetskonstruk, as teoreties-linguistiese konsep, wat spesifiek akademiese geletterdheid van nuweling eerstejaarstudente aan Suid-Afrikaanse universiteite meet. Dit moet verder gelees word binne konteks van dit wat in Hoofstuk 1 oor die bestek en probleem van die studie geskryf is, naamlik dat kommersieel-beskikbare toetse soos die ELSA-Plus om akademiese geletterdheid te meet nie aan

verwagtings (soos ook blyk uit die voorafgaande bespreking) voldoen het nie, en dus vervang moes word.

'n Verandering in toets en toetskonstruk was dus aangewese, aangesien die kommersiële toets eerstens op die praktiese vlak moeilik was om te administreer of uit te voer (logistieke beperkinge). Volgens Brown (2004: 19) is 'n toets slegs doeltreffend as dit uitvoerbaar is; met uitvoerbaarheid word bedoel dat dit bekostigbaar moet wees, dat dit met die inagneming van tydsbeperkinge afneembaar moet wees, dat dit eenvoudig moet wees om te administreer, en dat die nasienprosedure nie ruimte vir twyfel laat nie (objektiwiteit) en veral ook nie tydrowend moet wees nie.

Die konstruk van die ELSA-Plus het ten tweede oor die afgelope dekade telkens in die kruisvuur beland omdat dit op 'n oudmodiese begrip van taal, wat taalvaardigheid aan die kennis van klank, woordeskat, vorm en betekenis gelykstel, berus. In aansluiting by punt 3.3.3.2 hierbo, voer Bachman en Palmer (1996:61ff.) oortuigende argumente daarvoor aan dat toetse hierteenoor eerder op 'n interaktiewe siening van taalvermoë gegrond moet word, wat die onderhandeling van betekenis binne bepaalde kontekste benadruk. Dit is duidelik hoe hierdie argument van Bachman en Palmer (1996) aansluit by die vierde benadering (konstruktivisme) tot toegepaste taalkundige werk. Blanton (1994:221) beweer op haar beurt en in aansluiting hierby dat selfs al word akademiese diskoers hoofsaaklik in die produktiewe modus, veral aan die hand van skryfwerk, gevoer, vereis akademies geletterde gedrag ook interaksie met tekste. In hier voege voer sy (Blanton, 1994:228) aan dat ten einde studente vir die akademiese hoofstroom voor te berei

... we must foster the behaviours of 'talking' to texts, talking and writing about them, linking them to other texts, connecting them to their readers' own lives and experience, and then using their experience to illuminate the text and the text to illuminate their experience.

Dit is ook belangrik dat toetsontwerpers, met die inagneming van hul siening van taal, oorweging daaraan skenk dat 'n toets altyd vir 'n bepaalde doel geskep word

en dat toetsresultate onvermydelik 'n uitwerking kan hê op die toekoms van kandidate wat die toets aflê, soos reeds vermeld. Toetsontwerpers moet dus altyd daartoe in staat wees om aan te toon hoe 'n spesifieke taaltoetsresultaat met taalgebruik in 'n bepaalde situasie verband hou (Bachman & Palmer, 1996:61). Aangesien prestasie op taaltoetse gewoonlik as die uitkoms van taalvermoë vertolk word - die spesifieke getoetste kenmerk van die individu - is die omskrywing van taalvermoë van kernbelang vir toetskonstruksie.

Definisies van taalvermoë bring gewoonlik twee dimensies van ervaring van taaltoetsopstellers byeen, naamlik die talige én die formatiewe. Dit is daarom nie net taal *per se* wat omskryf word nie, maar ook die mate van beheer wat die individu daaroor het (verwys ook punt 3.3.3.2 hierbo). Laasgenoemde, die vormende sy van taalgebruik, word gewoonlik in terme van konsepte soos vermoë, bevoegdheid, bemeestering en beheer (Gee, 1998:56-57), kapasiteit, vaardigheid, gesag, mag (Blanton, 1994:230), ensovoorts uitgedruk – Weideman (2009c:170-178, spesifiek p. 172) bied 'n meer volledige oorsig hieroor. Omskrywings van taalvermoë berus egter op hul beurt, soos reeds genoem, ook op die toetsontwerper se siening van taal en word gewoonlik binne daardie konteks uiteengesit.

Vir bykans veertig jaar al, maar veral die afgelope ongeveer dertig jaar, word daar aanvaar dat die ontwerp van 'n taaltoets op 'n wyer, ryker siening van taal as vroeëre beskouings behoort te berus (verwys 3.3.3 hierbo). Tweede generasie perspektiewe in die toegepaste taalkunde vereis byvoorbeeld 'n siening wat verder strek as 'n beperkende beskouing van taal, een wat taal bloot as 'n kombinasie van klank, vorm en betekenis beskou, of in tegniese linguistiese terme, tot fonologiese, morfologiese, sintaktiese en semantiese elemente beperk (cf. ook Weideman, 2009c). In ooreenstemming met sosiaal-verrykte beskouinge van taal, is die implikasie van 'n verruimde raamwerk dat taal nie bloot ekspressief is nie, maar ook kommunikatief – dit is dus bedoel om menslike interaksie mee te bemiddel en te beding, soos in punte 3.2 en 3.3 hierbo reeds gesuggereer is. Die vernaamste

verskille tussen 'n beperkende en 'n oop of verruimde siening van taal, en die implikasies daarvan vir leer en toetsing, word in Tabel 3.1 hieronder opgesom.

Tabel 3.1: Opsomming van die verskille tussen 'n beperkende en 'n oop siening van taal (uit Weideman, 2003b:38)

Beperkende	Verruimde
Taal bestaan uit die volgende elemente: <ul style="list-style-type: none"> • klank; • vorm, grammatika; en • betekenis. 	Taal is 'n sosiale instrument om: <ul style="list-style-type: none"> • in spesifieke kontekste • menslike interaksie te bemiddel en • te beding.
Hooffunksie: ekspressie/uitdrukking	Hooffunksie: kommunikasie
Aanleer van taal = bemeestering van struktuur	Aanleer van taal = bedrewenheid in kommunikasie
Fokus: taal	Fokus: proses van taalgebruik

Onder die rubriek *Fokus*, sluit die tabel 'n punt, naamlik *proses*, in wat nie net vir leer en onderrig nie, maar ook vir toetsing van belang is. Soos Blanton (1994:231) tereg uitwys, is taalklasse, met 'n fokus op akademiese geletterdheid, klasse waar dit eerder oor taal as kommunikatiewe instrument handel en taal as sulks nie noodwendig van belang is nie. Dit toon bepaalde raakpunte met die Suid-Afrikaanse tersiêre werklikheid, naamlik dat daar nie aanvaar kan word dat klaarblyklike sukses in 'n taalvak op skool regstreeks met 'n eerstejaarstudent se akademiese geletterdheidsvlak verband hou nie. Voorts hang dit ten nouste saam met 'n punt wat Gee (1998:58) maak, naamlik dat die suksesvolle verwerwing van akademiese geletterdheid deur kinders in hoofstroomonderwys dikwels as 'n implisiete en 'n nie-regstreekse uitkoms of neweproduk van onderrig en leer plaasvind. Akademiese geletterdheid het volgens hierdie beskouing 'n doelbewuste, herformuleerde omskrywing nodig wat dit nie deur 'n streng afgebakende siening van taal wat dit tot vorm, klank en betekenis reduceer, sal laat beperk nie.

Vir studies in taalonderrig en die toegepaste linguistiek gaan hierdie verruimde en ryker beskouing so ver terug as die werk van Hymes (1972) oor kommunikatiewe bekwaamheid, en 'n tradisie gegrond op hierdie sosiolinguisties-pragmatiese idee wat onder andere ook tot uitdrukking gekom het in die werk van etlike ander taalkundiges en toegepaste taalkundiges soos Searle (1969), Halliday (1978, 1985)

en Wilkins (1976). Vir 'n geslag taalkundiges vir wie die klem op sintaktiese ontleding geval het, is dit miskien paslik om daarop te let dat hierdie breër perspektief vir ten minste sestig jaar reeds in taalteorie teenwoordig is. Weideman (2009c:183,184) merk op dat Hjelmslev reeds dekades gelede aangedui het dat taalteorie, ná 'n tydelike inperking daarvan, nie alleen daardeur gerig word om bloot die taalstruktuur te erken nie, maar ook deur die gebruikers van daardie taal, die samelewing waarbinne dit gebruik word, en die menslike kennissfeer agter taal te erken. Kortom: 'n sosiaal-verruimde perspektief op taal het ook 'n effek op wat die taaltoetsers, en wat die breër institusionele en samelewingskonteks aan wie die taaltoetsers onder meer verantwoordig is, verskuldig is, beskou as die moeite werd om te meet.

Die derde rede vir 'n heroorweging van die huidige toetskonstruk hou met die tweede verband, en gaan oor die kritiek teen 'n vaardighedsgebaseerde benadering tot die omskrywing van taalvermoë, asook die uitwerking van sodanige benadering op toetsing. Die taaltoetsers se perspektief op taal kan tereg as onaanvaarbaar gebrekkig beskou word as dit net op 'n samestelling van vaardighede (gewoonlik luister-, praat-, lees- en skryfvaardighede) gebaseer is, aangesien dit in taalonderrig maklik tot die veronderstelling aanleiding kan gee dat addisionele taal-leerders 'n taaltkort of -gaping het wat bloot, of selfs ideaal gesien, gevul kan word deur onderwysers wat hulle die nodige taalvaardighede by wyse van spreke kan gee. So 'n siening staan in teenstelling met die beskouing waarvolgens taal, ook akademiese diskoers, as 'n verwerfbare bevoegdheid beskou word of ten minste as iets wat aangeleer of ontwikkel moet word. Dit is ook in teenstelling dus met die beskouing dat akademiese geletterdheid as 'n besondere (sekondêre) diskoers verwerf en aangeleer behoort te word (Gee, 1998).

Met betrekking tot taaltoetsing is Van Dyk en Weideman (2004a:6) dit eens met Bachman en Palmer (1996:75) se kritiek teen die omskrywing van taalvermoë in terme van vaardighede (verwys ook punt 3.3.3.2 hierbo). Laasgenoemdes sonder een van die tekortkominge van so 'n siening uit, naamlik dat wyd uiteenlopende

take soos om na 'n gespreksgenoot of na die radio te luister beide as een aktiwiteit, naamlik luister, geklassifiseer sou kon word. Hulle kom derhalwe tot die gevolgtrekking dat taalvaardighede nie as deel van taalvermoë beskou kan word nie, maar wel as die konteksgebonde verwesenliking van die vermoë om taal in die uitvoering van bepaalde taalgebruikstake te kan aanwend. Volgens hulle is dit dus nie van enige nut om in terme van vaardighede te dink nie, maar eerder in terme van spesifieke aktiwiteite of take waarin taal doelbewus gebruik word. Dit is interessant om op te merk dat hierdie radikale kritiek van twee vooraanstaande deskundiges op 'n vaardigheidsbenadering nie veel prominensie geniet in diskussies oor die konstruksie wat taaltoets moet meet nie.

Van Dyk en Weideman (2004a:6) meen nogtans dat sodanige kritiek geldig is en wys daarop dat daar terselfdertyd aan akademiese taalvaardigheid as akademiese diskoers gedink moet word, sowel as aan die problematiek rondom die meet van hierdie vermoë in terme van die vlak waarop dit bemeester is. In hierdie verband merk Blanton (1994:230) tereg op dat

... individuals whom we consider academically proficient speak and write with something we call *authority*; that is one characteristic – perhaps the major characteristic – of the voice of an academic reader and writer. The absence of authority is viewed as powerlessness ...

Die hoofuitdaging vir 'n toets van akademiese taalvermoë is daarom die meting van die akademiese geletterdheidsvlak. Soos wat uit die onderstaande sal blyk, word daar nie saamgestem met die siening dat akademiese geletterdheid bloot uit vaardighede en strategieë vir akademiese sukses bestaan, of dat dit bloot uit luister-, lees- en skryfvaardighede bestaan, met studie en kritiese denkprosesse dalk op die koop toe aangelas nie. Van Dyk en Weideman (2004a:5,7) wys daarop dat 'n ryker, verruimde beskouing van taal in die algemeen en akademiese taalvermoë in die besonder, die ooglopende geldigheid van enige toetskonstruk aansienlik sal verhoog. Sodanige konstruk en die ontwikkeling daarvan word vervolgens in meer besonderhede ondersoek.

3.4.1 Die voorgestelde nuwe konstruk

3.4.1.1 Toetskonstrukte

Die toetskonstruk en gepaardgaande bloudruk (toetsspesifikasies – verwys Hoofstuk 4 en Bylaag B) omskryf in analities-wetenskaplike terme die kennis of vermoëns wat deur daardie bepaalde toets gemeet moet word. 'n Konstruk kom gewoonlik in terme van 'n teorie tot uiting - in hierdie geval 'n taalteorie, en meer spesifiek 'n akademiese-geletterdheidsteorie- en word deur onder andere die regverdiging, deur middel van artikulasies daarvan, en die implementering daarvan ontsluit (wat albei deel uitmaak van die proses van validering). Davies *et al.* (1999:31) omskryf die begrip toetskonstruk as "... an ability or set of abilities that will be reflected in test performance, and about which inferences can be made on the basis of test scores." Hierdie omskrywing veronderstel egter 'n algemene punt oor alle toetse: ten einde as geldig beskou te word, moet toetse dít meet wat dit veronderstel is om te meet, en nie enigiets anders nie. Terselfdertyd is dit wat gemeet word 'n hipotetiese, dit wil sê teoreties-wetenskaplik gekonspieerde, vermoë en is die geldigheid van die instrument waarmee gemeet word se resultate afhanklik van interpretasie. Die verbinding van toetsuitkomst of -resultate met 'n (geldige) interpretasie versterk verder een van die hoofargumente van die huidige ortodoksie in taaltoetsing, wat van mening is dat geldigheid altyd afhanklik is van interpretasie, want sonder 'n behoorlike, wetenskaplik-verantwoordbare idee van die hipotetiese taalvermoë wat deur 'n taaltoets gemeet word, is geen afdoende en teoreties-verdedigbare interpretasie moontlik nie.

Soos in die bespreking hieronder (verwys veral Hoofstukke 6 en 7) duidelik sal blyk, is hierdie afhanklikhede tussen byvoorbeeld die tegniese krag van die toets (soos uitgedruk in die objektiewe toetsresultate), en die interpretasie daarvan in die lig van die teoreties-gedefinieerde vermoë wat gemeet is, hoewel nou verweef en interafhanklik, wel konseptueel onderskeibaar. Die begrip tegniese krag of resultaat verwys na die verband tussen die tegniese, leidende funksie van die ontwerp en die fisiese aspek van ons ervaring. Die tegniese interpretasie van die resultate is konseptueel verbind aan die verwysing na die linguale dimensie. Die teoretiese

definisie van taalvermoë verwys weer na die funderende funksie wat teoretiese analise speel in die tegniese produk, wat in hierdie geval 'n taaltoets is.

In dieselfde gees as Bachman en Palmer (1996:61-79) se beskrywing van taalvermoë, omskryf Douglas (2000:111) 'n toetskonstruk as daardie deel van 'n toets wat die aard van die vermoë wat getoets word (insluitend grammatikale, tekstuele, funksionele en sosiolinguistiese kennis, strategiese bekwaamheid en agtergrondkennis) teoreties artikuleer. Konstrukt geldigheid verwys verder onder andere na die mate waartoe die onderliggende teorie in die toets verteenwoordig word (Davies *et al.*, 1999:33). Die kwessie van geldigheid en validering word elders in hierdie studie (verwys Hoofstukke 6 en 7) in meer besonderhede bespreek.

3.4.1.2 Konstrukte van akademiese geletterdheid

Akademiese geletterdheid is volgens Van Schalkwyk (2008:18) 'n komplekse begrip om te omskryf, aangesien dit vanuit verskeie velde gedefinieer en selfs toegeëien word, onder andere die linguistiek, die antropologie, die sosiologie, die sielkunde, en die opvoedkunde. Deesdae word selfs die tegnologie en die visuele bygereken (Thesen & Van Pletzen, 2006:10). In hierdie verband wys Kern (2000:23) daarop dat daar verskeie spesifieke definisies of konstrunkte (die term *definisie* word hier in die breedste betekenis van die woord gesien en kan ook as sinoniem vir die term *konstruk* gebruik word) van akademiese geletterdheid binne verskillende velde geld en dat dit moeilik is om 'n enkele definisie vas te pen, wanneer hy opmerk “literacy is an elastic concept: its meaning varies according to the disciplinary lens through which one examines it.”

Hierdie studie is veral geïnteresseerd in 'n konstruk vir akademiese geletterdheid vanuit die talige perspektief, maar binne konteks van dit wat in paragraaf 3.4 hierbo genoem is, naamlik dat dit nie beperk gesien behoort te word nie. Met erkentlikheid hieraan, en die feit dat die Suid-Afrikaanse hoëronderwyslandskap tans meer heterogeen daar uitsien as 'n dekade of twee gelede in terme van onder andere kulturele, sosio-ekonomiese en veral taalsamestelling, word daar vervolgens gepoog om 'n konstruk vir akademiese geletterdheid op grond van akademiese

geletterdheidsteorie tot stand te bring met die inagneming van veral die werk van Bachman en Palmer (1996), Blanton (1994), Boughey (2000), Gee (1990; 2003), Kress (1989), Lea en Street (1998), McKenna (2004a), Van Dyk en Weideman (2004a), Van Schalkwyk (2008) en Yeld (2001).

Boughey (2000:281) beskryf akademiese geletterdheid as die vermoë om aan akademiese diskoers deel te neem, maar wys daarop dat studente dit eers sal verwerf en ontwikkel as hulle saam met ander aan die diskoers blootgestel word en daaraan deelneem, met die inagneming daarvan dat elke dissipline by implikasie 'n eie diskoers het. Onderliggend aan hierdie definisie is daar egter twee potensieel aanvegbare kwessies. Die eerste is dat ondersoekers 'n meer generiese siening hieroor kan huldig, naamlik dat akademiese diskoers as sulks wel gedefinieer kan word en nie noodwendig net as dissipline-spesifieke diskoers nie. Die tweede is dat blote blootstelling aan die diskoers sal bydra tot die ontwikkeling van akademiese geletterdheid.

Nogtans word hier duidelik 'n nuwe dimensie tot die definiëring van akademiese geletterdheid toegevoeg, naamlik die van diskoers. Met die term *diskoers* word binne konteks van hierdie studie bedoel “systematically organised sets of statements which give expression to the meanings and values of an institution. Beyond that, they define, describe and delimit what ... is possible to say and [what] not...” (Kress, 1989:7). Gee (1990:143) beskryf dieselfde term weer as “... a socially accepted association among ways of using language, of thinking, feeling, believing, valuing and of acting that can be used to identify oneself as a member of a ... ‘social network’ or to signal ... a socially meaningful role”. Van Schalkwyk (2008:25) wys verder daarop dat, hoe die term *diskoers* ook al beskryf word, akademiese geletterdheid uiteindelik beoordeel sal word aan die hand van studente se bemeestering van die genoemde diskoers.

Gee (2003) neem sy definisie van akademiese geletterdheid 'n paar jaar later iets verder en wys daarop dat die kernelemente van geletterdheid *lees* en *skryf* is en dat akademiese geletterdheid by implikasie dus daardeur ten toon gestel moet word.

Die belang van lees en skryf in enige definisie van akademiese geletterdheid word ook gereflekteer deur Lea en Street (1998:160) se opmerking dat “academic literacy practices - reading and writing within disciplines - [are those that] constitute central processes through which students learn new subjects and develop their knowledge about new areas of study.”

In lyn met die vroeëre opmerking dat lees en skryf miskien nie so gemaklik in die werklikheid van die akademie isoleerbaar is nie, waarsku Williams (2004:576) dat daar tog gewaak moet word daarteen om akademiese geletterdheid te beperk tot bloot net “the ability to read and write.” Studente moet in staat wees om verworwe kennis en vaardighede op die gebied van akademiese geletterdheid te gebruik “[to] switch practices between one setting and another, to deploy a repertoire of linguistic practices appropriate to each setting, and to handle the social meanings and identities that each evokes” (Lea & Street, 1998:159). ’n Verruimde definisie van akademiese geletterdheid is volgens Williams (2004:594) nie geïnteresseerd in die “psycholinguistic process of reading and writing... [or] with the adequacy of reading comprehension or written production in terms of a prescribed educational standard” nie, “but rather with literacy as social practice.” Of, soos Leibowitz (2001:2) dit stel: “[it is a] culturally specific set of linguistic and discourse conventions, influenced by written forms utilised primarily in academic institutions.”

Sonder om enigsins afbreuk aan die bostaande te doen, kom die eerste werklik geïntegreerde definisie of konstruk van akademiese geletterdheid myns insiens voor in Blanton (1994:226) se beskrywing van wat akademies geletterde studente behoort te kan doen. Sy maak hierin korte metten van die opvatting dat ’n mens akademies taalvaardig raak deur bloot aan die akademiese omgewing blootgestel te word en deur geringe hoeveelhede woordeskat en taalkennis te leer. Sy erken ook by implikasie die belangrike rol van identiteit en die sosiale konteks waarin die diskoers deur middel van lees en skryf gevoer word. Indien ’n kommunikatiewe, interaktiewe opvatting oor akademiese diskoers gehuldig word, sou ’n akademiesegeletterdheidstoets volgens haar dus nie net waarde heg aan studente se

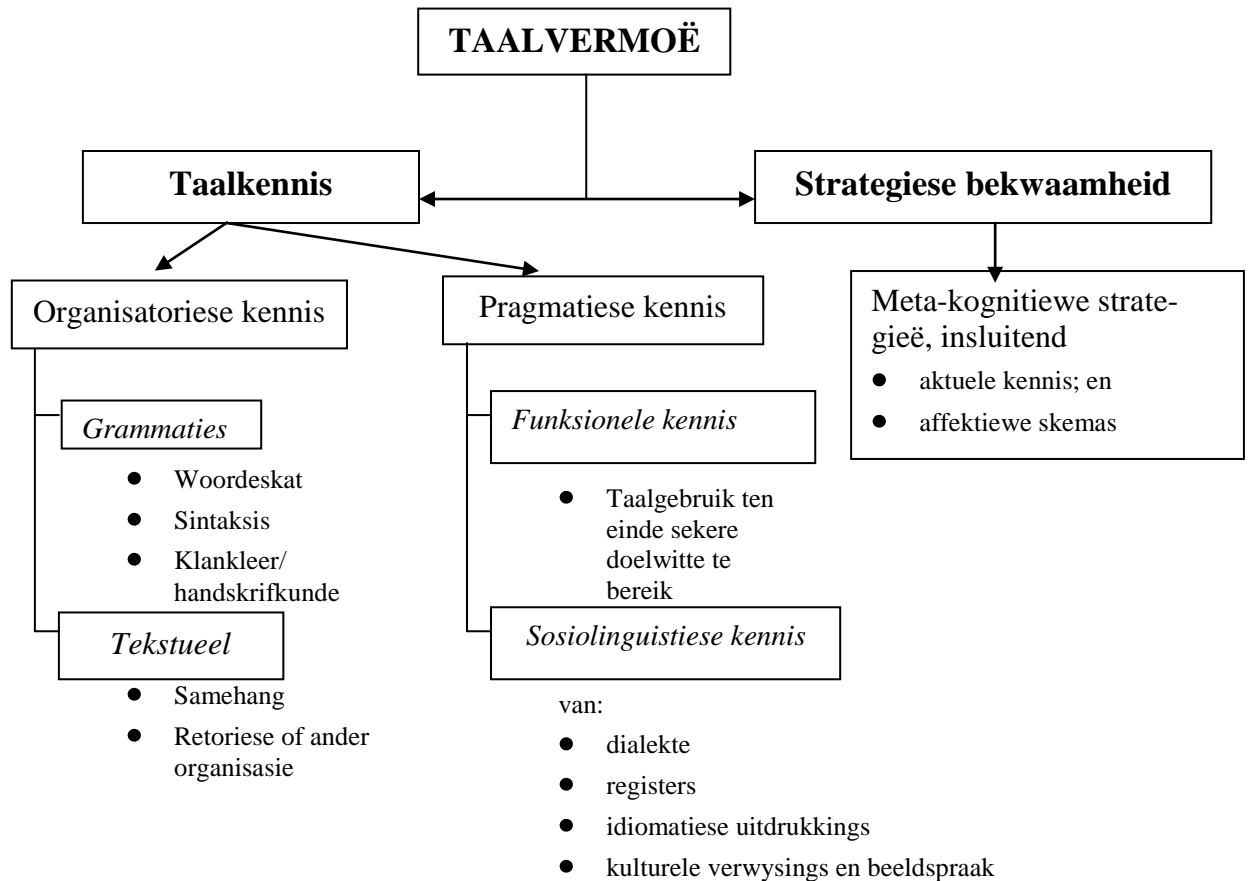
vermoë om nuwe woordeskat en taalkennis te verkry en te leer nie, maar ook na die mate waartoe studente die volgende akademiese aksies, in verhouding tot hulself en tot ander, suksesvol kan uitvoer:

- tekste in die lig van hul eie ervaring te kan vertolk, en ook eie ervarings in die lig van tekste te vertolk;
- in die lig van daardie ervaring met tekste saam te stem of te verskil;
- tekste met mekaar in verband te bring;
- tekste saam te voeg, en die nuutgeskepte eenheid te gebruik om nuwe menings te vorm of argumente te bou;
- tot opinies te kan kom deur van inligting uit tekste gebruik te maak (ekstrapolasie);
- hul eie tekste deur middel van enige of al bogenoemde te skep;
- oor die uitvoering van enige of al bogenoemde te kan praat en te kan skryf; en
- laasgenoemde twee punte ter bevrediging van die betrokke gehoor uit te voer (Blanton 1994:226).

Die waarde van Blanton se omskrywing lê daarin dat dit breed en inklusief is. Soos uit die onderstaande sal blyk, bevat dit komponente van en aksies geassosieer met akademiese geletterdheid, soos vervat in meer uitvoerige uiteensettings van hierdie idee. Die begrip van die student se “eie stem” figureerbyvoorbeeld sterk in meer as een komponent van Blanton se omskrywing. Eweneens word die taalfunksies en - vermoëns wat met akademiese geletterdheid gepaardgaan, soos ekstrapolering, samevoeging en redevoering, spesifiek hier vermeld. Soos hieronder gemerk sal word, is hierdie funksies daardie wat telkens prominent figureer in ander definisies van akademiese geletterdheid.

In Bachman en Palmer (1996) se werk word ’n selfs meer gedetailleerde omskrywing gevind, waarin aspekte van lees en skryf en van identiteit opgesluit lê. Volgens hulle omskrywing, en soos ook in figuur 3.2 hieronder aangetoon, staan

taalvermoë (of die meting van taalvermoë) op twee bene, naamlik taalkennis en strategiese bevoegdheid (1996:67):



Figuur 3.2: Die Bachman en Palmer-konstruk

Hierdie meer omvattende uiteensetting laat egter ’n aantal vrae ontstaan. Kan daar byvoorbeeld werklik tussen retoriese organisasie en funksionele keuse ’n onderskeid gemaak word? Is dit moontlik om taalgebruik, ten einde kommunikatiewe doelwitte te bereik (sogenaamde funksionele kennis), van die gebruik van idioome, metafore en register (sogenaamde sosiolinguistiese kennis) te onderskei? Hoe verskil die keuse van ’n register (as ’n deel van pragmatiese, sosiolinguistiese kennis) op sy beurt van die taalgebruiker se keuse van ’n bepaalde organisatoriese vorm (retoriese kennis)? Hoe kan daar voorts tussen aktuele kennis (onder strategiese bevoegdheid) en die toepaslike keuse en gebruik vir ’n betrokke situasie (vermoedelik die uitkomst van sosiolinguistiese kennis) ’n onderskeid

getref word? Laastens, hoe kan argumentatiewe skemas (vermoedelik hier deel van strategiese bevoegdheid), registerkeuse (pragmatiese kennis) en tekstuele kennis (taalorganisasie) afsonderlik van mekaar beskou word?

Hierdie potensiële verwarring wat daar kan ontstaan tussen die aard van al hierdie komponente word inderdaad deur sommige ander vertolkings van die Bachman en Palmer-konstruk geïllustreer. In 'n bespreking hiervan bring Brindley (2002:461) byvoorbeeld die vermoë om bepaalde kommunikatiewe doelwitte te bereik, met strategiese, en nie met pragmatiese bevoegdheid nie, in verband. Soos Brindley (2002:462) in der waarheid dan ook uitwys, is die beskrywing van strategiese bekwaamheid een aspek in terme waarvan hierdie konstruk in die kruisvuur beland het.

Daar sou egter ook aangevoer kon word dat die klaarblyklike deursyfering en oorvleueling tussen kategorieë en subkategorieë van die konstruk dit nie noodwendig minder nuttig maak nie. Bachman en Palmer (1996:66) waarsku juis dat hoewel hulle voorgestelde raamwerk vir enige taaltoetsituasie van nut kan wees, dit steeds vir elke bepaalde situasie en gebruik binne konteks geplaas en vertolk sal moet word, wanneer hulle sê dat

[w]e ... need to define language ability in a way that is appropriate for each particular testing situation, that is, for a specific purpose ... The way we define language ability for a particular testing situation, then, becomes the basis for the kinds of inferences we can make ... (1996: 66).

Volgens hulle moet taalvermoë dus vir elke bepaalde toetsituasie toepaslik omskryf word ten einde geldige afleidings daaruit te kan maak alvorens besluite (oor studente) geneem kan word na aanleiding van die meting van hul taalvermoë.

Só 'n herversolking van die oorspronklike Bachman en Palmer-konstruk vir die besondere konteks van hoëronderwys, word gevind in die werk van Nan Yeld, verbonde aan die Universiteit van Kaapstad (UK), en word beskryf in Yeld (2001). Hoewel hulle die oorspronklike Bachman en Palmer-verdeling behou het, met die besware daarteen soos hierbo genoem, het hulle 'n interpretasie en begrip van tipiese akademiese take, hoofsaaklik gegrond op insette deur kundige panele,

daarby gevoeg(Yeld, 2001). 'n Begrip van Bachman en Palmer (1996:68) se konstruk word daarom onder andere deur die identifisering van 'n hele aantal taalfunksies (aanvanklik onder “pragmatiese kennis” ingedeel) en akademiese geletterdheidstake verryk. Dit sluit onder andere inligtingsbegrip, omskrywing, opsomming, beskrywing, redevoering, klassifisering, kategorisering, vergelyking, en teenstelling in. Hierdie hervertolking is dus inderdaad verenigbaar met ander se gebruik van hierdie bepaalde raamwerk.

Die perspektiewe op akademiese geletterdheid(stoetsing) hierbo uiteengesit, word vervolgens in meer besonderhede bespreek en maak grootliks gebruik van interpretasies en insigte wat reeds verwoord is in voorstudies vir hierdie ondersoek in Van Dyk en Weideman (2004a, b).

3.4.1.3 Van teorie na praktyk: die ontwerp van 'n konstruk vir akademiese geletterdheid

Uit die definisies wat hierbo beskryf en ontleed is, blyk die Alternative Admissions Research Project (AARP) van die Universiteit van Kaapstad se definisie van akademiese geletterdheid vir doeleindes van akademiese geletterdheidstoetsing aan Suid-Afrikaanse universiteite die nuttigste te wees (Yeld, 2001). Deur middel van institusionele samewerking met die Universiteit van Kaapstad en 'n aantal ander hoëronderrysinstellings met die oog op die ontwikkeling van toelatingstoets vir Suid-Afrikaanse universiteite, is daar eerstehands met hierdie toets (PTEEP – Placement Test in English for Educational Purposes) kennis gemaak wat die meetinstrument is wat uit die konstruk gevloei het. Maar ook híérin, met ander woorde in die doel waarvoor die toets gesamentlik ontwikkel is, het 'n potensiële probleem gelê, naamlik dat die bedoeling waarmee die toets ontwerp is, geheel en al van die redes waarom 'n akademiese geletterdheidstoets aan die Universiteit van Pretoria (en later ook sy vennote) nodig was, verskil het. Kortweg gestel, moes en moet die akademiesegeletterdheidstoets binne die AARP-toetsbattery die potensiaal vir akademiese sukses billik meet, en is dit ook nog altyd so gebruik. Aangesien hierdie toets as 'n akademiese potensiaaltoets voorgehou word, moet dit noodwendig nie net in 'n beperkte sin na taalvermoë kyk nie, maar ook na begrip

(veral die kandidaat se aanwending van kognitiewe strategieë). Die metings hiervan deur middel van PTEEP vind gewoonlik plaas gedurende die jaar voordat toegang tot universiteitstudie verlang word, en die toetsresultate word onder andere gebruik om sodanige kandidate toegang tot deelnemende instellings te bied (of te weier). As sulks is dit 'n toets wat onmiddellike politieke implikasies inhou, en waarvan die gevolge binne 'n hoërondewysomgewing met toegang tot studiegeleenthede as wagwoord, moontlik aanvegbaar is. Alhoewel die rol van PTEEP en AARP aan die vervaag is, omdat dit oorspoel sal word deur die National Benchmark Tests, wat waarskynlik binnekort op nasionale vlak naas die huidige Nasionale Senior Sertifikaat-resultate vir toelating tot hoërondewys gebruik sal word, was dit tog baanbrekerswerk waarvan die invloed (ook op TAG en TALL) nie onderskat moet word nie.

Aan die ander kant is daar die institusionele bedoelinge van die Universiteit Stellenbosch, die Noordwes Universiteit en die Universiteit van Pretoria wat in die algemeen eers 'n akademiese geletterdheidstoets na 'n student se toelating afneem, wat nie net die doel van die toets (om reeds toegelate risikostudente te identifiseer) verander nie, maar dit ook versag in terme van die politieke gevolge daarvan. Toetsresultate word ook verskillend aangewend: in die een geval word dit gebruik om toegang te verleen (of te weier); in die ander om op 'n toepaslike kursus geplaas te word om die mislukingsrisiko te verminder. Ter opsomming is eersgenoemde duidelik as 'n toelatingstoets, en laasgenoemde ten minste aanvanklik as 'n plasingstoets ontwikkel.

Daar is egter 'n verdere komplikasie in die aanwending van TAG in die opsig dat die Universiteit Stellenbosch in 2007 die besluit geneem het om die akademiese geletterdheidstoets (aanvanklik gebruik as 'n plasingstoets) vanaf 2008 as deel van hierdie universiteit se toegangstoetsbattery (wat uit potensieel vyf toetse bestaan) te gebruik, aangesien die twee bestaande taaltoetse (in Afrikaans en Engels) van hierdie battery statisties en empiries beskou nie bevredigend gemeet het nie. Dit word dus onmiddellik 'n toets met medium- tot hoërbelang, aangesien dit 'n student

die geleentheid kan ontnem om aan tersiêre studie te kan deelneem. Ter verdediging van hierdie besluit en vir doeleindes van hierdie studie, moet dit egter duidelik gestel word dat die akademiese geletterdheidstoets, soos reeds genoem, deel van 'n battery is en nie alleen vir toelatingsdoeleindes gebruik word nie, maar altyd in kombinasie met ten minste een of meer van die ander toetse in die battery. Verder word toegang ook bepaal met inagneming van die matriekresultaat, in 'n 40:60 verhouding, waarvan laasgenoemde die gewig van die gemiddelde matriekresultaat sal wees. Die akademiese geletterdheidstoets word terselfdertyd steeds as plasingstoets by hierdie instelling gebruik en die fokus van hierdie studie is eweneens ook op die verantwoordelike gebruik van die toets as plasinginstrument.

Aangesien daar dus in die reël meer op die spel is by eersgenoemde toets (die AARP-toetse wat as toelatingstoetse aangewend word), regverdig dit 'n uiters versigtige benadering. Dit word daarom ook in die algemeen as deel van 'n groter stel akademiese potensiaaltoetse afgeneem- die AARP-toetse sluit byvoorbeeld ook wiskunde- en wetenskapbegrips- en -prestasietoetse in. Een van die gevolge hiervan is dat die huidige taal- en begripstoets binne die AARP-stel langer as twee en 'n half uur neem om te voltooi, en langer as drie uur om af te neem. Aangesien die doelwitte van die Noordwes Universiteit en die Universiteite van Pretoria en Stellenbosch hiermee verskil in die opsig dat dit ook as plasinginstrument gebruik word en objektiewe resultate so spoedig moontlik na voltooiing beskikbaar moet wees, en in die lig van 'n populasie van byvoorbeeld meer as 7 000 studente by die Universiteit van Pretoria alleen wat op een dag getoets moet word, is tyd veel meer beperk. Bogenoemde het dus gelei tot 'n herkonseptualisering van hoe 'n toets wat op bogenoemde drie instellings se behoeftes afgestem is, ontwerp behoort te word.

Gegewe die logistieke en ander beperkinge, was die hoofuitdaging by die ontwikkeling van 'n nuwe toets dus om te besin of dieselfde konstruk as die van AARP op die voorgestelde akademiese geletterdheidstoets toegepas kon word, met dien verstande dat 'n reeks toepaslike spesifikasies na aanleiding van bogenoemde

konstrukte ook ontwikkel moes word. Ervaring oor die voorafgaande paar jaar met die ontwikkeling van 'n aantal weergawes van die AARP se taal- en begripstoets het in hierdie verband handig te pas gekom, en het daartoe gelei dat 'n aantal van die produktiefste elemente van die oorspronklike bloudruk geïdentifiseer en herbenut is. Op grond van hierdie ervaring is daar tot die slotsom gekom dat daar met redelike vrymoedigheidvoortgegaan kan word om die oorspronklike bloudrukke te rasionaliseer, te herskik en te herformuleer ten einde 'n meer doeltreffende weergawe te ontwikkel wat dit makliker sou maak om akademiese geletterdheidsvlakke betroubaar en binne veel strenger tydsbeperkings te toets.

Die voorgestelde konstruk (Weideman, 2003a:xi) vir 'n toets van akademiese geletterdheid, vereis daarom die volgende van studente:

- begrip binne konteks van 'n omvattende akademiese woordeskatreeks;
- die vertolking en gebruik van metafore en idiome, en begrip van bybetekenis, woordspeling en dubbelsinnigheid;
- begrip van verhoudings tussen verskillende teksdele, bewustheid van die logiese ontwikkeling van 'n (akademiese) teks vanaf inleiding tot gevolgtrekking, en kennis van taalgebruik wat die verskillende teksdele samehangend aanmekaar bind;
- die vertolking van verskillende teksoorte (genre), en sensitiwiteit vir die betekenis wat daardeur oorgedra word en die gehoor op wie dit gemik is;
- die vertolking, gebruik en skep van inligting in grafiese of visuele formaat;
- onderskeid tussen belangrike en nie-belangrike inligting, feit en opinie, stellings en argumente, en oorsaak en gevolg, en die indeling en hantering van data wat vergelykings tref;
- die raaksien van opeenvolging en volgorde, en die doen van eenvoudige berekenings met betrekking tot akademiese inligting waardeur vergelykings getref en 'n argument gevoer kan word;
- kennis van wat as staving vir 'n argument dien, en die vermoë om deur middel van afleidings tot opinies te kan kom / dit te kan ekstrapoleer, en dit, of die implikasies daarvan, op ander gevalle as die een voor hande toe te pas;

- begrip van die kommunikatiewe funksie van verskillende uitdrukkingswyses in akademiese taal (soos omskrywing, die verskaffing van voorbeelde, redenering, definiëring); en
- die soeke na betekenis (van byvoorbeeld 'n akademiese teks) buite die grense van die sin, asook die vergelyking van een deel van 'n teks met 'n ander, en een teks met 'n ander.

Soos wat duidelik uit hierdie omskrywing van akademiese taalvermoë blyk, is daar van die ietwat kunsmatige en ongetwyfeld aanvegbare verdeling van die oorspronklike konstruk afstand gedoen. Daar sou ook geargumenteer kon word dat, deur die definisie van akademiese geletterdheid in terme van komponente uit te druk, 'n mens juis in bogenoemde vermoëns dít kan vasvang wat op tersiêre vlak in terme van veral lees en skryf van studente vereis word. Hierdie siening van akademiese geletterdheid is by verskeie seminare en in kongresaanbiedings deur Van Dyk en Weideman (cf. 2004a) vir bespreking voorgelê, waar dit wye en positiewe reaksie uitgelok het. In die lig van die ontwikkelingsgeskiedenis van die konstruk – ontwikkeling wat oorlegpleging met panele akademici van regoor die verskillende dissiplines behels het – behoort dít nie te verras nie. Die algemene reaksie van die gehore het dié van die aanvanklike konsultasies bevestig. Die reaksie was dat bogenoemde elemente inderdaad 'n aantal wesentliche komponente van akademiese geletterdheid verteenwoordig. Akademici van regoor die dissiplinêre spektrum vind daarom sterk aanklank by die voorgestelde konstruk, wat daarop dui dat Van Dyk en Weideman (2004a) se siening van die konstruk vir 'n toets van akademiese geletterdheid gepas blyk te wees of ten minste van so 'n aard is dat dit wyd gunstig ontvang behoort te word. Dit word bevestig deur die aantal instellings, bo en behalwe die drie hierbo genoem, wat óf 'n behoefte uitgespreek het om deel te neem aan die ontwikkeling of gebruik van die nuwe toets, óf belangstelling getoon het om studente op dieselfde manier by te staan.

'n Ongetwyfelde verdere voordeel daaraan verbonde om met 'n konstruk soos bogenoemde te werk, is die positiewe terugspoeleffek (Brindley 2002:467). In

hierdie geval bied die toets reeds 'n aanduiding van wat uiteindelik onderrig sal word; so ook weerspieël die kursus die toetskonstruk. Op sy beurt verbeter dit die ooglopende geldigheid van die toets. Aangesien die toetskonstruk in lyn met die huidige uitkomsgebaseerde benadering van Suid-Afrikaanse hoërsonderwys in terme van 'n reeks uitkomst geformuleer is, toon dit boonop 'n aantal soortgelyke voordele, soos 'n nouer band tussen assessering en leer, groter deursigtigheid in verslaglewering, en beter kommunikasie tussen belanghebbendes (Brindley 2002:465). Inderdaad, in die lig van die oorwegend negatiewe effekte van die terugspoeleffek by ander toetse, en spesifiek akademiese geletterdheidstoetse, is dit moeilik om die positiewe terugspoeleffek wat by TAG waargeneem kan word volledig na waarde te skat.

Die meer stroombelynde konstruk wat vir TAG gebruik word, lok tans ook waardering uit. Van Dyk en Weideman (2004a) verruim byvoorbeeld volgens Van Schalkwyk (2008:27) bestaande definisies of beskrywings van akademiese geletterdheid in die konstruk wat hulle vir 'n toets van akademiese geletterdheid (vanuit die talige perspektief) voorhou. Van Schalkwyk (2008:27-28) meen verder hierdie definisie bied "... a sensible, coherent and apparently objectively measurable description of how students might demonstrate the extent to which they are academically literate." Sy (2008:28) spreek egter ook kritiek teen hierdie definisie van akademiese geletterdheid uit in die opsig dat dit verleidelik kan wees, juis omdat dit so omvattend en waardevol blyk te wees: "The fact that the list lies so neatly in black and white on the page can, unless care is taken, undermine the haze and fusion that results from the dynamic social context within which learning takes place..."

Myns insiens is hierdie kritiek geldig, veral gesien in die lig van 'n breër definisie vir akademiese geletterdheid waar die konteks en die persone betrokke in daardie konteks nie buite rekening gelaat kan word nie. Die kwessie van groter deursigtigheid in terme van die sosiale belang (die feit dat hierdie konstruk op 'n gegewe moment op sekere kandidate onder sekere omstandighede vir plasings- en

toelatingsdoeleindes gebruik word) kom daarom weer op die voorgrond. Hoewel uiters belangwekkend, lê dit egter eerder op die afgebakende terrein van die studie wat Rambiritch tans onderneem, en waarna daar reeds vantevore verwys is.

Let daarop dat woorde ‘nuut’ of ‘nuwe’ telkens in die voorafgaande bespreking voorkom. Met hierdie woorde word slegs aangedui dat die konstruk van TAG en TALL die ELSA-Plus se konstruk vervang het en dus in hierdie opsig nuut is. Dit word nie geïmpliseer dat die konstruk vir TAG en TALL van nuuts af geskep is nie; intendeel, die werk wat by AARP en deur prof. Nan Yeld van die Universiteit Kaapstad (veral in haar doktorsale tesis) in die verband gedoen is (en wat weer op die werk van Bachman en Palmer gebaseer is), was die fondasie waarop die konstruk van TAG en TALL gebou is en beide prof. Yeld en die AARP-span verdien hier spesiale erkenning.

3.5 Samevatting

Hierdie hoofstuk het hoofsaaklik op die redes vir ’n verandering van konstruk binne die konteks van akademiese geletterdheidstoetsing gekonsentreer, asook hoe die keuse van ’n konstruk regverdig kan word in die lig van paradigmas in die toegepaste linguistiek in die algemeen, en van benaderings tot taaltoetsing in die besonder. In die gees van Shohamy (2001) se aansporing aan taaltoetsers om elke toetsresultaat te ondersoek en elke voorsienbare implikasie te heroorweeg, verteenwoordig dit ’n poging om die konseptualisering van ’n toets te verduidelik en binne konteks te plaas. Ten einde uiteindelik openbaar verantwoordbaar te wees soos wat in hierdie era van groter deursigtigheid van alle maghebbende instellings vereis word, is dit van wesenlike belang dat daar dus begin word om die storie van die hoofinstrumente in die uitoefening van die genoemde mag te vertel – in hierdie geval, die storie van TAG. Die wye belangstelling wat hierdie nuwe toetse reeds by ander hoërsonderwysinstellings uitgelok het, beteken dat dit nie tot ’n enkele, plaaslike konteks beperk kan wees nie, en dat die toets se potensiaal om al hoe meer studente se lewens te raak ongetwyfeld groter sal wees as wat aanvanklik daarmee beoog is. Dit blyk duidelik uit die Universiteit Stellenbosch se besluit om die toets

as deel van hul toegangstoetsbattery in te sluit, soos hierbo bespreek – hierdie besluit word egter tans heroorweeg in die lig van die nuutste verwickelinge rondom toelating op die hoëronderwyslandskap. Dít alleen behoort ’n openbare beskrywing van die konseptualisering van die toets te noodsaak, wat hierdie hoofstuk dan ook gepoog het om te doen, naamlik om die begin van hierdie toets se storie -’n geskiedenis van waar en hoe en teen watter konteks dit ontstaan het, en watter oorwegings die konseptualisering daarvan beïnvloed het – te vertel.

In die volgende hoofstuk word daar gefokus op die geïmpliseerde spesifikasies in bostaande konstruk wat geoperasionaliseer word as toepaslike toetstaaktipes wat die akademiese geletterdheidsvlakke van studente moet meet. Sodanige taaktipes funksioneer egter nooit afsonderlik van ’n begrip van die sogenaamde fundamentele oorwegings by toetsontwerp nie wat in die daaropvolgende hoofstukke aan die orde kom. Die bespreking sal, in die hoofstukke wat volg, telkens dus weer terugverwys na spesifieke komponente van die filosofiese raamwerk wat in Hoofstuk 2 uitgestippel is.

Op soek na die toepaslike maatstaf: van konstruk na toetsspesifikasie na taaktipe

4.1 Inleidend

Een van die uitdagings in die proses van toetsontwerp is die verkryging van artikulasie of belyning tussen toetskonstruk, eksperimentering met taaktipes en 'n uiteindelijke stel toetsspesifikasies wat hieruit voortvloei. Hierdie tegniese belyning of harmonisering van komponente van die meetinstrument maak konseptueel staat op die eggo van die estetiese dimensie van ons ervaring in die leidende funksie van die toets as tegniese instrument, en is streng gesproke aan die een kant dus 'n regulatiewe voorwaarde vir die verantwoordelike ontwerp van 'n taaltoets, wat dien as illustrasie vir hoe konstitutiewe en regulatiewe elemente en voorwaardes met mekaar vervleg is.

Aan die ander kant, val die klem in hierdie hoofstuk op die tegniese operasionalisering van die teoreties-gedefinieerde vermoë wat gemeet moet word. As begrip verwys die tegniese en fisiese dimensies van ons ervaring na die werking van 'n tegniese proses. Tegniese operasionalisering is dus op sy beurt 'n konstitutiewe voorwaarde vir toetsontwikkeling. Hierdie hoofstuk fokus derhalwe op die geïmpliseerde spesifikasies van die konstruk, wat in Hoofstuk 3 voorgehou is, deur spesifieke aandag te vestig op die operasionalisering van die genoemde

toetskonstruk en toetsspesifikasies in die vorm van toepaslike toetsitems wat die akademiese geletterdheidsvlakke van studente moet meet.

Al die afdelings van TAG sowel as 'n aantal taaktipes word daarom ondersoek om sekerheid oor die belyningskwessie te kry. Daar sal in hierdie verband in die besonder op die herkonseptualisering van 'n spesifieke taaktipe, naamlik Cloze/Sinluiting, ingegaan word wat tradisioneel as 'n goeie en betroubare taaktipe gesien word, maar wat weens die logistieke beperkinge van TAG, veral die feit dat dit 'n toets is wat slegs uit meervoudige keuse items kon bestaan, aangepas moes word.

Hierdie hoofstuk is derhalwe ook 'n goeie illustrasie van die punt wat in Hoofstuk 2 (verwys 2.3) gemaak is, naamlik dat die toegepaste linguistiek 'n dissipline van ontwerp is. In hierdie opsig sal daar hier ingespeel word op die herontwerp van 'n tradisioneel goeie taaktipe, maar een wat weens bepaalde beperkinge, soos reeds hierbo aangedui, aangepas moes word om uiteindelik 'n bruikbare, betroubare en steeds geldige oplossing vir die probleem te bied. Hierdie hoofstuk is dus nie net 'n illustrasie van hoe die leidende of kwalifiserende funksie van 'n ontwerp, as 'n toegepaste linguistiese oplossing, te vinde is in die tegniese aspek daarvan nie, maar ook 'n aanduiding van hoe die ontwerp bestudeer, ondersoek, verklaar en selfs verander kan word vanuit die analitiese perspektief, wat as rasionaal daarvoor aangebied kan word.

4.2 Die konteks

In Hoofstuk 1 is die rasionaal vir die ontwerp van 'n nuwe konstruk vir 'n toets van akademiese geletterdheid vir gebruik, aanvanklik slegs aan die Universiteit van Pretoria, uiteengesit. Hoofstuk 3 het daarop voortgebou deur onder andere 'n beskrywing te bied van hoe die nuwe konstruk tot stand gekom het. Daar is ook aangevoer dat 'n toetsontwerp vir hierdie konstruk nodig was wat eerstens as bloudruk daarvoor sou kon dien, maar wat ook tweedens die logistiese beperkinge wat veral met die administrasie van die toets in 'n beperkte tyd aan groot getalle

studente die hoof sou bied deur 'n bruikbare produk te ontwikkel, sonder om die betroubaarheid en geldigheid daarvan te kompromitteer. In hierdie verband is dit belangrik om daarop te let dat die doel van die toets aanvanklik slegs was om plasing van studente op verskillende taalkursusse te fasiliteer, maar dit het gou, soos reeds vermeld, ook 'n hoërbelangfunksie begin vervul, naamlik dat dit as deel van die toegangstoetsbattery aan die Universiteit Stellenbosch gebruik sou word. Dit alles bring mee dat daar met selfs groter omsigtigheid met die toetsontwerp omgegaan moes word om te verseker dat letterlik duisende uitslae steeds binne 'n tydperk van 24 tot 48 uur beskikbaar gestel kan word ten einde ingeligte besluite oor plasing en selfs toelating te kan neem. Die etiese dimensie van toetsontwerp kom dus by implikasie ook weer hier ter sprake, alhoewel dit, soos reeds in Hoofstuk 2 aangedui, nie die primêre fokus van hierdie studie is nie.

Die grootte van die toetspopulasie en die dringendheid van die beskikbaarstelling van uitslae het die besluit op 'n volledig meervoudige keuse toets (wat elektronies nagesien kon word) 'n vanselfsprekende keuse gemaak. Die essensiële aard van die gebruik van meervoudige keuse items vir TAG het die toetsontwerpers ook gedwing om die kritiek wat teen hierdie tipe toetse uitgespreek word te heroorweeg, sonder om op die versigtigheid waarmee toetsitems normaalweg geskep word, in te boet. Dit het trouens daartoe gelei dat daar op ondernemende en kreatiewe wyse met die skep van toetsitems omgegaan is, eerder as om van meet af aan toe te gee aan die vooroordeel dat hierdie of daardie aspek van taal nie in meervoudige keuse formaat getoets kan word nie (Van Dyk & Weideman, 2004b:16). Die vindingrykheid waarmee sekere toetsitems geskep is, het trouens net gedemonstreer dat bevooroordeeldheid presies net dit is, naamlik vooroordele wat die verantwoordelike, professionele taaltoetsers se tegniese verbeelding inperk en verhinder éérder as om aan te moedig om na kreatiewe oplossings vir probleme te soek. Ter illustrasie word die volgende vrae uit TAG 2007 aangebied:

In 'n teks getiteld *GM-Voedsel* met subtitel "*Frankenstein-kos*" of "*manna*"?, is die vertolking van metafore byvoorbeeld as volg getoets:

Die subtitel van die artikel verwys na twee

- A. probleemassekte van GM-voedsel in Afrika.
- B. probleemareas waarmee biotegnoloë worstel.
- C. alternatiewe oplossings vir wêreldverhongering.
- D. opponerende sienings oor die rol van GM-voedsel.

In dieselfde teks is stylfigure op die volgende manier getoets:

Watter woordpaar illustreer duidelik die ironiese teenstrydigheid in die houding van die teenstanders van biotegnologie (paragraaf 4) ?

- A. enersyds > andersyds
- B. Chihuahua > wolf
- C. nekhare > drogbeelde
- D. houtgerus > afgryse

Begrip van die onderskeid tussen oorsaak en gevolg, asook vir die verhouding tussen verskillende teksdele / taalgebruik wat teksdele samehangend aan mekaar bind, is weer só getoets:

Watter woordpaar stel NIE 'n oorsaak-gevolgverhouding binne die konteks van die artikel nie?

- A. erosie > minder landbougrond
- B. GM-voedsel > oorlewing
- C. wêreldaanwas > laer oesopbrengste
- D. min industrialisering > biodiversiteit

Dit behoort reeds uit hierdie enkele voorbeelde te blyk dat die vaardighede wat van studente verwag word wanneer hulle TAG aflê, van so 'n aard is dat dit veel meer insluit as 'n eng, of talig beperkende, definisie wat akademiese geletterdheid tot 'n stel vaardighede beperk of dit inkort tot die bemeestering van klank, vorm en betekenis (vergelyk 1.4.1 en 3.4).

In Hoofstuk 3 (vergelyk die aanloop tot Tabel 3.1) is verskeie redes aangevoer waarom 'n oop of verruimde siening van taal noodsaaklik is vir die basis van 'n toets van akademiese geletterdheid. Een daarvan is dat toetsitems groter ooreenstemming met die werklike taalgebruiksituasie moet toon. Die drie uittreksels hierbo is dus nie net illustrasies van hoe komponente van die konstruk verbeeldingryk getoets kan word nie, maar is ook voorbeelde van die werklike taalgebruiksituasie – teks en konsepte, asook die maak van onderskeide en gepaardgaande kognitiewe bewerkings - waarmee studente daaglik gekonfronteer word. Trouens, wanneer die konstruk van TAG en verbandhoudende taaktipes

(soos ook in Hoofstuk 3 bespreek), by kongresse en in publikasies voorgehou word, is daar telkens resonansie met die ervaring en indrukke van kollegas oor wat 'n akademies geletterde persoon tipies behoort te kan doen.

4.3 Taaktipes

Die uiteindelijke seleksie van taaktipes en items vir 'n bepaalde toets is volgens Van Dyk en Weideman (2004b:17) nie net afhanklik van die belyning van toetskonstruk met toetsspesifikasies nie, maar ook van die oordeel van die toetsontwerper in terme van die produktiwiteit daarvan. Binne die konteks van die model van die toegepaste linguistiek as dissipline van ontwerp, wat in Hoofstuk 2 voorgehou is, maak hierdie opmerking sin, aangesien daar, in die meeste gevalle retrospektief, geoordeel moet word of items bruikbaar is al dan nie en toetsspesifikasies word dienooreenkomstig aangepas. Toetsontwikkelaars moet derhalwe besluite oor items neem na gelang van byvoorbeeld hul kwalitatiewe ervaring met tipes take, of die geantisepeerde waarde wat aan bepaalde, dikwels nuutgeskepte, items geheg word, of selfs die kwantitatiewe waarde (ook retrospektief) van taaktipes. Van Dyk en Weideman (2004b:18) voer in hierdie verband aan dat taaktipes in terme van hul beoordeling in vier kategorieë verdeel kan word, naamlik *aanvaarbaar* ('n hoë vlak van belyning met die toetskonstruk, maar oënskynlik nie produktief nie), *onaanvaarbaar* (lae produktiwiteit en 'n lae vlak van belyning met die toetskonstruk), *wenslik* (produktief sowel as 'n hoë vlak van belyning met die toetskonstruk), en *nie ideaal nie* (potensieel produktief, maar nie goeie belyning met die toetskonstruk nie). Figuur 4.1 is 'n voorstelling van hierdie kategorisering (Van Dyk & Weideman, 2004b:18).

belyning	hoog	<i>aanvaarbaar</i>	<i>wenslik</i>
	laag	<i>onaanvaarbaar</i>	<i>nie ideaal nie</i>
		laag	hoog

produktiwiteit

Figuur 4.1: Matriks vir beoordeling van toetsitems

Van Dyk en Weideman (2004b:18) wys egter daarop dat indien daar geoordeel word dat 'n item nie ideaal is vir gebruik nie, dit nie onmiddellik van die databank van items verwyder hoef te word nie, aangesien daar geleenthede kan opduik waar die bepaalde item aangepas of herkonseptualiseer kan word om byvoorbeeld 'n hoër vlak van belyning met die konstruk of die toetsspesifikasies te toon. In dieselfde opsig kan 'n item wat as aanvaarbaar gekategoriseer is, meer produktief gemaak word deur van loodstoetse gebruik te maak en telkens aanpassings aan byvoorbeeld die afleiers van die bepaalde item te maak, sodat dit uiteindelik as wenslik gekategoriseer kan word, met ander woorde beide produktief én belyn met die konstruk.

Ter illustrasie van bostaande onderskeidings word daar vervolgens gekyk na enkele taaktipes waarmee daar al in die ontwikkelingsproses van TAG geëksperimenteer is, naamlik:

- Skommelteks ('n paragraaf waarvan die sinsvolgorde verander is)
- Register en tekstipe ('n taak wat van kandidate verwag om taaluitinge in een lys met ander in 'n tweede lys af te paar in terme van die register/styl/tekstipe)
- Woordeskatkennis (gekontekstualiseer deur middel van kort sinne en gebaseer op internasionaal aanvaarbare woordeskatlyste, soos die van Coxhead [2000] en Nation [2001] – beide hierdie lyste word gereeld opgedateer en die nuutste weergawes is op die internet beskikbaar; vir doeleindes van TAG word die nuutste lys gewoonlik van die internet gehaal, indien nodig).
- Woordeboekdefinisies ('n afgeleide vorm van woordeskatkennis, maar gebaseer op outentieke woordeboekdefinisies van algemene akademiese woordeskat)
- Fouterkenning (spesifiek anglisismes)
- Interpretasie van grafiese en visuele inligting (vrae oor tendense; eenvoudige numeriese berekeninge; skattings)

- Langer tekste (leesbegrip – ’n verskeidenheid van akademiese taalvaardighede)
- Akademiese skryftake (spesifiek daardie tipe skryftake wat klassifikasie en vergelyking vereis)
- Sinsluiting/Cloze (die herstel van ’n sistematies-gemutileerde teks)
- C-prosedure (’n afgeleide vorm van sinsluiting, waar die helfte of net meer as die helfte van ongeveer elke tweede woord verwyder is)

Van al hierdie taaktipes val net fouterkenning in die sogenaamd onaanvaarbare kwadrant; waarskynlik omdat sogenaamde anglisismes eerder problematies is binne konteks van algemene taalvermoë as in akademiese taalvermoë. Al die ander taaktipes belyn op die een of ander manier met die konstruk, en het minstens die potensiaal om met verloop van tyd ook vrugbaar gebruik te kan word.

Wanneer die oorblywende taaktipes (al die items waarmee geëksperimenteer is minus die onaanvaarbare taaktipe) vergelyk word met die oorspronklike sub-komponente van die konstruk, ontvou ’n beeld soos in Tabel 4.1 vervat:

Tabel 4.1: Vergelyking van potensieel bruikbare taaktipes met konstruk

Sub-komponente van konstruk	Taaktipes
Akademiese woordeskat	Woordeskatkennis Woordeboekdefinisies Sinsluiting C-prosedure
Metaforiese taalgebruik	Teksbegrip
Tekstualiteit – kohesie en teksgrammatika	Skommelteks Sinsluiting C-prosedure Teksbegrip Akademiese skryftake

Tabel 4.1 (vervolg...)

Tekstipe, styl en register	Register en tekstipe Interpretasie van die grafiese en visuele Skommelteks Sinsluiting C-prosedure Teksbegrip Akademiese skryftake
Interpretasie van visuele en grafiese inligting	Interpretasie van die grafiese en visuele Klassifikasie, toepassing en berekening
Onderskeid tussen die essensiële en die nie-essensiële	Teksbegrip Interpretasie van grafiese en visuele Akademiese skryftake
Klassifikasie, vergelyking en berekeninge	Interpretasie van die grafiese en visuele Teksbegrip
Afleiding en toepassing	Teksbegrip Akademiese skryftake Woordeskatkennis
Kommunikatiewe funksie: definiëring, beredenering	Teksbegrip Sinsluiting Skommelteks
Betekeniswaarde groter as sinsvlak	Teksbegrip Register en tekstipe Skommelteks Interpretasie van grafiese en visuele

Uit 'n elementêre klassifikasie soos in Tabel 4.1 hierbo, was dit gou duidelik dat sekere taaktipes hul eerder daartoe leen om akademiese geletterdheid te meet as ander. Die vraag wat gevolglik ontstaan het en wat Van Dyk en Weideman (2004b:19-20) verder ondersoek het, was waarom daar dan nie net op daardie taaktipes gefokus word wat beter met die konstruk belyn en produktief blyk te wees nie, met die oënskynlike verwagting dat daar dan met 'n meer vaartbelynde toets vorendag gekom kan word.

Die antwoord op hierdie vraag is kompleks en open weer die debat van of dit wenslik is om te poog om 'n enkele aanduiding, of 'n verskeidenheid van aanduidings, van taalvermoë te hê, soos wat die geval was met die totstandkoming van geïntegreerde toetsing (vergelyk in hierdie verband veral ook punt 3.3.3 in die vorige hoofstuk en die bespreking daar). Die voor die hand liggende redenasie is

gewoonlik dat 'n beperkte aantal afdelings of sogenaamde sub-toetse ideaal vanuit die toetsontwikkelaar se perspektief is, maar dan moet dit ook in ag geneem word dat 'n toets gewoonlik 'n hoër betroubaarheidswaarde het as daar meer sub-toetse en derhalwe vrae in 'n toets is. Daar is 'n verdere, indirekte rede vir die ontwerp van 'n toets as bestaande uit 'n verskeidenheid sub-toetse, soos wat daar gesien sal kan word in Hoofstukke 6 en 7 hieronder, wanneer die validering van TAG in meer besonderhede bespreek word. Dit is naamlik dat die vergelyking van die relatiewe bydraes van elke sub-toets (en, daarmee saam, die komponent of komponente van die konstruk wat deur die sub-toets gemeet word) tot die geheel gemakliker bepaal kan word. Daar word egter later weer hierop teruggekom.

Vir doeleindes van TAG is daar besluit om nie net op daardie taaktipes te fokus wat klaarblyklik hoë vlakke van produktiwiteit en belyning met die konstruk getoon het nie – praktiese oorwegings was die hoofrede hiervoor. Terwyl langer tekste byvoorbeeld meer wenslik is om te gebruik as korter tekste in die meting van iets soos akademiese geletterdheid, omdat die relevansie daarvan tot taalgebruik in die domein wat gemeet word, naamlik akademiese diskoers, onmiddellik patent is, is dit ook waar dat kandidate heelwat meer tyd bestee aan die beantwoording van vrae wat op langer tekste gebaseer is. In 'n toets wat aan streng tydsbeperkinge moes voldoen (soos in die toetsspesifikasies duidelik sal blyk), soos reeds elders bespreek, was dit nodig om ook ondersoek in te stel na ander, minder tydrowende, maar steeds produktiewe en belynde taaktipes, sonder om op die waarde van langer tekste in te boet.

4.4 Taaktipes en ontwerpbeginsels

Een van die regulatiewe ontwerpbeginsels wat uit die teoretiese raamwerk van die toegepaste linguistiek, soos in Hoofstuk 2 voorgehou, voortvloei, kan juis afgelei word uit die verhouding tussen die leidinggewende tegniese ontwerpfunksie van 'n toets, en die ekonomiese modus van die ervaring van toetsontwikkelaars. Hierdie beginsel vind neerslag in die konsep van tegniese afweging (van nuttige of waardevolle en minder nuttige of minder waardevolle komponente). Omdat sulke

tegniese afweging ook die belyning en balans in die ontwerp ondersteun, word daar in die soeke na 'n tegniese balans dus ook 'n antisipasie van die tegniese op die estetiese ontdek – dit gaan hier dus oor 'n tegniese afweging van toetstipes wat nuttig en minder nuttig kan wees in die ontwerp, ten einde alles in die beste, teoreties-regverdigbare en mees verdedigbare balans te kry.

'n Balans tussen korter en langer afdelings moes derhalwe gevind word en daarom is daar verskeie afdelings in TAG opgeneem (verwys afdeling 4.6 vir 'n uiteensetting). Daar is egter nie met die langer leesteks weggedoen nie, maar die aantal vrae word beperk en die gewig per vraag vir hierdie afdeling is gewoonlik swaarder as by ander afdelings; hierdie afdeling bly die grootste enkele afdeling van die toets. Hoewel dit dus meer ruimte vir produktiewe en belynde items bied, word hierdie afdeling ingeperk en in kombinasie met ander gebruik.

Insgelyks moes daar oordele oor die opneem van 'n akademiese skryftaak in TAG gevel word. Wat aan die een kant gewen is deur uiteindelik daarmee weg te doen, omdat die verskeidenheid van itemtipes in meervoudige keuse formaat konstant goed gekorreleer het ($r = 0.85$, $p \leq 0.05$) met die totaal van die toets, is aan die anderkant weer ingeboet op gesigsgeldigheid. Om egter weer groter gesigsgeldigheid te bewerkstellig, is daar besluit om akademiese skryftake steeds in te sluit, maar dit slegs na te sien / in ag te neem by grensgevalle.

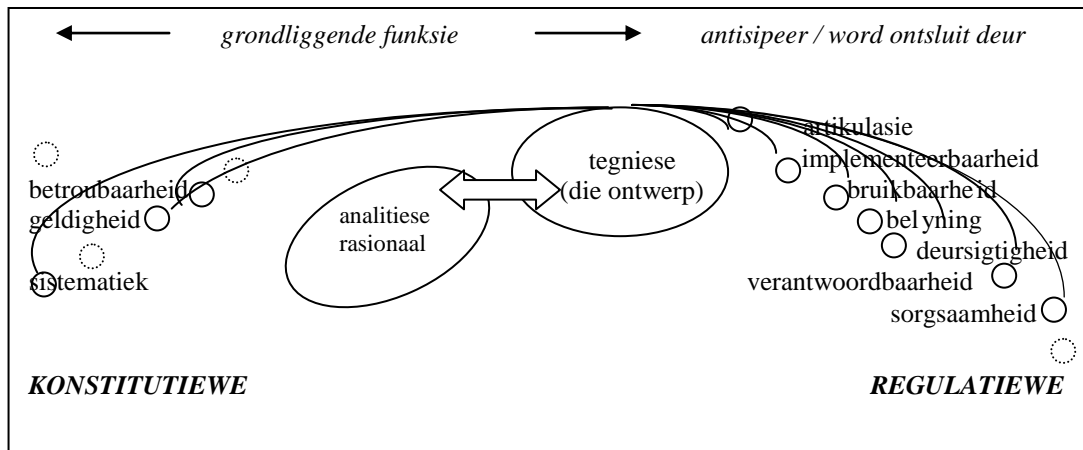
'n Verdere onbeantwoorde vraag, wat vir doeleindes van hierdie studie nie ter sprake sal kom nie en dus nie in besonderhede bespreek sal word nie, omdat dit nog nie empiries ondersoek is nie, is die kwessie van gewig per vraag of afdeling. Soos reeds in Tabel 4.1 hierbo gesien is, leen sekere item- / taaktipes hulle meer tot produktiwiteit en belyning met die konstruks as ander. Die premis waarmee daar dus gewerk kan word, is dat daar 'n swaarder gewig aan hierdie tipe vrae toegeken behoort te word. So byvoorbeeld kan daar geargumenteer word dat kennis van akademiese woordeskat en die vermoë om betekenis uit konteks af te lei belangrik is om in die akademiese omgewing te oorleef en daarom op hierdie vlak swaarder behoort te weeg as ander vrae wat byvoorbeeld begrip vir metaforiese taalgebruik in

akademiese skryfwerk meet. Nog 'n voorbeeld, en van kardinale belang, is die klem wat die vermoë om onderskeid tussen belangrike en minder belangrike inligting te tref in 'n toets behoort te hê teenoor iets soos die vermoë om te kan ekstrapoleer en te kan toepas, indien daar aanvaar word dat eersgenoemde laasgenoemde voorafgaan in akademiese redenering (Van Dyk & Weideman, 2004b:20). Gewigte word dus in TAG versprei volgens taaktipes wat in afdelings opgeneem is. Daardie itemtipes wat nie onder bepaalde afdelings pas nie, kom normaalweg egter wel in die leesbegripafdeling aan die bod, afhangend van die taak- en itemtipes waartoe die teks homself verleen.

Op hierdie punt blyk dit in orde te wees om bostaande bespreking eers weer binne konteks van die teoretiese raamwerk (soos in Hoofstuk 2 bespreek) wat vir hierdie studie gekies is, te plaas. Daar is in Hoofstuk 2 aangevoer dat daar 'n balans moet wees tussen die konstitutiewe en regulatiewe aard van 'n toets en dat daar derhalwe afweging sal moet plaasvind om uiteindelik 'n bruikbare, teoreties regverdigbare, deursigtige, betroubare en geldige toets (om maar enkele eienskappe van 'n goeie toets te noem) daar te stel. Daar moet dus vooruitskouend (antisipatories) en terugskouend te werk gegaan word wanneer oplossings vir taalprobleme ontwerp word, of soos Weideman (2009a:20) dit stel:

[i]n addition to having foundational concepts that relate the technical dimension to preceding aspects of experience, [the] guiding aspect of applied linguistic designs also anticipates, in the other direction, its linkages with the lingual, social, economic, aesthetic, juridical and ethical aspects that follow it. These forward-looking, anticipatory links between the technical, qualifying function of the test design and these other aspects yield the ideas of technical articulation, test implementation or use, its utility, its alignment with, for example, learning and teaching language, its public defensibility or accountability, and its fairness or care for those taking tests.

Artikulasie tussen die basisfunksies (van konstitutiewe aard) van 'n toets en die ontslote, regulatiewe aard daarvan is dus van belang en word die beste geïllustreer deur Figuur 4.2 (oorgeneem en verwerk uit Weideman, 2009a:20).



Figuur 4.2: Illustrasie van artikulasie tussen konstitutiewe en regulatiewe elemente

Bostaande kan ook anders verwoord word, naamlik deur 'n lys van sogenaamde ontwerpbeginsels saam te stel waarop toetsontwikkelaars behoort te let. So 'n lys sou die volgende terugskouende en antisipatoriese aspekte kon insluit:

- | | |
|---------------------|---|
| Terugskouend
{ | <ul style="list-style-type: none"> • Toetse behoort 'n teoretiese begroning te hê (konstruk). • Die betroubaarheidswaardes van die toets as geheel behoort aanvaarbaar te wees (volgens algemeen aanvaarde praktyke, 'n alpha-waarde van bo 0.7). • Toetse moet deurlopend en sistematies met verwysing na 'n veelheid datastelle en argumente geregverdig en gevalideer word, aangesien dit daardie persone wat die toetse aflê, affekteer. |
| Vooruitskouend
{ | <ul style="list-style-type: none"> • Daar moet artikulasie tussen konstruk, taaktipes, items en toetsspesifikasies wees. • Die toets behoort implementeerbaar te wees (min logistieke beperkinge). • Die toets moet bruikbaar wees. • Die toets(proses) moet deursigtig, legitiem en verdedigbaar wees. • Die toets(proses) moet verantwoordbaar wees in die openbare sfeer. |

Wat op hierdie stadium duidelik uit die bostaande behoort te blyk, is dat dit terugskouend nodig was om enkele taaktipes van TAG verder te ondersoek met die doel om dit meer produktief te maak en 'n beter belyning met die konstruk te verseker, en dit is presies ook wat die afgelope aantal jare gebeur het – daar is herhaaldelik met taaktipes geëksperimenteer in loodstoetse en dit is telkens verfyn tot op 'n aanvaarbare vlak van produktiwiteit en belyning met die konstruk. Die res van hierdie hoofstuk sal daarom ter illustrasie op een spesifieke taaktipe konsentreer, naamlik hoe daar geëksperimenteer is met 'n sinsluiting-taaktipe wat normaalweg produktief blyk te wees, maar wat in 'n meervoudige keuse formaat 'n deel van sy produktiwiteit verloor het.

4.5 Eksperimentering ter verbetering van taak- en itemtipes

Soos reeds genoem is die fokus van hierdie afdeling op 'n sinsluiting-taaktipe. Tabel 4.1 was ook 'n aanduiding daarvan dat sinsluitingstake normaalweg die potensiaal het om produktiewe taaktipes te wees wat goed met die konstruk belyn is. Die rede hiervoor is dat dit byvoorbeeld die vlak van hantering van woordeskat, grammatika, samehang, begrip vir tekstipe en potensieel selfs die kommunikatiewe funksies van taal kan meet.

Aanvanklike toetse met die sinsluiting-taaktipe wat in meervoudige keuse formaat gemeet is weens die logistieke beperkinge van TAG, soos reeds elders vermeld, het egter teleurstellende resultate opgelewer. Van vyf items wat in Maart 2003 op meer as 1000 toetslinge in 'n loodstoets gebruik is, het slegs een item binne die statistiese parameters wat vir TAG en TALL gestel is, geval – 'n diskriminasiewaarde (die graad waartoe 'n item tussen sterker en swakker kandidate diskrimineer) van meer as 0.3 en 'n fasiliteitswaarde (aanduiding van moeilikheidsgraad) van tussen 0.3 en 0.7. Slegs toe die parameters verstel is na 'n diskriminasiewaarde van 0.26 en na 'n fasiliteitswaarde van tussen 0.25 en 0.84, was drie van die vyf items wenslik vir gebruik, hoewel nog nie ideaal nie.

Uit die aard van die saak kon die genoemde waardes verbeter / binne die grense van die parameters val deur byvoorbeeld aan die afleiers te verander. Met die kundigheid wat betrokkenheid by ander toetsontwikkelingsgeleenthede (spesifiek die AARP-toetse wat in Hoofstuk 3 ook ter sprake was) egter opgelewer het, is daar besluit om nie hierdie roete te volg nie, maar eerder aan die formaat van die vraag te verander. Vervolgens is 'n aangepaste vorm van hierdie taaktipe opgestel wat nie meer net van kandidate verwag om 'n keuse uit te oefen van elke n'de woord wat weggelaat is nie, maar ook dat hulle moet aandui op watter plek die woord weggelaat is, wat uit die aard van die saak 'n hoërorde kognitiewe handeling vereis en dus potensieel 'n beter meting van akademiese geletterdheid sou gee.

Die res van hierdie hoofstuk is 'n bespreking van die herhaalde aanpassings (en gevolge daarvan) van die genoemde sinsluitingstaak en die gepaardgaande rubriek. Die eerste voorbeeld uit TAG 2004 is op die volgende bladsy:

Afdeling 7: Teksredigering

Sommige woorde is uitgelaat in die onderstaande teks. Kies eers die **plek** waar die woord weggelaat is, en daarna die **woord**. Die korrekte antwoorde vir die eerste twee word vir u gegee, en is in vetdruk gemerk in die tweede en derde kolomme. Byvoorbeeld, by die eerste een merk (b) die **plek** van die woord wat weg is, en (c) is die regte **woord**.

Bier en glase

TEKS	PLEK	WOORD
In ^(a) /België ^(b) / feitlik ^(c) /elke ^(d) /bier uit sy	(a)/ (b) / (c)/ (d)/	(a) is (b) kan (c) word (d) gaan
^(a) / eie, ^(b) / glas ^(c) /gedrink en ^(d) /geniet.	(a)/ (b) / (c)/ (d)/	(a) kenmerkende (b) grootte (c) persoonlike (d) word
In die ^{58&59} (a)/ geval ^(b) / van ^(c) / De Coninck ^(d) / die	58. (a)/ (b)/ (c)/ (d)/	59. (a) eerste (b) bier (c) gaan (d) wees
^{60&61} (a)/ drinkery ^(b) / boonop met 'n ^(c) / eie ^(d) / gepaard. ^{62&63} Oor ^(a) / die	60. (a)/ (b)/ (c)/ (d)/	61. (a) vlytig (b) ritueel (c) benewens (d) aardig
^(b) / jare heen ^(c) / is De Coninck ^(d) / ^{64&65}	62. (a)/ (b)/ (c)/ (d)/	62. (a)/ (b)/ (c)/ (d)/
^(a) /verskillende ^(b) / glase geskink, ^(c) / waarvan die twee ^(d) / die bolglas en die fluit is.	64. (a)/ (b)/ (c)/ (d)/	63. (a) baie (b) in (c) is (d) vir
^{66&67} (a)/ glas ^(b) / het sy ^(c) / eie ^(d) / tradisie. ^{68&69} (a)/ die	66. (a)/ (b)/ (c)/ (d)/	64. (a)/ (b)/ (c)/ (d)/
fyner ^(b) / van ^(c) / 25 cℓ ^(d) / word as ^{70&71} (a)/ slegs ^(b) / vir ^(c) / vroue ^(d) / beskou. In	68. (a)/ (b)/ (c)/ (d)/	65. (a) bekendstes (b) vernaamstes (c) belangrikstes (d) oulikstes
Vlaandere bestel 'n mens dus 'n "fluitjie", 'n woord met die konnotasie "fallus".	70. (a)/ (b)/ (c)/ (d)/	66. (a) hierdie (b) persoonlik (c) elke (d) vetterige 67. (a) in (b) fluitglas (c) bolglas (d) deur 68. (a) jonger (b) bekyk (c) geskik (d) vaardige

Die aanloop tot hierdie formaat het heelwat ondersoekstyd in beslag geneem en daar is begin met 'n eksperiment onder Engelse studente, waar TALL vir doeleindes van die ondersoek gebruik is (Van Dyk & Weideman, 2004b:21). Vir hierdie eksperiment is 'n soortgelyke, gemutileerde teks as die voorbeeld hierbo gebruik, maar wel in Engels. 'n Toets bestaande uit 50 van hierdie tipe items is ontwikkel en die resultate hiervan was, in teenstelling met vorige sinsluitingspogings, bevredigend vanuit verskeie hoeke. Toetsontwikkelaars is eerstens gedurig opsoek na subtoetse of afdelings wat 'n betroubaarheidswaarde (sogenaamde alpha-waarde, wat aangedui word deur die simbool: α) het wat hoër is as die van ander subtoetse – dit is normaalweg 'n aanduiding dat 'n bepaalde taaktipe beter diskrimineer as ander en dat dit dus 'n beter en meer produktiewe tipe taak is om by 'n toets in te sluit. Die oorspronklike toets met 50 items is op 'n groep sogenaamde akademies sterk studente ($n=43$) gebruik en het inderdaad 'n hoër betroubaarheidswaarde ($\alpha=0.749$) getoon as 'n vorige loodstoets wat ook 50 items bevat het, maar wat oor sewe taaktipes versprei was. Let egter daarop dat onderafdelings nooit die toets as geheel kan ewenaar nie en dit is nie wat hier gesuggereer word nie.

Sommige van die items is na afloop van die loodstoets verwyder – daardie items wat nie binne die parameters wat vir TAG en TALL gestel is, geval het nie (dus 'n te lae diskriminasiewaarde en / of 'n te lae of 'n te hoë moeilikheidsgraad). Ander items is weer aangepas deur byvoorbeeld aan die afleiers te verander of dit rond te skuif, indien dit nodig geblyk te wees het. 'n Volgende poging met 30 items en basies dieselfde teks is gevolglik gebruik in 'n tweede loodstoets wat op twee groepe eerstejaarstudente gebruik is. Die een groep was aangemerkt as akademies sterk ($n=23$) en die ander akademies swak ($n=21$) in terme van akademiese geletterdheidsvlakke.

Die resultate op hierdie tweede loodstoets was selfs meer bevredigend as die eerste. Tabel 4.2 is 'n vergelyking van die gemiddelde prestasie, die betroubaarheid, die gemiddelde moeilikheidsgraad, en die item-totaal-korrelasie wat 'n aanduiding van

die gemiddelde diskriminasiewaarde van die toets is, van die twee genoemde loodstoetse wat op drie groepe studente op die proef gestel is.

Tabel 4.2: Vergelyking van toetsgemiddeld, betroubaarheid, gemiddelde moeilikheidsgraad en gemiddelde diskriminasiewaarde vir 'n eerste en tweede loodsfase van sinsluitingsubtoets

Weergawe	Vermoë	Gemiddeld	Alpha	Gemiddelde moeilikheidsgraad	Gemiddelde diskriminasie
<i>Eerste</i>	bo gemiddeld	64.42%	0.75	64%	0.23
<i>Tweede</i>	bo gemiddeld	66.96%	0.82	67%	0.36
<i>Tweede</i>	onder gemiddeld	58.57%	0.81	59%	0.33

Uit bostaande tabel blyk dit duidelik dat veral die betroubaarheidswaarde van die een toetsgeleentheid na die volgende 'n goeie toename getoon het. Dit dui verder daarop dat hierdie vorm van sinsluiting 'n taaktipe is wat potensiaal toon en wat verder ondersoek kon word. Verdere eksperimente in Engels én Afrikaans, waar hierdie taaktipe deel van 'n langer toets van akademiese geletterdheid is, het aangetoon dat dit beslis belofte inhou: dit is op sogenaamde swak eerstejaarstudente (aldus die akademiese geletterdheidstoets wat vroeër genoem is: PTEEP) getoets en het die resultate, soos vervat in Tabel 4.3, opgelewer. Wat egter hier in ag geneem moet word, is dat daar 'n reeds onderskeid getref is tussen studente met bo gemiddelde en studente met onder gemiddelde vermoëns en daar is dus in effek reeds vir 'n bron van diskriminasie gekontroleer. Beide Tabelle 4.3 en 4.4 is dus slegs ter illustrasie van die argument gebruik.

Tabel 4.3: Vergelyking van toetsgemiddeld, betroubaarheid, gemiddelde moeilikheidsgraad en gemiddelde diskriminasiewaarde vir 'n derde en vierde loodsfase van sinsluitingsubtoets

Weergawe	Vermoë	Gemiddeld	Alpha	Gemiddelde moeilikheidsgraad	Gemiddelde diskriminasie
<i>Derde (Engels)</i>	onder gemiddeld	48.34%	0.86	48%	0.35
<i>Vierde (Afrikaans)</i>	onder gemiddeld	70.19%	0.83	70%	0.33

Toe TAG en TALL vir die eerste keer in 2004 op 'n groep van 6247 eerstejaarstudente met gemiddelde akademiese vermoëns (die totale eerstejaarspopulasie; dus sterk én swak studente) gebruik is, het hierdie taaktipe selfs nog beter resultate getoon, en het dit bevestig wat met die afneem van die loodstoetse bevind is, soos aangedui in Tabel 4.4:

Tabel 4.4: Vergelyking van toetsgemiddeld, betroubaarheid, gemiddelde moeilikheidsgraad en gemiddelde diskriminasiewaarde van sinsluitingssubtoets van TAG en TALL 2004

Weergawe	Vermoë	Gemiddeld	Alpha	Gemiddelde moeilikheidsgraad	Gemiddelde diskriminasie
Engels (TALL)	gemiddeld	61.94%	0.92	62%	0.66
Afrikaans (TAG)	gemiddeld	58.19%	0.88	58%	0.53

Soos met enige eksperimentele benadering het hierdie aangepaste sinsluitingstaaktipe egter steeds bepaalde probleme, veral in terme van die gesigsgeldigheid daarvan – anekdotiese getuienis deur taaldosente en studente dui daarop dat dit oor die algemeen nie as gebruikersvriendelik ervaar word nie, desnieteenstaande die goeie statistiese resultate daarvan. In 'n poging om dit meer gebruikersvriendelik te maak en die gesigsgeldigheid daarvan te verhoog, is die rubriek herskryf deur tussenstappe in te bou (sogenaamde 'scaffolding'). Hieronder en op die volgende bladsy volg 'n voorbeeld uit TAG 2007:

Afdeling 6: Teksredigering

Hierdie afdeling bestaan uit drie dele. Merk u keuse vir elke vraag op die los antwoordblad.

Deel I

Sommige woorde is uit die onderstaande teks weggelaat. Die teks verskyn in die linkerkantste kolom. U moet die **plek** (a, b, c, d) kies waar die woord weggelaat is. Die korrekte antwoord vir die eerste een word vir u as voorbeeld gegee en is duidelik gemerk in die regterkantste kolom.

Teks	Woord
Weervoorspellings (a)/word gegrond op die (b)/ van die beweging en (c)/groei van stelsels (d)/van lae- en hoëdruk wat siklone en antisiklone genoem word.	(A)/ (B) / (C)/ (D)/
49. Dit is ontsaglike (a)/groot lugdraaikolke (b)/honderde	49. (A)/ (B)/ (C)/ (D)/

tot (c)/duisende kilometers in (d)/wat voortdurend aan die beweeg en verander is. Die siklone, laedrukstelsels, waarvan die sirkulasie kloksgewys in die Suidelike halfrond is, word gekenmerk deur stygende lugstrome wat wolke en reën veroorsaak. 50. In (a)/die antisiklone, hoëdrukstelsels, (b)/is die sirkulasie (c)/met (d)/dalende lugstrome wat wolke laat verdamp en droë weer veroorsaak.

Deel II

Sommige woorde is uit die onderstaande teks weggelaat. Die teks verskyn in die linkerkantste kolom. Die plek waar die woord weggelaat is, word in die teks aangedui met 'n skuinsstreep (/). U moet die **woord** kies wat op daardie plek weggelaat is uit die moontlikhede wat in die tweede kolom gegee word. Die korrekte antwoord vir die eerste een word vir u as voorbeeld gegee, en is duidelik gemerk in die regterkantste kolom.

Teks	Plek
Die kern / Suid-Afrika se weardiens is die Weerburo 51/ onder 'n staatsdepartement funksioneer. Hiervandaan gaan inligting 52/ wat aan die boer, stedeling, reisiger, internasionale vlieëniërs en aan skepe in die Indiese en Atlantiese oseaan sê wat hulle te 53/ kan wees: sonskyn, reën, oorstromings, hitte, koue of storms.	(A)/ op (B)/ by (C)/ deur (D)/ van 51. (A)/watter (B)/waarvan (C)/waar (D)/wat 52. (A)/verby (B)/op (C)/uit (D)/in 53. (A)/wyte (B)/wagte (C)/wete (D)/werke

Deel III

Sommige woorde is uit die onderstaande teks weggelaat. Die teks verskyn in die linkerkantste kolom. Kies eers die **plek** waar die woord weggelaat is, en daarna die **woord**. Die korrekte antwoord vir die eerste een word weer vir u as voorbeeld gegee en is duidelik gemerk in die middelste en regterkantste kolomme.

Teks	Plek	Woord
i&ii By die (A)/word (B)/lugdruk (C)/kwikbarometers (D)/gemeet.	i. (A)/ (B)/ (C)/ (D)/	ii. (A)/werk (B)/winkel (C)/wind (D)/weerstasies
Lugtemperatuur 54&55 (A)/deur (B)/termometers, (C)/windsnelhede en (D)/rigting deur anemometers, 56&57 (A)/lugvogtigheid (B)/higrograaf, (C)/en nat en (D)/droë lug deur boltermometers gemeet. 58&59 So (A)/meet (B)/hulle (C)/ook wolktipes (D)/wolkhoogtes visueel 60&61 (A)/of (B)/wolkhoogtemeters, reënval (C)/reënmeters, (D)/sonskyn 62&63 met (A)/glassonskynmeters (B)/sonstraling (C)/met (D)/sol-rimeters.	54. (A)/ (B)/ (C)/ (D)/ 56. (A)/ (B)/ (C)/ (D)/ 58. (A)/ (B)/ (C)/ (D)/ 60. (A)/ (B)/ (C)/ (D)/ 62. (A)/ (B)/ (C)/ (D)/	55. (A)/sal (B)/is (C)/was (D)/word 57. (A)/sekere (B)'/n (C)/daardie (D)/een 59. (A)/en (B)/in (C)/op (D)/by 61. (A)/aan (B)/gee (C)/met (D)/deur 63. (A)/of (B)/en (C)/met (D)/op

Hoewel bostaande poging statisties gesproke selfs nog beter blyk te gewerk het (vergeelyk Tabel 4.5), was daar steeds ongemak onder veral taaldosente oor die dekoderingsvaardighede waaroor studente moet beskik om hierdie afdeling sogenaamd te kan ontsyfer. Die gevoel was dat dit tot die nadeel van studente kan wees, hoewel die teenargument, naamlik dat dekoderingsvaardighede waarskynlik ook 'n eienskap van akademiese geletterdheid is, binne hierdie konteks nie geredelike instemming van die dosente gekry het nie.

Dekoderingsvaardighede word soms as onderdeel van intelligensie geag en dan kan daar met die aanname gewerk word dat studente met bo gemiddelde intelligensie die vaardigheid vinniger as studente met onder gemiddelde intelligensie sal aanleer en daarom beter op 'n subafdeling soos die een hier ter sprake sal presteer. Myns insiens is intelligensie egter iets waaroor 'n mens in 'n mindere of 'n meerdere mate beskik, iets wat nie aangeleer kan word nie, maar slegs ontwikkel en verfyn kan word.

Tabel 4.5: Vergelyking van toetsgemiddeld, betroubaarheid, gemiddelde moeilikheidsgraad en gemiddelde diskriminasiewaarde van sinsluitingsubtoets vir TAG 2007

Weergawe	Vermoë	Gemiddeld	Alpha	Gemiddelde moeilikheidsgraad	Gemiddelde diskriminasie
<i>Afrikaans (TAG)</i>	gemiddeld	57.31%	0.89	57%	0.57

Daar is derhalwe besluit om die uitleg van die taaktipe weereens aan te pas en dit word vir die eerste keer in TAG en TALL 2009 opgeneem. 'n Voorbeeld uit 'n loodstoets word op die volgende aantal bladsye aangebied:

Afdeling 6: Teksredigering

Hieronder volg 'n stukkie teks oor blogs wat verwerk is uit *Die By* ('n bylae tot *Beeld*) van 2 Junie 2007, p. 8. In hierdie teks is daar enkele woorde weggelaat. Lees dit eers om 'n geheelbeeld van die teks te vorm voordat u enige van die vrae aanpak.

Net soos Yahoo en Google sou die woord “blog” ’n paar dekades gelede soos babataal geklink het. Nou is dit dié taal van die jonger generasie. Wêreldwyd daar meer as 100 miljoen blogs en die syfer groei teen 120 000 blogs per dag. Hierdie stelling gemaak deur die blogsoekenjin *Technorati* wat die ontwikkeling van blogs en ander gebruikersgedrewe inhoud die web volg. *Technorati* sê verder dat 1,4 blogs elke sekonde van elke dag geskep word. Die blogs wêreldwyd het in minder as ’n jaar ’n geweldige toename getoon en het glo verdubbel van 35 na 70 miljoen. Die blogosfeer, die wêreldwye blog-gemeenskap genoem word, is sterk en besig om volwasse te word. Volgens *Technorati* is reeds ’n invloedryke en belangrike deel van die web. Baie van die blogs is (mense wat ’n blog skep en dan belangstelling verloor) en baie is ook *spam*-blogs of s-blogs, maar wat saak, is die miljoene aktiewe, geesdriftige bloggers wat nie kan wag dat mense deur hul sleutelgat nie.

Die onderstaande vrae het betrekking op gedeeltes van bostaande teks. Daardie dele wat betrekking het, is telkens in ’n blokkie geplaas. In elkeen van hierdie blokkies is daar ’n aantal woorde weggelaat. U moet die plek aandui waar die woorde weggelaat is, sowel as die woord wat daar weggelaat is. Hier volg ’n voorbeeld:

Wêreldwyd ⁽ⁱ⁾ daar meer ⁽ⁱⁱ⁾ as 100 miljoen blogs ⁽ⁱⁱⁱ⁾ en die syfer groei teen 120 000 ^(iv) blogs per dag.

Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii)
- D. By posisie (iv).

Watter woord is op hierdie plek weggelaat?

- A. was
- B. word
- C. is
- D. sal wees

Hier volg die vrae:

^{23&24} Hierdie ⁽ⁱ⁾ stelling ⁽ⁱⁱ⁾ gemaak ⁽ⁱⁱⁱ⁾ deur die ^(iv) blogsoekenjin *Technorati* wat ^{25&26} die ontwikkeling ⁽ⁱ⁾ van blogs ⁽ⁱⁱ⁾ en ander gebruikersgedrewe inhoud ⁽ⁱⁱⁱ⁾ die web ^(iv) volg.

23. Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii)
- D. By posisie (iv).

24. Watter woord is op hierdie plek weggelaat?

- A. sal
- B. was
- C. kan
- D. word

25. Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii)
- D. By posisie (iv).

26. Watter woord is op hierdie plek weggelaat?

- A. op
- B. in
- C. deur
- D. by

27&28 Die ⁽ⁱ⁾ blogs wêreldwyd ⁽ⁱⁱ⁾ het in ⁽ⁱⁱⁱ⁾ minder as 'n ^(iv) jaar 'n geweldige toename getoon en het glo verdubbel van 35 na 70 miljoen.

27. Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii)
- D. By posisie (iv).

28. Watter woord is op hierdie plek weggelaat?

- A. nommer
- B. aantal
- C. klomp
- D. menigte

29&30 Die blogosfeer, ⁽ⁱ⁾ die ⁽ⁱⁱ⁾ wêreldwye blog-gemeenskap genoem ⁽ⁱⁱⁱ⁾ word, is sterk en besig om ^(iv) volwasse te word.

29. Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii).
- D. By posisie (iv).

30. Watter woord is op hierdie plek weggelaat?

- A. waar
- B. kan
- C. soos
- D. hoe

31&32 Volgens ⁽ⁱ⁾ Technorati is ⁽ⁱⁱ⁾ reeds 'n invloedryke en ⁽ⁱⁱⁱ⁾ belangrike deel ^(iv) van die web.

31. Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii)
- D. By posisie (iv).

32. Watter woord is op hierdie plek weggelaat?

- A. hy
- B. dit
- C. ons
- D. sy

33&34 Baie ⁽ⁱ⁾ van die ⁽ⁱⁱ⁾ blogs is ⁽ⁱⁱⁱ⁾ (mense wat 'n blog ^(iv) skep en dan belangstelling verloor) ^{35&36} en ⁽ⁱ⁾ baie ⁽ⁱⁱ⁾ is ook ⁽ⁱⁱⁱ⁾ spam-blogs of ^(iv) s-blogs, ^{37&38} maar ⁽ⁱ⁾ wat saak, ⁽ⁱⁱ⁾ is die miljoene ⁽ⁱⁱⁱ⁾ aktiewe, ^(iv) geesdriftige bloggers ^{39&40} wat nie ⁽ⁱ⁾ kan wag ⁽ⁱⁱ⁾ dat mense ⁽ⁱⁱⁱ⁾ deur hul sleutelgat ^(iv) nie.

33. Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii)
- D. By posisie (iv).

34. Watter woord is op hierdie plek weggelaat?

- A. lewendig
- B. skeppend
- C. dormant
- D. bruikbaar

35. Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii)
- D. By posisie (iv).

36. Watter woord is hierdie plek weggelaat?

- A. sodoende
- B. sommer
- C. sowat
- D. sogenaamde

37. Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii)
- D. By posisie (iv).

38. Watter woord is op hierdie plek weggelaat?

- A. is
- B. maak
- C. word
- D. bly

39. Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii)
- D. By posisie (iv).

40. Watter woord is op hierdie plek weggelaat?

- A. kyk
- B. sien
- C. loer
- D. skreeu

Hoewel laasgenoemde iterasie van hierdie itemtipe onlangs eers bedink is (vyfde en sesde loodsfase van 'n toets met 88 items) en daar tot op hede nog nie veel statistiese inligting oor hierdie formaat beskikbaar is nie, is eerste aanduidings (vergeelyk Tabel 4.6) wel dat hierdie 'n stap in die regte rigting is, veral as in ag geneem word dat hierdie iterasie aan die einde van die eerste studiejaar onder studente met beide gemiddelde en onder gemiddelde vermoëns geloods is en hul akademiese geletterdheidsvlakke teen hierdie tyd reeds 'n verbetering toon (Van Dyk, 2005; en Van der Slik & Weideman, 2008). Wat wel opmerklik is, is dat die alpha-waardes van 'n groep met onder gemiddelde vermoëns (n=531) beduidend laer is as by die groep met gemiddelde vermoëns (n=277). Daar kan egter op hierdie stadium maar net daarvoor gespekuleer word, omdat daar 'n veelheid van veranderlikes ter sprake kan wees, byvoorbeeld die feit dat die een groep meer homogeen in terme van vaardigheidsvlakke daar uitsien as die ander, die feit dat dit twee verskillende instellings is waar die toetse afgeneem is, die feit dat hierdie toetse teen die einde van die akademiese jaar afgeneem is (studente se vaardigheidsvlakke het dus al verbeter en hulle kan teen hierdie tyd ook al bekend wees met die formaat van hierdie vraag), en die feit dat hierdie toetse slegs nog in 'n loodsfase is. Die gemiddelde diskriminasiewaarde vir beide groepe is insgelyks ook

laag. Beide hierdie punte word vir doeleindes van hierdie studie nie verder ondersoek nie, maar sal wel in die slothoofstuk by verdere navorsingsmoontlikhede genoem word.

Tabel 4.6: Vergelyking van toetsgemiddeld, betroubaarheid, gemiddelde moeilikheidsgraad en gemiddelde diskriminasiewaarde van sinsluitingsubtoets vir 'n verdere loodsfasie in 2008

Weergawe	Vermoë	Gemiddeld	Alpha	Gemiddelde moeilikheidsgraad	Gemiddelde diskriminasie
<i>Vyfde(Afrikaans)</i>	gemiddeld	62.89%	0.82	63%	0.24
<i>Sesde(Afrikaans)</i>	onder gemiddeld	55.20%	0.68	55%	0.13

Anekdotiese getuienis van dosente dui wel daarop dat die gesigsgeldigheid van hierdie formaat aansienlik hoër is as by vorige weergawes van hierdie taaktipe. Daar word tans met selfs 'n volgende iterasie geëksperimenteer, waaroor daar uit die aard van die saak nog nie besonderhede beskikbaar is nie, maar die verwagting is dat dit selfs nog beter sal toets. 'n Voorbeeld hiervan sien as volg daar uit:

Afdeling 7: Teksredigering

Hieronder volg 'n stukkie teks. In hierdie teks is daar enkele woorde weggelaat. Lees dit eers om 'n geheelbeeld van die teks te vorm voordat u enige van die vrae beantwoord.

Die Suid-Afrikaanse rolprentbedryf is die naasoudste die wêreld, maar 'n dwerg op die wêreldverhoog. 'n Gebrek aan finansiële en regeringsondersteuning sy groei en Suid-Afrikaanse gehore verkies om eerder vir Hollywood-flieks te betaal. Volgens Roberta du Randt van Penguinfilms al verskeie treffer-produksies vervaardig het, is die Suid-Afrikaanse filmbedryf egter besig om te herleef. “Met die koms van digitale tegnologie het dit vir Suid-Afrikaners makliker geword om ons eie te vertel. Dis 'n goedkoper manier van films maak en dit bring nuwe moontlikhede vir die plaaslike bedryf.” Maar sy glo Suid-Afrikaanse gehore speel 'n groot rol in die sukses van die filmbedryf. “Daar bestaan 'n dat Suid-Afrikaanse rolprente swak is, en dit moet verkeerd bewys word. Verlede jaar het ons 'n rolprent vervaardig wat uitstekend in die gevaar het, maar wat glad nie in Suid-Afrika ondersteun is nie.” Inisiatiewe M-Met se Movie of the Month-projek waar produksieaatskappye geld kan kry om plaaslike rolprente te vervaardig en so die plaaslike bedryf te stimuleer, doen ook sy deel”, sê Roberta.

Die onderstaande vrae het betrekking op gedeeltes van bostaande teks. Daardie dele wat betrekking het, is telkens in 'n blokkie geplaas. In elkeen van hierdie blokkies is daar 'n aantal woorde weggelaat. U moet die plek aandui waar die woorde weggelaat is, sowel as die woord wat daar weggelaat is. Hier volg 'n voorbeeld:

Die [i] Suid-Afrikaanse [ii] rolprentbedryf is die [iii] naasoudste [iv] die wêreld, maar 'n dwerg op die wêreldverhoog.

Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii).
- D. By posisie (iv).**

Watter woord is op hierdie plek weggelaat?

- A. op
- B. van
- C. in**
- D. vir

Hier volg die vrae:

51&52 'n [i] gebrek aan [ii] finansiële en [iii] regeringsondersteuning [iv] sy groei en Suid-Afrikaanse gehore verkies om eerder vir Hollywood-flieks te betaal.

53&54 Volgens Roberta du Randt [i] van Penguinfilms [ii] al verskeie trefferproduksies [iii] vervaardig het, is die Suid-Afrikaanse filmbedryf [iv] egter besig om te herleef.

51. Op watter plek is die woord weggelaat?

- A. By posisie (i).
- B. By posisie (ii).
- C. By posisie (iii).
- D. By posisie (iv).

52. Watter woord is op hierdie plek weggelaat?

- A. bevorder
- B. bewys
- C. belemmer
- D. kwyd

'n Ander onopgeloste probleem, wat vir doeleindes van hierdie studie nie aangespreek is nie, maar wat ook by verdere navorsingsmoontlikhede in die slothoofstuk van hierdie studie weer uitgewys sal word, is eerstens die kwessie van tekskeuse vir hierdie taaktipe – dit is besonder moeilik om 'n gepaste, koherente teks te vind vir hierdie taaktipe en dit verg heelwat verwerking en aanpassing van outentieke materiaal alvorens dit bruikbaar is. Tweedens sal die frekwensie waarteen woorde weggelaat word in hierdie gemutileerde formaat moontlik ook met vrug ondersoek kan word. Meer frekwente weglating sal die teks waarskynlik te ingewikkeld maak vir eerstejaarstudente en dit sou waarskynlik ook moeiliker wees om dit op te stel in hierdie formaat. Minder frekwente weglating sal uitleggewys weer van waarde wees, maar dit sal dan waarskynlik weer te maklik wees vir studente.

4.6 Uiteindelijke toetsspesifikasies

'n Uitdaging waar toetsontwikkelaars egter deurlopend voor te staan kom, is die belyning van die toetskonstruk met die toetsspesifikasies wat logieserwys uit hierdie raamwerk behoort voort te vloei. Volgens Alderson, Clapham en Wall (1995:9) is die toetsspesifikasies die stel vereistes wat aandui wat die toets veronderstel is om te toets en ook hoe dit getoets behoort te word. Davies *et al.* (1999:207) beskryf dit as “[a] document which sets out what a test is designed to measure and how it will be tested.” Davidson en Lynch (2002:4) verduidelik verder dat 'n toetsspesifikasie nodig is “... to provide an efficient blueprint by which many similar instances of the same assessment task can be generated.” Toetsspesifikasies word met ander woorde daargestel om die produksie van groot getalle toetsitems te monitor en te kontroleer ten einde daartoe by te dra dat verskillende weergawes van dieselfde konstruk konsekwent bly in terme van byvoorbeeld moeilikheidsgraad, lengte, formulering, ensovoorts – dit hou verband met die betroubaarheid van 'n toets waarop daar meer in die volgende hoofstuk uitgebrei sal word (Davidson & Lynch, 2002:4-5). Alderson *et al.* (1995:11) wys in aansluiting hierby daarop dat toetsspesifikasies vir standaardiseringsprosedures in terme van toetsontwikkeling gebruik kan word. Die belyning van toetskonstruk met toetsspesifikasie is volgens Van Dyk en Weideman (2004b:17) ook 'n eerste en betekenisvolle stap in die rigting van toetsgeldigheid (waaroor daar in Hoofstukke 6 en 7 meer inligting sal wees), omdat dit 'n aanduiding daarvan kan wees of die toets inderdaad meet wat dit veronderstel is om te meet. Davies *et al.* (1999:207) wys in hierdie verband insgelyks daarop dat “... specifications are important in the establishment of the test's construct validity.” Konstrukgeldigheid en die gedetailleerde beginsels of voorwaardes waarvolgens toetsitems voorberei moet word, hou dus direk met mekaar verband.

4.6.1 Die skryf van toetsspesifikasies

Die skryf van toetsspesifikasies is so 'n belangrike deel van die ontwerpproses dat dit die metode van toetsing (soos reeds hierbo gesuggereer) en die inhoud van toetse bepaal. Dit is dus by implikasie die riglyn of stel riglyne waarvolgens 'n toets

ontwikkel word. Dit kan op basis van 'n voorafbepaalde formaat wat aan bepaalde formele vereistes moet voldoen, geskoei wees, maar dit kan ook net enkele eienskappe van toetse insluit, waarvan die volgende voorbeelde is: die doel van die toets, die teikengroep, die struktuur van die toets op makro- en mikrovlak, die lengte van elke afdeling, die tydsduur, instruksies vir voltooiing, aanwysers vir bepunte, die verwagte respons(e), die konstruksie waarop dit gebaseer is, fasiliteite benodig, ensovoorts (McNamara, 2000:31; Davies *et al.*, 1999:207).

Dit is egter belangrik om daarop te let dat toetsspesifikasies nie noodwendig vas is nie, aangesien dit voortdurend kan verander na gelang van die omstandighede – loodstoetse kan byvoorbeeld 'n goeie aanwyser wees dat bepaalde items in 'n sekere formaat nie gewerk het nie, of dalk juis baie goed gewerk het (soos wat die geval was met die drie voorbeelde van toetsitems wat hierbo voorgehou is en soos wat ook elders in hierdie hoofstuk geïllustreer sal word). Net so kan groeiende kundigheidsvlakke van toetsontwikkelaars ook 'n invloed op toetsspesifikasies hê.

Toetsontwerp is dus 'n proses wat deur verskeie fases gaan en vooruitskouend én retrospektief benader behoort te word (verwys ook Hoofstuk 2) alvorens spesifikasies geskryf kan word, soos wat die res van hierdie hoofstuk sal illustreer. Dit beteken dat daar op 'n sistematiese wyse te werk gegaan moet word om 'n eenheid in 'n veelheid van vereistes te bewerkstellig deur byvoorbeeld die stel van sekere verbandhoudende statistiese parameters, die bepaling van die betroubaarheid van die toets, die validering van die toets, die skep van 'n teoretiese regverdiging vir die toets, die skep van 'n implementeerbare en uitvoerbare toetsstelsel, die sorgsaamheid waarmee die toets ontwikkel is, en om hierdie as ontwerpbeginsels van 'n goeie toets te aanvaar.

As daar uit die hoek van die sistematiese raamwerk wat ten grondslag van hierdie studie lê, na bogenoemde beginsels gekyk word, verwys dit telkens na een of ander verhouding tussen die leidende, tegniese funksie van ons ervaring en 'n ander dimensie daarvan. So byvoorbeeld verwys die tegniese sistematiesing van vereistes na die verhouding tussen die tegniese en die aritmetiese, binne konteks van

'n tegniese eenheid in 'n veelheid van (statisties-bepaalde/tegniese) parameters. Die vereistes vir die tegniese konsistensie of betroubaarheid van 'n toets verwys na die samehang tussen die tegniese en kinematiese dimensies, en die tegniese geldigheid/krag van 'n toets na die verhouding van die tegniese met die fisiese aspek van ons ervaring. Soos telkens reeds vermeld, bring die begrip van teoretiese regverdiging die verband uit tussen die tegniese bo-bou en teoreties-analitiese onderbou van 'n toets, terwyl implementeerbaarheid as idee weer die sosiale antisipasie in die tegniese eggo. Tegniese uitvoerbaarheid en sorgsaamheid is idees wat op hul beurt weer verbind is aan die verhouding tussen die tegniese en, respektiewelik, die ekonomiese en etiese aspekte. Elkeen van hierdie dra 'n ontwerpbeginsel by tot die debat oor verantwoordelike taaltoetsing.

Dit is aan sulke ontwerpbeginsels wat toetsspesifikasies gekoppel moet wees en bly. Davidson en Lynch (2002:14e.v.), Bachman en Palmer (1996:171e.v.) en Alderson *et al.*(1995:9e.v.) kan geraadpleeg word vir 'n uitgebreide uiteensetting van voorbeelde van formate van toetsspesifikasies. Vir doeleindes van hierdie studie is daar egter besluit op 'n bruikbare kombinasie van elemente uit bogenoemde werke en dit word in Tabel 4.7 hieronder saamgevat. Let egter daarop dat ook hierdie voorgestelde wyse van toetsspesifikasies dokumenteer, plooibaar is na gelang van die omstandighede; die bedoeling is bloot dat dit as riglyn vir toetsontwikkelaars dien, soos gesuggereer in Figuur 4.1 en die voorafgaande. Volgens Alderson *et al.* (1995:11, 16, 20) kan daar egter ook spesifikasies vir toetsgebruikers en sogenaamde toetsvalideerders wees. Vir doeleindes van hierdie studie is daar aanvanklik nie spesifikasies vir toetsgebruikers opgestel nie, aangesien die algemene inligtingsblad vooraan TAG, tesame met die rubrieke (wat wel in die tabel hieronder opgeneem is) voldoende hiervoor is. Die Universiteit Stellenbosch het egter 'n stel toetsspesifikasies vir toetsgebruikers van TAG en TALL ontwikkel en dit verskyn in Bylaag B.

Tabel 4.7: Toetsspesifikasies van TAG vir toetsontwikkelaars

Doel van toets			
TAG is die Afrikaanse akademiese geletterdheidstoets van die Universiteite van Pretoria, Stellenbosch en die Vrystaat, en die Noordwes Universiteit. Dit word gebruik om (i) voorbereidingsvlakke van studente in terme van akademiese geletterdheid (akademiese taalvaardigheid) te meet as deel van 'n toelatingsmeganisme, en/of (ii) studente met hoë akademiese potensiaal, maar vir wie akademiese geletterdheid 'n risiko in die suksesvolle afhandeling van hul studies kan wees, op gepaste taalondersteuningsprogramme/akademiese geletterdheidskursusse te plaas. TALL is die Engelse weergawe van TAG en is gebaseer op dieselfde konstruksie.			
Teikengroep			
Die teikengroep van TAG is voornemende eerstejaarstudente wanneer die toets as deel van 'n toelatingsmeganisme gebruik word, en/of reeds toegelate eerstejaarstudente wat op gepaste akademiese taalkursusse geplaas moet word na gelang van hul risikolakke.			
Voorgestelde tyd	50-minute – 1-uur	Aantal vrae	Tussen 60 en 70
Puntetotaal	100	Toetsformaat	Pen en papier (skriftelik)
Toetsmetode	Slegs meervoudige keuse vrae → los antwoordblad	Tekstipes / materiaal	Ouentiek, Graad 11- / eerstejaarsvlak, ongeveer 45- 60% leesgemak
Makro-samestelling			
Ses afdelings (wat rondgeskuif en aangepas kan word):			
Afdeling 1: Skommelteks → ongeveer 5 vrae (5 punte)			
Afdeling 2: Interpretasie van visuele inligting → ongeveer 7 vrae (7 punte)			
Afdeling 3: Teksbegrip → ongeveer 22 vrae (50 punte)			
Afdeling 4: Akademiese woordeskat → ongeveer 9 vrae (18 punte)			
Afdeling 5: Tekstipe → ongeveer 5 vrae (5 punte)			
Afdeling 6: Teksredigering → ongeveer 15 vrae (15 punte)			
Mikro-samestelling			
Voorbeelde van rubrieke (kan aangepas word)	Artikulasie met konstruksie (kan aangepas word)		
Afdeling 1: Skommelteks Die korrekte volgorde van die onderstaande sinne is doelbewus verander. Dui aan wat die korrekte volgorde behoort te wees deur u keuse op die los antwoordblad te merk.	<ul style="list-style-type: none"> • Tekssamestelling en samehang • Betekeniswaarde groter as sinsvlak 		
Afdeling 2: Interpretasie van visuele inligting Bestudeer die onderstaande teks en grafiek en beantwoord die vrae wat daarop volg deur u keuse op die los antwoordblad te merk.	<ul style="list-style-type: none"> • Interpretasie van grafiese inligting • Klassifikasie en vergelyking 		
Afdeling 3: Teksbegrip Die berig op die volgende bladsy is verwerk uit <i>Die Burger</i> , 20 Februarie 2004, p.13. Dit word in twee dele verdeel wat inhoudelik direk by mekaar aansluit. Daarom moet u eers Deel I (vrae 13-29) beantwoord voordat u Deel II (vrae 30-34) aanpak. Merk telkens u keuse op die los antwoordblad.	<ul style="list-style-type: none"> • Akademiese woordeskat • Metaforiese taalgebruik • Tekssamestelling en samehang • Bekendheid met tekstipes, styl en register • Onderskeid tussen hoofsaak en bysaak, oorsaak en gevolg en feite en menings • Klassifikasie en vergelyking • Afleivermoë en toepassingsvermoë • Definiëring, beredenering en bewyslewing • Betekeniswaarde groter as sinsvlak 		

Afdeling 4: Akademiese woordeskat

Kies die beste moontlike antwoord uit die lys van opsies en merk u keuse op die los antwoordblad.

- Akademiese woordeskat
- Afleivermoë en toepassingsvermoë

Afdeling 5: Tekstipe

Kolom A hieronder bevat vyf taaluitinge wat elk 'n ander tekstipe (advertensietaal, koeranttaal, literêre taal, ensovoorts) verteenwoordig. Paar elke item in kolom A af met die item in kolom B wat dieselfde tekstipe verteenwoordig. Merk u keuse op die los antwoordblad.

- Bekendheid met tekstipes, styl en register
- Afleivermoë en toepassingsvermoë

Afdeling 6: Teksredigering

Hierdie afdeling bestaan uit drie dele. Merk u keuse vir elke vraag op die los antwoordblad.

Deel I

Sommige woorde is uit die onderstaande teks weggelaat. Die teks verskyn in die linkerkantste kolom. U moet die **plek** (a, b, c, d) kies waar die woord weggelaat is. Die korrekte antwoord vir die eerste een word vir u as voorbeeld gegee.

- Tekssamestelling en samehang
- Onderskeid tussen hoofsaak en bysaak, oorsaak en gevolg en feite en menings
- Afleivermoë en toepassingsvermoë
- Betekeniswaarde groter as sinsvlak

Deel II

Sommige woorde is uit die onderstaande teks weggelaat. Die teks verskyn in die linkerkantste kolom. Die plek waar die woord weggelaat is, word in die teks aangedui met 'n skuinsstreep (/). U moet die **woord** kies wat op daardie plek weggelaat is uit die moontlikhede wat in die tweede kolom gegee word. Die korrekte antwoord vir die eerste een word vir u as voorbeeld gegee, en is duidelik gemerk in die regterkantste kolom.

Deel III

Sommige woorde is uit die onderstaande teks weggelaat. Die teks verskyn in die linkerkantste kolom. Kies eers die **plek** waar die woord weggelaat is, en daarna die **woord**. Die korrekte antwoord vir die eerste een word weer vir u as voorbeeld gegee en is duidelik gemerk in die middelste en regterkantste kolom.

4.7 Samevatting

In Suid-Afrika word daar besonder min oor toetsontwikkeling gepubliseer. 'n Siniese waarnemer sou moontlik kon opmerk dat dit die geval is, omdat baie taaltoetse hier, soos elders, nog in die sogenaamde prewetenskaplike fase is. Tog is daar in hierdie studie na 'n aantal taaltoetse (ELSA Plus, PTEEP, TALL, en TAG) verwys wat oënskynlik met groot noukeurigheid saamgestel is. Nogtans is uiteensettings en besprekings van hoe selfs hierdie toetse ontwerp en ontwikkel word, heel skaars. Dit is uit die aard van die saak nie gewens nie, aangesien daar 'n verantwoordelikheid aan die kant van toetsontwikkelaars is om te verseker dat

toetse, veral die tipe waaroor daar in hierdie studie gerapporteer word, op deursigtige wyse ontwerp en saamgestel moet word. Daar moet volgens Shohamy (2001) in elk geval meer gereeld diskoers gevoer word oor hoekom toetse lyk soos dit uiteindelik lyk, veral wanneer daar besluit word hoe om van toetskonstruk te beweeg na eksperimentering met taaktipes ten einde toetsspesifikasies te kan formuleer vir toetsitems.

In Hoofstuk 2 is genoem dat die kwalifiserende funksie van 'n ontwerp, as 'n toegepaste linguistiese oplossing, nie net te vinde is in die tegniese aspek daarvan nie, maar ook 'n aanduiding gee van hoe die ontwerp bestudeer, ondersoek, verklaar en selfs verander kan word vanuit die analitiese perspektief wat as rasionaal daarvoor aangebied kan word. Daar is in hierdie hoofstuk op beide hierdie funksies gefokus in die opsig dat die analitiese basisfunksie (byvoorbeeld deur die byeenbring en analise van empiriese data) van die ontwerp gebruik is om die leidende of kwalifiserende funksie te informeer. Die kwalifiserende funksie van TAG (soos duidelik in hierdie hoofstuk geïllustreer) is met ander woorde te vinde in die tegniese aspek van die ontwerp daarvan, wat vanuit die analitiese basis ondersoek, verklaar en verander is. Dit blyk dus in orde te wees om aan te voer dat 'n holisties-geïntegreerde benadering binne konteks van die bostaande gevolg moet word eerder as om net op hierdie of daardie aspek van 'n toets te fokus (soos ook duidelik in Hoofstuk 6 sal uitkom). Let egter daarop dat hoewel hierdie studie hoofsaaklik met die konstitutiewe konsepte van linguistiese ontwerpe te make het, dit nie die regulatiewe idees uitsluit nie. Die twee is voortdurend in wisselwerking met mekaar en daarom sal daar, soos ook in hierdie hoofstuk die geval was, telkens na die regulatiewe aspekte van linguistiese ontwerpe verwys word.

Hierdie hoofstuk was dus 'n poging om die narratief van ontwerp oor TAG se beweging van konstruk na spesifikasie na taak- en itemtipe te vertel binne konteks van die teoretiese raamwerk wat in Hoofstuk 2 geskep is. Die ontwerpbeginsels vir taaltoetse wat deur hierdie raamwerk gegenereer word, het telkens in hierdie hoofstuk aan die orde gekom in die byeenbring van die konsepte van die tegniese

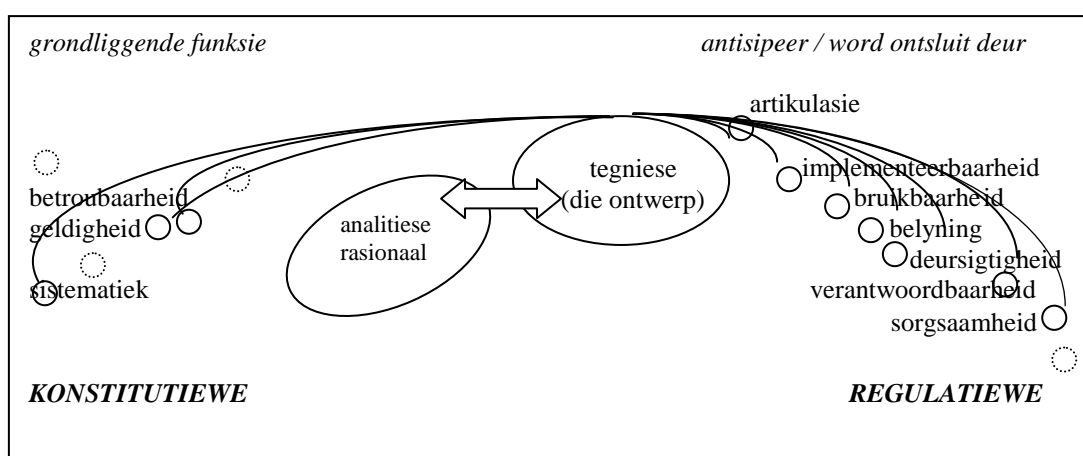
parameters (op toets- en itemvlak), tegniese konsistensie of betroubaarheid, tegniese-analitiese regverdigbaarheid (konstrukgeldigheid), tegniese nuttigheid / waarde, tegniese balans, en tegniese belyning. Die volgende drie hoofstukke bespreek in meer besonderhede twee van die sogenaamde fundamentele oorwegings by toetsontwerp volgens die huidige ortodoksie, wat die verdere konteks waarteen konstruksie, toetsspesifikasies en taaktipes / itemtipes staan, sal inkleef. Dit is betroubaarheid (Hoofstuk 5) en geldigheid (Hoofstuk 6) wat deur 'n proses van validering (Hoofstuk 7) bewerkstellig word,

Betroubaarheid: gister, vandag en môre dieselfde

5.1 Inleidend

In die vorige hoofstuk is daar melding gemaak van 'n balans wat tussen die konstitutiewe en die regulatiewe aspekte van toegepaste linguistiese artefakte, as tegniese ontwerpe ten einde taalprobleme op te los, bewerkstellig moet word. Binne die konteks van 'n postmoderne ingesteldheid blyk dit dat die hedendaagse neiging, ook in Suid-Afrika, egter die fokus eerder weg van die konstitutiewe en nader aan die regulatiewe verskuif. Daar word dus oënskynlik meer klem op politiese en etiese oorwegings, deursigtigheid, en verantwoordbaarheid (alles regulatiewe aspekte) gelê, met die gevolg dat die konstitutiewe aspekte of grondliggende funksies van ontwerpe soms moontlik van minder belang geag kan word. Hierin is egter 'n stukkie ironie opgesluit, aangesien die kern van 'n postmodernistiese tradisie, in die woorde van Weideman (2003d:15-17), juis is om 'n benadering te volg wat krities dog ook verantwoordbaar is. As die term verantwoordbaarheid gebruik word, beteken dit noodwendig dat die grondliggende funksies soos die teoretiese en empiriese onderbou nie onderspeel (of oorspeel) kan / moet word nie. Die basiese premis is dus om met 'n sosiaal verantwoordbare en teoreties regverdigbare ontwerp vir die probleem vorendag te kom. Figuur 5.1 (oorgeneem en verwerk uit Weideman, 2006 en wat min of meer ooreenstem met Figuur 4.2, maar gerieflikheidshalwe weer hier aangebied word) is 'n illustrasie

hiervan, waarin dit blyk dat die grondliggende funksies waarop toegepaste linguisties-tegniese ontwerpe berus, ontsluit word deur regulatiewe aspekte en dat die twee kante van die model (soos in Hoofstuk 2 voorgehou) in dinamiese interaksie met mekaar is en dus in balans móét wees – die een kan nie sonder die ander nie. Indien die pendulum dus vashaak by óf die een óf die ander kant, sal die ontwerp vir die oplossing van die probleem gevolglik nie meer met integriteit aangebied kan word nie.



Figuur 5.1: Illustrasie van die balans tussen die konstitutiewe en die regulatiewe

Let daarop dat hoewel die uitspraak hierbo gemaak word dat die neiging plaaslik én internasionaal is om die fokus weg van die konstitutiewe na die regulatiewe te skuif, dit nie noodwendig die geval by alle taaltoetse is nie, veral nie daardie met medium- tot hoëbelang funksies en wat op groot skaal bedryf word nie. Die hedendaagse Suid-Afrika is en bly egter 'n polities gedrewe omgewing en bostaande stelling moet dus as 'n vermaning gelees word, eerder as 'n feit, naamlik dat die ontwikkelaars van toetse in Suid-Afrika daarop moet let om deurlopend die een te doen, sonder om die ander na te laat.

Hoewel hierdie studie meer gemoeid is met die konstitutiewe aspekte van toegepaste linguistiese ontwerpe, moet die leser daarop let dat dit nie enigsins probeer impliseer dat die een kant van die model meer aandag as die ander verdien nie. Soos hierbo uiteengesit, moet die konstitutiewe en regulatiewe kante van enige

ontwerp in konjunksie met mekaar wees en mekaar eerder komplimenter as teenwerk. In hierdie hoofstuk word 'n bespreking van een van die basiskonsepte in toetsing, naamlik betroubaarheid, aangebied. Ofskoon hierdie begrip aan die konstitutiewe kant van die skaal lê, moet dit nie uit die oog verloor word nie dat betroubaarheid ook gaan bydra tot die bruikbaarheid van, die regverdiging vir, en die sorgsaamheid waarmee TAG ontwikkel en gebruik word. Die argument van hierdie hoofstuk is dus 'n bevestiging daarvan dat die konstitutiewe en regulatiewe komponente van bostaande model in interaksie en in enige goeie en regverdigbare ontwerp in harmonie met mekaar is.

5.2 Betroubaarheid gedefinieer

Naas geldigheid, wat in die volgende hoofstuk bespreek sal word, is betroubaarheid die belangrikste kenmerk waaroor 'n toets kan beskik. Die rede hiervoor is dat betroubaarheid in die eerste plek konsistensie vir 'n toets bied wat aanleiding gee tot geldigheid. Ten tweede is betroubaarheid 'n aanduiding daarvan dat veralgemenings van toetsresultate minstens empiries-analities, teoreties regverdigbaar kan wees. In lekerterme gestel, wys betroubaarheid op die konsistensie en stabiliteit van 'n toets. Met konsistensie word bedoel die mate waartoe al die dele van die toets saamhang om dieselfde saak / kwessie (byvoorbeeld akademiese geletterdheid) te meet. Stabiliteit weer, is die mate waartoe 'n toets dieselfde saak / kwessie op verskillende tye en in verskillende kontekste meet. Dit gaan dus oor die onderlinge verbande of samehang van 'n toets of toetse, sowel as die herhaalbaarheid daarvan. Hatch en Faraday (1982:244) verwoord dit as volg: “[reliability is] the extent to which a test produces consistent results when administered under similar conditions.” Davies *et al.* (1999:168) stel dit myns insiens selfs nog beter met die stelling “[reliability is the] actual level of agreement between the results of one test with itself or with another test.” In terme van die teoretiese onderbou wat vir hierdie studie gebruik word, is die tegniese konsistensie van 'n toets as meetinstrument, en die stabiliteit van die resultate wat verkry word, konseptueel verwant aan die eggo van die kinematiese dimensie van ons ervaring in die tegniese modus. In die oorspronklik-kinematiese ervaringswyse

tree fenomene soos konsistensie, uniforme of reëlmatige beweging, en stabiliteit na vore. In die analogies-tegniese sin waarin daar hierna verwys word, verkry die reëlmatigheid 'n tegniese stempel, wat uitdrukking vind in die konsep van tegniese betroubaarheid.

Beide Davies *et al.* (1999:169) en Nitko (2004:60) stel dit egter ook dat betroubaarheid 'n noodsaaklike, maar nie voldoende kenmerk van 'n toets is nie. Binne konteks van die teoretiese raamwerk waarop hierdie studie gebaseer is en dit wat in paragraaf 5.1 hierbo gestel is, is dit hieruit duidelik dat betroubaarheid alleen nie genoegsaam is vir aannames oor die sogenaamde kwaliteit van 'n toets nie. Dit moet in ag geneem word dat toetsresultate niks meer is as kwantitatiewe beskrywings van individuele eienskappe of vermoëns op 'n gegewe moment nie. Hierdie kwantitatiewe beskrywings stel belanghebbendes, soos universiteitsowerhede (vir doeleindes van hierdie studie), in staat om vergelykings te kan maak, om voorspellings oor toekomstige gedrag te kan waag, en om die invloed van 'n verskeidenheid van faktore op gedrag te kan bepaal. 'n Betroubaarheidskoëffisiënt verwys dus nie slegs na die toets as geheel nie, maar eerder en veral na resultate behaal op 'n bepaalde toets.

Die basiese veronderstelling van betroubaarheid is derhalwe dat 'n presiese en konsistente beskrywing van die toetsling gegee kan word ten opsigte van daardie aspek, byvoorbeeld akademiese geletterdheid, waarin belanggestel word, en dat voorspellings oor toekomstige toetsresultate gemaak kan word. De Zeeuw (1978:152) se definisie is selfs nog meer genuanseerd as hy betroubaarheid beskryf as:

... de mate van nauwkeurigheid, dit is foutloosheid, waarmee op grond van de gegewens ontleend aan de test als meetinstrument uitsprake kunnen worden gedaan. Nauwkeurigheid in drieërlei sin: nauwkeurigheid in die betekenis van die mate van presisie welke gebonden is aan het instrument; nauwkeurigheid als de mate van constantie of stabiliteit van het met de test gemetene in die tijd; en de nauwkeurigheid als de mate van generaliseerbaarheid vanuit de gegeven test resultaten naar mogelijke gelijksoortige toetsresultaten.

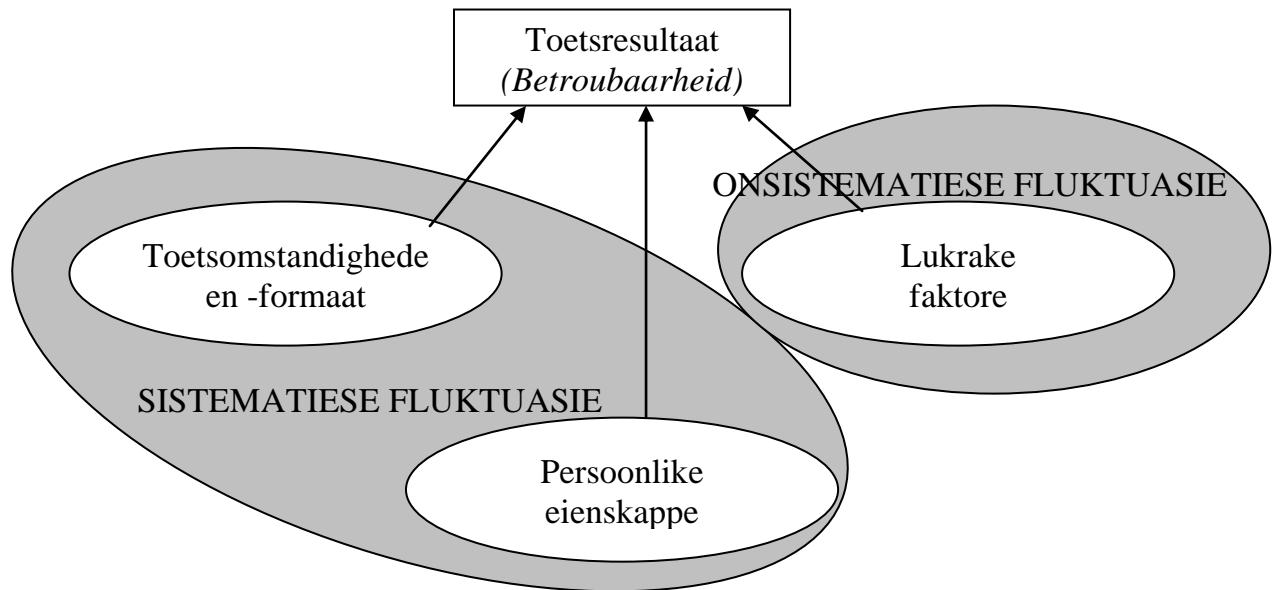
Vrae wat dus deurlopend oor meetnaukeurigheid, oftewel betroubaarheid, gevra moet word, is:

- Is die resultaat wat persoon X vandag op die toets behaal dieselfde as die resultaat wat hy / sy oor 'n maand of oor 'n jaar sal behaal as hy / sy weer die toets moet skryf en nog oor dieselfde vermoë as tans beskik? Is daar met ander woorde stabiliteit oor tyd heen?
- Is die resultaat wat persoon X behaal 'n ware refleksie van sy / haar vermoë? Hoe 'n akkurate weerspieëling is die resultate van die ware vermoë van die toetsling?
- Hoe veralgemeenbaar is die toetsresultate?

Dit is belangrik om daarop te let dat geen toets ooit 100% betroubaar is nie en dat toetsresultate van die een geleentheid na die volgende kan wissel, omdat daar 'n aantal veranderlikes is wat daarop kan inspeel, soos die omstandighede en formaat waarin die toets afgeneem is, persoonlike eienskappe van die toetsling, en ander toevallige faktore wat onvoorspelbaar en tydelik of wisselend van aard is (Bachman, 1990:164). Sou 'n toetsling byvoorbeeld by een toetsgeleentheid 'n lae resultaat behaal en by 'n volgende geleentheid 'n aansienlik hoër resultaat, is dit 'n aanduiding daarvan dat die toets nie stabiel is oor tyd heen nie en die resultate derhalwe nie betroubaar is as indikator van die persoon se vermoë nie (Bachman, 1990:24), veral in gevalle waar daar geen intervensie plaasgevind het nie. Daar is dus 'n fluktuasie in die response van die toetsling en dit kan aan die een of ander veranderlike, soos hierbo genoem, toegeskryf word.

Fluktuasie word in twee kategorieë verdeel, naamlik sistematiese variasie en onsistematiese variasie (Smit, 1991:25-26; Bachman, 1990:163-164). Eersgenoemde het te make met faktore wat uniform is of deurlopend min of meer konstant bly, byvoorbeeld die toetsomstandighede en -formaat, en persoonlike eienskappe soos ras, geslag, vakkennis, ensovoorts. Dit is daardie faktore waaroor daar tot 'n mate beheer is. Sou 'n toetsling byvoorbeeld agtergrondkennis het oor die tema waaromheen die toets gebou is, behoort dit argumentsonthelwe konstant te bly van die een toetsgeleentheid na die volgende en kan daarom as sistematiese variasie gekategoriseer word (Smit, 1991:25; Bachman, 1990:164).

Variasie vind egter ook, soos reeds genoem, op sogenaamd onsistematiese wyse plaas waar daar geen patroon sigbaar is nie, of waar daar nie werklik beheer oor die fluktuasie is nie (Bachman, 1990:164). Onsistematiese variasie kan dus aan verskillende sogenaamde metingsfoute toegeskryf word. Smit (1991:26) se verklaring vir onsistematiese variasie is dat enige faktor anders as dit wat die toets veronderstel is om te meet en wat 'n fluktuasie in toetsresultaat tot gevolg het, as onsistematiese fluktuasie en dus 'n oorsaak van fouttelling of metingsfout gesien moet word. Bachman (1990:163) se definisie blyk egter oper te wees en daar word vir doeleindes van hierdie studie daarby volstaan. Die rede hiervoor is dat dit meer in lyn is met die teoretiese raamwerk wat in Hoofstuk 2 van hierdie studie uiteengesit is en wat die agtergrond is waarteen bogenoemde gelees moet word. In 'n postmoderne perspektief kan daar verder nie by 'n beperkende definisie soos die van Smit, naamlik dat enige faktor anders as dit wat die toets veronderstel is om te meet onsistematiese variasie tot gevolg het, volstaan word nie. Bachman se interpretasie artikuleer dus beter met die genoemde teoretiese raamwerk van hierdie studie, naamlik dat daar balans en belyning tussen die konstitutiewe en die regulatiewe aard van ontwerpe moet wees. Figuur 5.2 hieronder wat verwerk en aangepas is uit Bachman (1990:165) is 'n illustrasie hiervan in die opsig dat regulatiewe en konstitutiewe konsepte konstant in interaksie met mekaar is of mekaar kan beïnvloed: hoewel betroubaarheid aan die grondliggende kant van die skaal lê, kan dit beïnvloed word deur aspekte aan die kwalifiserende kant van die skaal. So gebeur dit dan dat faktore wat tot sistematiese en onsistematiese fluktuasie aanleiding gee, 'n invloed kan hê op 'n konsep soos betroubaarheid; net so kan faktore wat tot variasie bydra ook 'n invloed op 'n regulatiewe konsep soos bruikbaarheid hê.



Figuur 5.2: Illustrasie van die invloed van sistematiese en onsistematiese fluktuasie op betroubaarheid

Hierdie interaksie tussen die konstitutiewe en die regulatiewe word verder geïllustreer deur Bachman (1990:24) se verduideliking dat

[i]n any testing situation, there are likely to be several different sources of measurement error, so that the primary concerns in examining the reliability of test scores are first, to identify the different sources of error, and then to use the appropriate empirical procedures for estimating the effect of these sources of error on test scores. The identification of potential sources of error involves making judgements based on an adequate theory of sources of error.

As 'n holisties geïntegreerde benadering, soos in die vorige hoofstuk voorgestel, dus ook vir die bepaling van betroubaarheid gevolg word, blyk Bachman (1990:163-166) se uiteensetting van sistematiese en onsistematiese faktore wat betroubaarheid beïnvloed 'n werkbare opsie te wees. Net soos wat individue se vermoëns byvoorbeeld varieer, sal die mate waartoe hul deur enige van die drie faktore geraak word ook varieer. Sommige toetslinge kan byvoorbeeld uitstekend vaar met meervoudige keuse toetse (soos TAG en TALL), terwyl dit nie die geval by ander is nie. Dit kan ook wees dat een etniese groep of geslag weer beter of swakker as 'n ander groep of geslag vaar. Sommige kan goed uitgerus wees wanneer die toets afgelê word en ander weer nie, wat 'n invloed op toetsprestasie (of fluktuasie) kan hê. Alle faktore wat dus 'n invloed op die toetsresultaat kan hê, dra potensieel tot

die metingsfout by en moet so ver moontlik vermy word, of soos Bachman (1990:166) dit stel: “[a] major concern in the design and development of ... tests ... is to minimize the effects of test method, personal attributes that are not part of [the] ability [to be measured], and random factors on test performance.” Foutfaktore is dus daardie veranderlikes wat nie relevant tot die doel van die toets is nie en wat van tyd tot tyd fluktuasies in toetstellings veroorsaak (Smit, 1991:27).

Dit is dus belangrik om van die sogenaamde fouttelling kennis te neem binne konteks van die definiëring van betroubaarheid, aangesien dit die metingsfout (oftewel die ‘Standard Error of Measurement’ - SEM) konstitueer en die SEM ’n invloed op die betroubaarheid van ’n toets het. Die rede hiervoor is, soos Smit (1991:26) dit stel: “[h]oe groter die rol van die fouttelling/metingsfout in prestasie in ’n bepaalde toets is, hoe groter sal die onsistematiese fluktuering van individuele toetsprestasie by herhaalde toepassings wees, en hoe minder betroubaar sal die toets gevolglik wees.” Bachman (1990:24, 160) som dit as volg op: “[r]eliability is a quality of test *scores*, and a perfectly reliable score would be one which is free from errors of measurement ... [and therefore we] must be concerned about errors of measurement ... because we know that test performance is affected by factors other than the abilities we want to measure.” Dit hou in dat wanneer bostaande drie faktore (fluktuasie in toetsomstandighede en -formaat, fluktuasie in persoonlike eienskappe en fluktuasie in toevallige faktore) tot die minimum beperk word, die metingsfout verlaag en die betroubaarheid dus sal verhoog.

Opgesom kom dit daarop neer dat ’n betroubare toets ’n toets is waar die waargenome telling ’n akkurate weergawe van die ware telling is met ’n baie lae foutkomponent. In ’n poging om resultate te verkry wat dus so na as moontlik aan die ware telling is, moet daardie faktore (verwys Figuur 5.2 en die bespreking daaromheen) wat waargenome tellings beïnvloed so ver moontlik gekontroleer word – deur kontrole word die metingsfout beperk en die betroubaarheid uiteindelik verhoog.

Miller, Linn en Gronlund (2009:108) stel dit dat die begrip betroubaarheid (ook soms as die veralgemeenbaarheid van 'n toets beskryf) verklaar kan word aan die hand van die volgende vier punte, wat artikuleer met dit wat hierbo gestel is:

- Betroubaarheid het te make met die resultate van 'n toets en nie die toets self nie.
- 'n Betroubaarheidswaarde verwys altyd na 'n tipe konsistensie of stabiliteit.
- Betroubaarheid is 'n belangrike en nodige eienskap van 'n toets, maar dit is nie voldoende vir 'n argument vir waarom dit 'n kragtige (geldige) toets is of nie.
- Betroubaarheid word primêr aan die hand van statistiese analyses bepaal.

5.2.1 Die wiskundige onderbou van die konsep betroubaarheid

'n Definisie vir betroubaarheid is nie volledig sonder 'n verduideliking van die wiskundige beginsels wat grondliggend daartoe lê nie. Hierdie beginsels vind hul beslag in die sogenaamde klassieke toetsteorie (KTT) en daar is drie terme wat hier ter sprake is: waargenome telling, ware telling en fouttelling. Met die KTT as basis word daar met die aanname gewerk dat die fluktuasie of variasie in die waargenome telling die fluktuasie of variasie in die ware telling en die fouttelling insluit. Volgens Smit (1991:26) bestaan 'n toetsresultaat (anders gestel die waargenome telling) uit twee komponente, naamlik 'n ware telling en 'n fouttelling. Hy gee dit as volg weer:

$$X = W + F$$

waar X = die waargenome telling / toetsresultaat ('obtained score')

W = die ware telling ('true score')

F = die fouttelling ('error score')

Dit is belangrik om daarop te let dat toetsadministreerders slegs die waargenome telling sien en nooit die ware of die fouttellings nie. Hulle moet dan daarvolgens veralgemenings maak en ten einde hierdie veralgemenings so akkuraat as moontlik te maak, is dit nodig om 'n deeglike begrip te hê vir dit wat onder 'n

betroubaarheidswaarde lê. Dit kan byvoorbeeld so wees dat 'n student 70% op TAG behaal wat hom / haar in 'n lae risiko-kategorie plaas, maar sy / haar werklike akademiese geletterdheidsvlak is eintlik 74%. Dit beteken dat die fouttelling -4 is. In leketaal beteken dit die student se akademiese geletterdheidsvermoëns is eintlik 74%, maar daar is faktore wat 'n invloed gehad het op die fluktuasie, soos die toetsomstandighede wat nie optimaal was nie, of die feit dat die student moeg was. Net so kan daar ook faktore wees wat bevorderlik was vir toetsomstandighede en kan die student se ware telling op 66% lê. Faktore soos hierdie dra by tot metingsfoute wat 'n student se waargenome telling óf hoër óf laer kan laat neig.

Omdat een van die doelwitte van enige toets wat met integriteit ontwerp is, is om fouttellings tot die absolute minimum te beperk, is die volgende vergelyking van kardinale belang vir 'n beter begrip van die konsep betroubaarheid:

$$1 = \frac{\sigma_t^2}{\sigma_0^2} + \frac{\sigma_e^2}{\sigma_0^2}$$

σ_0^2 = variansie van die waargenome telling ('obtained score')

waar σ_t^2 = variansie van die ware telling ('true score')

σ_e^2 = variansie van die fouttelling ('error score')

Hierdie vergelyking dui daarop dat die verhouding van die ware telling tot die waargenome telling (σ_t^2/σ_0^2) plus die verhouding van die fouttelling tot die waargenome telling (σ_e^2/σ_0^2) gelyk aan 1 moet wees. Die verhouding van die ware telling tot die waargenome telling gee dus 'n aanduiding van die foutvariensie in 'n toets, oftewel daardie proporsie variensie in waargenome tellings wat bydra tot ware tellings. Hierdie konsep is die basiese definisie vir betroubaarheid. Die simbool vir betroubaarheid (r_{tt}) kan dus die verhouding (σ_t^2/σ_0^2) vervang in die vergelyking:

$$1 = r_{tt} + \frac{\sigma_e^2}{\sigma_0^2}$$

Indien die ratio van die fouttelling tot die waargenome telling (σ_e^2/σ_0^2) afgetrek word, is die basiese formule vir betroubaarheid die volgende:

$$r_{tt} = 1 - \frac{\sigma_e^2}{\sigma_0^2}$$

Hoe nader die fouttelling dus aan 0 kom, hoe nader is die betroubaarheid aan 1.00. Fouttellings is egter nooit totaal en al kontroleerbaar nie (verwys weer die bespreking by Figuur 5.2) en kan daarom nooit gelyk aan 0 wees nie. Dit beteken dus ook dat betroubaarheid nooit 1.00 kan wees nie. In enige metingsituasie word daar egter altyd gepoog om die fouttelling so laag moontlik te hou. Die mate waartoe dit bereik kan word, is ook die mate waartoe daar op akkurate wyse (dus betroubaar) 'n toetsling se werklike kennis of vaardighede gemeet kan word.

5.2.2 Betroubaarheidskoëffisiënte

Soos hierbo uiteengesit is daar in enige toetsituasie waarskynlik meer as een faktor wat tot metingsfoute sal bydra. As 'n groep studente byvoorbeeld 'n toets soos TAG aflê kan daar aanvaar word dat daar verskillende vaardigheidsvlakke is en dat daar variansie in hul uitslae ten opsigte van mekaar sal wees. Net so kan daar, soos reeds genoem, ook variansie wees as gevolg van makliker en moeiliker kontroleerbare omstandighede, byvoorbeeld hoe die toetsling die dag gevoel het, die tyd wat vir die toets toegelaat is (ongeveer 'n uur in TAG en TALL se geval), die inhoud van die toets, die formaat van die toets (meervoudige keuse items in TAG en TALL se geval), ensovoorts.

In die literatuur (cf. Alderson, Clapham & Wall, 1995; Bachman, 1990; Bachman & Palmer, 1996; Brown, 2004; Davies *et al.* 1999; Miller, Linn & Gronlund, 2009; Sluijter, 1998; Smit, 1991) word daar normaalweg tussen vier benaderings tot die bepaling van betroubaarheid onderskei wat elk 'n ander aspek van metingsfoute aanspreek. Hierdie vier benaderings is gemoeid met toetsstabiliteit (sogenaamde toets-hertoets-betroubaarheid), toetsekwivalensie (sogenaamde ekwivalente-vorm-betroubaarheid), interne konsistensie (sogenaamde inhoudsbetroubaarheid), en

betroubaarheid tussen beoordelaars. Tabel 5.1 is 'n voorlopige opsomming hiervan en is oorgeneem en aangepas uit Miller, Linn & Gronlund (2009:110) – dit word egter ook in meer besonderhede hieronder bespreek:

Tabel 5.1: Benaderings tot die bepaling van toetsbetroubaarheid

Metode	Tipe betroubaarheid	Prosedure
Toets-hertoets-betroubaarheid	Meting van stabiliteit	Selfde toets word twee keer op dieselfde populasie geadministreer met 'n tydsinterval tussenin
Ekwivalente-vorm- betroubaarheid	Meting van ekwivalensie	Twee vorme van die toets aan dieselfde populasie
Inhoudsbetroubaarheid	Meting van interne konsistensie	Een toets, verdeel resultate in twee gelyke dele en korreleer die twee toetsselftes
Inter-beoordelaar-betroubaarheid	Meting van konsistensie van beoordelaars	Een toets, twee of meer onafhanklike beoordelaars; beoordeel ooreenstemming

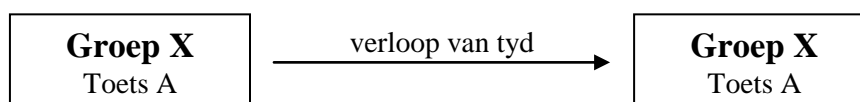
Laasgenoemde betroubaarheidsmeting, naamlik die inter-beoordelaar-betroubaarheid, word slegs gedoen in gevalle waar toetse byvoorbeeld met die hand nagesien word en beoordeling subjektief en nie meer objektief van aard is nie. Aangesien TAG 'n toets is wat volledig uit meervoudige keuse items bestaan wat deur 'n rekenaar nagesien word, is laasgenoemde betroubaarheidsmeting nie vir doeleindes van hierdie studie van belang nie en dit sal daarom nie verder bespreek word nie. Daar sal dus net op die eerste drie benaderings tot toetsbetroubaarheid gefokus word.

Voordat bogenoemde drie oorblywende metodes egter bespreek word, is dit ter agtergrond belangrik om kennis daarvan te neem dat die meeste indekse van toetsbetroubaarheid uitgedruk word as korrelasiekoëffisiënte en wat bekendstaan as die betroubaarheidskoëffisiënt (r_{tt}). 'n Betroubaarheidskoëffisiënt is dus die korrelasie tussen twee stelle gegewens verkry van dieselfde groep individue en is 'n aanduiding van die konstantheid van meting. Of, anders gestel, is dit die graad van verwantskap of ooreenstemming tussen twee stelle resultate wat veronderstel is om

dieselfde aspek, byvoorbeeld akademiese geletterdheid, te meet (Miller, Linn & Gronlund, 2009:109).

5.2.2.1 Toets-hertoets-betroubaarheid

Hierdie vorm van betroubaarheid is presies wat die naam sê, naamlik dat 'n toets op 'n groep studente geadminestreer word en na 'n tydperk word presies dieselfde toets weer geadminestreer, soos in Figuur 5.3 geïllustreer.



Figuur 5.3: Illustrasie van toets-hertoets-betroubaarheid

Die metode wat gebruik word om toets-hertoets-betroubaarheid te bepaal, is die korrelasie van een toetsuitslag met dieselfde toets na 'n verloop van tyd. Die sogenaamde Pearson-korrelasiekoëffisiënt is die statistiese prosedure wat hiervoor gebruik word. Die korrelasiekoëffisiënt tussen die twee toetse is dus 'n aanduiding van die mate waartoe individue oor tyd heen min of meer dieselfde resultate daarop behaal het. Hoe hoër die korrelasiekoëffisiënt, hoe hoër is die betroubaarheidskoëffisiënt - die betroubaarheidskoëffisiënt (r_{tt}) is volgens Smit (1991:29) die koëffisiënt van stabiliteit.

Daar moet verkieslik nie 'n te lang tydsverloop tussen die twee administrasies wees nie, aangesien daar 'n verskeidenheid van veranderlikes (blootstelling, intervensie, ontwikkeling van dekodeeringsvaardighede, ensovoorts) is wat 'n invloed op die betroubaarheid kan hê. Smit (1991:29) verwys in hierdie verband na die sogenaamde gedragsaspek (vir doeleindes van hierdie studie verwys gedragsaspek na akademiese geletterdheidsvermoëns) wat konstant behoort te bly met verloop van tyd. Hy voer aan dat

... indien daar aanvaar word dat 'n bepaalde gedragsaspek stabiel bly met verloop van tyd en die koëffisiënt van stabiliteit is laag, is dit onmoontlik om vas te stel of die aanname van stabiliteit van die gedragsaspek foutief is en of die lae koëffisiënt van stabiliteit slegs binne die konteks van die kennis van die bepaalde gedragsaspek wat gemeet is, geïnterpreteer moet word.

Wat hier van belang is, is dat resultate van metings en analyses nooit sonder meer patent is nie – dit is altyd, as die resultate van tegniese prosesse, onderworpe aan tegniese interpretasie. Daar word veral in Hoofstuk 8 weer na hierdie punt teruggekeer.

Bachman (1990:182) verduidelik dit aan die hand van die volgende voorbeeld: “[i]f there is a considerable time lapse between test administrations, some individuals’ [ability] may actually improve or decline more than that of others, causing them to perform differently the second time.” Sou twee administrasies egter weer te na aan mekaar wees, is daar altyd die risiko van oordrag van die een toetsgeleentheid na die volgende, wat volgens Smit (1991:29) uitgeskakel behoort te word. Volgens Bachman (1990:182) hang dit egter eerstens af van dit wat prakties uitvoerbaar is en tweedens van die oordeel van besluitnemers – die twee aspekte moet dus teen mekaar opgeweeg word alvorens daar met ’n weloorwoë besluit na vore gekom word, wat toetsresultate kontroleerbaar maak. Wanneer toetsresultate betroubaar is oor tyd heen (goeie toets-hertoets-betroubaarheid) is die fouttelling wat aan tyd toegeskryf kan word dus gekontroleer in soverre dit moontlik is. Dit kan soos in Tabel 5.2 opgesom word:

Tabel 5.2: Opsomming van die toets-hertoets-betroubaarheidskoëffisiënt

Vrae deur koëffisiënt beantwoord	Tipe fouttelling waarvoor voorsiening gemaak word
1. Hoe word resultate geaffekteer deur ’n identiese toets by ’n ander geleentheid op ’n identiese populasie te gebruik?	Tyd en geleentheid Populasiesamestelling
2. Hoe stabiel is die resultate op ’n spesifieke toets oor tyd heen?	Impak van intervensie

Vervolgens word enkele voorbeelde van toets-hertoets-betroubaarheidsmetings wat vir TAG en TALL gedoen is, voorgehou.

As eerste voorbeeld word eerstejaarstudente in die Fakulteit Ingenieurswese aan die Universiteit Stellenbosch se toets-hertoets-betroubaarheidsmeting vir 2005 voorgehou. Hierdie kohort studente het TAG en TALL afgelê aan die begin van die

jaar (Februarie) tydens die amptelike verwelkomingsweek van die Universiteit – die week voor klasse begin het. Na aanleiding van die toetsuitslae is hul op verskillende taalkursusse geplaas: studente met 'n hoë risiko moes 'n taalintervensie *Afrikaans 143* of *Engels 153* volg met aanvanklik 'n sterker klem op taalverwerwing, maar met 'n latere klemverskuiwing na akademiese geletterdheidsaspekte. Studente het nie 'n keuse gehad van watter kursus hulle wou volg nie; hulle is verplig om die kursus te volg vir die taal waarin hulle die hoogste risiko getoon het. Studente met lae risiko is verplig om die taalintervensie *Professionele Kommunikasie 113* te volg, waar die fokus is op kommunikasie vir die ingenieursprofessie, met elemente van akademiese geletterdheidsontwikkeling daarby ingebou. Omdat hierdie studie spesifiek met TAG gemoeid is, word daar slegs na studente wat die Afrikaanse kursusse gevolg het, gekyk.

In die onderstaande analise is daar deurlopend van gepaarde t-toetse gebruik gemaak - dit is gedoen om te bepaal of daar 'n beduidende toename in die gemiddelde prestasie van Februarie na November is, en of die rangorde min of meer dieselfde gebly het. Gepaarde waarnemings is gebruik, omdat al die studente in die kohort 'n voor- én 'n natoets afgelê het. Studente wat net een toets geskryf het, se toetspunte is nie in ag geneem nie. Daar is besluit om hierdie benadering te volg, aangesien dit die voordeel het dat daar deurlopend met dieselfde populasie gewerk word en dat verskille as gevolg van verskille in die populasie noodwendig uit die weg geruim word. Dit is dus 'n meer presiese toets as die nie-gepaarde toetse.

Eenkantige of tweekantige hipotesetoetse kan gedoen word. Vir laasgenoemde is die nulhipotese dat die gemiddeld van toets 1 gelyk is aan die gemiddeld van toets 2, oftewel $H_0: \text{gem}_1 = \text{gem}_2$. Die alternatiewe hipotese is dat die twee gemiddeldes verskil, oftewel, $H_A: \text{gem}_1 \neq \text{gem}_2$. Die eenkantige hipotesetoets word dus gebruik wanneer daar 'n sterk aanduiding is dat toets 2 se gemiddeld hoër is as toets 1 se gemiddeld. Die nulhipotese sien in hierdie geval as volg daar uit: $H_0: \text{gem}_1 \leq \text{gem}_2$, en die alternatiewe hipotese: $H_A: \text{gem}_1 < \text{gem}_2$. Vir doeleindes van hierdie analise is 'n eenkantige hipotesetoets gedoen, wat beteken dat die aanname vir die

nulhipotese was dat die gemiddeld in alle gevalle tussen die Februarie- en November-toetse dieselfde gebly het. Die alternatiewe hipotese was dat die gemiddeld vir die November-toets hoër moes wees as vir die Februarie-toets.

Die nulhipotese word verwerp indien die (absolute) berekende t-waarde (= t Stat) groter is as die (absolute) kritieke t-waarde, of anders gestel, die nulhipotese word verwerp as die berekende p-waarde kleiner is as die 95% vertrouensinterval / betekenispeil ($\alpha = 0.05$). Tabel 5.3 is 'n opsomming van die beskrywende statistiek vir daardie studente wat die Afrikaanse kursusse gevolg het (dus *Afrikaans 143* of *Professionele Kommunikasie 113* in Afrikaans). Let ook daarop dat Pearson se korrelasiekoëffisiënt telkens bereken is om die verband tussen die voor- en die natoetse te bepaal; hiervoor was die nulhipotese telkens dat daar geen korrelasie is nie, en die alternatiewe hipotese dat daar wel 'n beduidende korrelasie is, oftewel $H_0: \rho (\text{rho}) = 0$, en $H_A: \rho (\text{rho}) \neq 0$.

Tabel 5.3: Beskrywende statistiek vir studente wat taalintervensie in Afrikaans gevolg het (Afrikaans 143 én Professionele Kommunikasie 113 in Afrikaans)

	Afrikaans (Toets A: Februarie)	Afrikaans (Toets A: November)
Mean	66.32	70.18
Variance	438.09	348.53
N	111	111
DF	110.00	
t Stat	-3.80	
p(T<=t) one-tail	<0.001	
t Critical one-tail	1.66	
p(T<=t) two-tail	<0.001	
t Critical two-tail	1.98	

Pearson se korrelasiekoëffisiënt vir bogenoemde datastelle was 0.8708 met $p < 0.0001$ en daar is dus steeds 'n hoë en beduidende verband tussen die voor- en die natoets. Die hipotesetoets toon weer aan dat die nulhipotese verwerp is en dat

die alternatiewe hipotese bevestig is – daar is dus statisties gesproke ’n verskil in die gemiddeld tussen die voor- en die natoets.

Tabel 5.4 is ’n opsomming van die beskrywende statistiek vir daardie studente wat slegs *Afrikaans 143* gevolg het, met ander woorde daardie studente vir wie taal (lees Afrikaans) hoë risiko in die suksesvolle voltooiing van hul studies inhou.

Tabel 5.4: Beskrywende statistiek vir studente wat Afrikaans 143 gevolg het

	<i>Afrikaans 143</i> (Toets A: Februarie)	<i>Afrikaans 143</i> (Toets A: November)
Mean	35.04	44.13
Variance	221.52	233.48
N	24	24
DF	23	
t Stat	-3.60	
p(T<=t) one-tail	0.001	
t Critical one-tail	1.71	
p(T<=t) two-tail	0.002	
t Critical two-tail	2.07	

Pearson se korrelasiekoëffisiënt vir bogenoemde datastelle was 0.729 met $p < 0.0001$ en daar is dus steeds ’n hoë en beduidende verband tussen die voor- en die natoets. Die hipotesetoets toon weer aan dat die nulhipotese verwerp is en dat die alternatiewe hipotese bevestig is – daar is dus statisties gesproke ’n verskil in die gemiddeld tussen die voor- en die natoets.

Tabel 5.5 is ’n opsomming van die beskrywende statistiek vir daardie studente wat slegs *Professionele Kommunikasie 113* in Afrikaans gevolg het, met ander woorde daardie studente vir wie taal (lees Afrikaans) nie ’n hoë risiko in die suksesvolle voltooiing van hul studies inhou nie.

Tabel 5.5: Beskrywende statistiek vir studente wat Professionele Kommunikasie 113 in Afrikaans gevolg het

	<i>Professionele Kommunikasie 113 (Toets A: Februarie)</i>	<i>Professionele Kommunikasie 113 (Toets A: November)</i>
Mean	74.95	77.37
Variance	152.67	132.47
N	87	87
DF	86	
t Stat	-2.34	
p(T<=t) one-tail	0.011	
t Critical one-tail	1.66	
p(T<=t) two-tail	0.022	
t Critical two-tail	1.99	

Pearson se korrelasiekoëffisiënt vir bogenoemde datastelle was 0.676 met $p < 0.0001$ en daar is dus steeds 'n hoë en beduidende verband tussen die voor- en die natoets. Die hipotesetoets toon weer aan dat die nulhipotese verwerp is en dat die alternatiewe hipotese bevestig is – daar is dus statisties gesproke 'n verskil in die gemiddeld tussen die voor- en die natoets.

Uit die bostaande is dit eerstens duidelik dat daar 'n relatief hoë korrelasie tussen die aanvanklike akademiese geletterdheidstoets en 'n herhaling van dieselfde toets later in die jaar was vir al drie die groepe (*Afrikaans 143* en *Professionele Kommunikasie 113* gesamentlik, en *Afrikaans 143* en *Professionele Kommunikasie 113* afsonderlik). Al drie hierdie korrelasies was beduidend en daar kan dus met sekerheid gesê word dat daar 'n verband tussen die toetse is en dit stabiel oor tyd heen blyk te wees. In terme van die gemiddelde kan dit gestel word dat die nulhipotese, naamlik dat die natoets se gemiddeld statisties gesproke gelyk behoort te wees aan die voortoets se gemiddeld, telkens verwerp en die alternatiewe hipotese telkens bevestig is. Die styging in gemiddeldes was ook te verwagte, omdat daar hier sprake van 'n intervensie en blootstelling aan die akademiese omgewing was. Dit is egter nie 'n aanduiding van stabiliteit nie.

Die steekproefgroottes kon 'n invloed op die korrelasiekoëffisiënte gehad het, aangesien die aantal studente wat *Afrikaans 143* gevolg het slegs 24 was. Verder was hierdie groep studente ook besonder homogeen in samestelling (almal nie-moedertaalsprekers van Afrikaans; trouens, minstens 60% van die groep het Afrikaans as tweede addisionele of as vreemde taal gebesig). Die groep studente wat *Professionele Kommunikasie 113* gevolg het, was ook besonder homogeen in die opsig dat die meerderheid van hulle moedertaalsprekers van Afrikaans was en hulle het besonder lae tot selfs geen risiko getoon in terme van hul akademiese geletterdheidsvlakke nie – hier moet in ag geneem word dat studente wat *Professionele Kommunikasie 113* gevolg het, die toetse vrywillig in November kom aflê het en dit was in elk geval die top studente wat dit gedoen het. Hoewel daar dus tekens van betroubaarheid hier ter sprake is, is daar ter wille van sekerheid nog 'n steekproef gedoen op 'n meer heterogene en groter groep.

Eerstejaarstudente in die Fakulteit Natuurwetenskappe aan die Universiteit Stellenbosch is sedert 2007 almal verplig om die kursus *Wetenskapskommunikasievaardigheid 172* te volg. Alle studente, ongeag die risikovlak, volg dus hierdie kursus vir 'n jaar lank (twee periodes per week). Hierdie kursus is 'n akademiese geletterdheidskursus en die uitkomst daarvan is belyn met die konstruk van die akademiese geletterdheidstoets, soos in Hoofstuk 3 bespreek. Die 2008-kohort, wat hieronder bespreek sal word, het TAG en TALL reeds in 2007, tydens hul matriekjaar, afgelê as deel van die Universiteit se toegangsbeleid wat bepaal dat studente 'n kombinasie van toetse in die toegangstoetsbattery moet aflê wat saam met die destydse matriekpunte in berekening gebring word vir toelating (verwys weer Hoofstuk 1). Studente het 'n keuse of hulle hierdie kursus in Afrikaans of in Engels wil volg. Dieselfde statistiese tegnieke as hierbo is vir hierdie steekproef gebruik (hipotesetoetsing aan die hand van gepaarde t-toetse). Tabelle 5.6 en 5.7 hieronder is 'n opsomming van die beskrywende statistiek vir daardie studente wat die kursus in onderskeidelik Afrikaans en Engels gevolg het (dus alle studente wat beteken dat TAG én TALL ingesluit word).

**Tabel 5.6: Beskrywende statistiek vir alle studente wat
Wetenskapskommunikasievaardigheid 172 (WKV 172) in Afrikaans
gevolg het**

	<i>WKV 172 (AFR)</i> (Toets A: in 2007)	<i>WKV 172 (AFR)</i> (Toets A: November)
Mean	61.14	71.00
Variance	261.5	197.5
N	242	242
DF	241	
t Stat	-14.495	
p(T<=t) one-tail	<0.0001	
t Critical one-tail	2.342	
p(T<=t) two-tail	<0.0001	
t Critical two-tail	2.596	

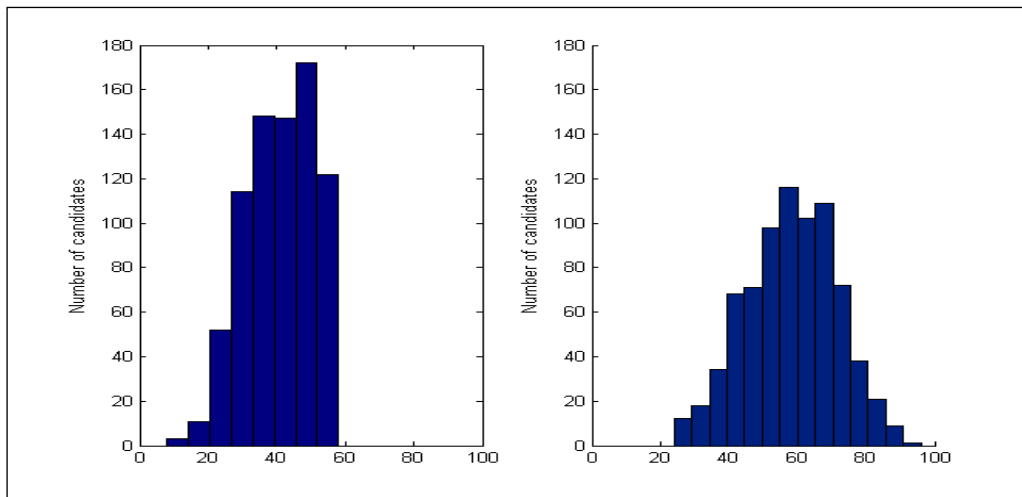
Pearson se korrelasiekoëffisiënt vir bogenoemde datastelle was 0.765 met $p < 0.0001$ en daar is dus hier ook 'n beduidende verband tussen die voor- en die natoets. Die hipotesetoets toon weer aan dat die nulhipotese verwerp is en dat die alternatiewe hipotese bevestig is – daar is dus statisties gesproke steeds 'n verskil in die gemiddeld tussen die voor- en die natoets.

**Tabel 5.7: Beskrywende statistiek vir alle studente wat
Wetenskapskommunikasievaardigheid 172 (WKV 172) in Engels gevolg
het**

	<i>WKV 172 (ENG)</i> (Toets A: in 2007)	<i>WKV 172 (ENG)</i> (Toets A: November)
Mean	69.19	75.59
Variance	317.39	220.11
N	192	192
DF	191	
t Stat	-7.761	
p(T<=t) one-tail	<0.0001	
t Critical one-tail	2.346	
p(T<=t) two-tail	<0.0001	
t Critical two-tail	2.6018	

Pearson se korrelasiekoëffisiënt vir bogenoemde datastelle was 0.770 met $p < 0.0001$, wat weereens 'n beduidende verband tussen die voor- en die natoets aantoon. Die hipotesetoets toon weer aan dat die nulhipotese verwerp is en dat die alternatiewe hipotese bevestig is – daar is dus statisties gesproke steeds 'n verskil in die gemiddeld tussen die voor- en die natoets, wat blyke daarvan is dat die intervensie daartoe kon bygedra het.

Die derde voorbeeld wat voorgelê word, is oorgeneem uit Van Dyk (2005). Hierdie kohort studente is ingeskryf aan die Universiteit van Pretoria (UP) en is verplig om 'n akademiese geletterdheidskursus (*EAG 110* en *EAG 120*) vir 'n jaar lank met twee periodes per week te volg. Hierdie kursus is ook belyn met die konstruk van die akademiese geletterdheidstoets (TAG en TALL). Studente wat vir hierdie kursus geregistreer is, is daardie studente wat hoë risiko in terme van taal vir die suksesvolle voltooiing van hul studies, aldus TAG of TALL, getoon het. Let daarop dat ander statistiese tegnieke as die wat hierbo gebruik is, vir die ontleding van die UP-data gebruik is. Figuur 5.4a en 5.4b (oorgeneem uit Van Dyk, 2005) is histogramme van die totale prestasie van dieselfde groep (hoë risiko) studente ($N = 769$). Figuur 5.4a is 'n aanduiding van die verspreiding van hierdie studente se akademiese geletterdheidsvermoëns in Januarie en Figuur 5.4b van die November-verspreiding. Hieruit is dit duidelik dat daar 'n verskuiwing na regs is van Januarie na November – die gemiddelde prestasie ('mean') in Januarie was byvoorbeeld 41% en in November was dit 58%, wat weer goed met die gemiddeld van die totale populasie wat die toets daardie jaar afgelê het, vergelyk. Daar is dus weereens blyke van die impak van die intervensie.



Figuur 5.4a: Januarie-prestasie

Figuur 5.4b: November-prestasie

Hoewel bostaande histogramme alreeds ’n aanduiding daarvan is dat daar ’n oorvleueling tussen die Januarie- en November-toets bestaan, is ’n hipotesetoets ook gedoen om te bepaal of daar ’n liniêre korrelasie tussen die twee datastelle was. ’n Positiewe resultaat is na aanleiding van die Pearson-korrelasie gevind ($N = 769$, $r = 0.654$, en $p \leq 0.001$). Hierdie resultate is in lyn met bostaande wat vir die Universiteit Stellenbosch gedoen is; daar kan dus verwag word dat prestasie in die November-toets beduidend sal korreleer met prestasie in die Januarie-toets. Wat veral weereens hier in ag geneem moet word (net soos by die analyses hierbo), is dat die groep studente wat in hierdie geval die intervensie ondergaan het slegs die data is van daardie studente wat ’n gefokusde intervensie ondergaan het. Die data van daardie studente wat nie gefokusde intervensie ontvang het nie, is nie by hierdie datastel ingesluit nie (sien ook die slotparagraaf van hierdie afdeling vir ’n meer breedvoerige verduideliking hiervoor), en daar kan nie met sekerheid uitsprake oor die toets-hertoets-betroubaarheid van TAG (en TALL) gemaak word nie.

Op hierdie punt is dit interessant om te let op ’n meer omvattende ondersoek van Weideman en Van der Slik (2008:161-182) na die stabiliteit van die toetsresultate van TAG en TALL oor verskeie administrasies heen. In hierdie artikel dui hulle juis aan dat ’n veelheid van veranderlikes in ag geneem moet word by die interpretasie van resultate. Die fokus van hierdie ondersoek was, in teenstelling met die analyses hierbo, op verskillende populasies, naamlik die van die Universiteit

Stellenbosch, die Noordwes Universiteit en die Universiteit van Pretoria. Hoewel dit dus nie dieselfde populasie was wat die toetse 'n tweede keer afgelê het nie, kan dit as toets-hertoets-betroubaarheid geag word, omdat dieselfde toets as't ware 'n tweede keer en selfs 'n derde keer by die onderskeie instellings afgelê is (verwys Tabel 5.2). Hul bevinding was dat beide TAG en TALL konsistente aanduidings gee van studente se taalvermoëns (Weideman & Van der Slik, 2008:167, 168, 173, 177).

Dieselfde outeurs (Van der Slik & Weideman, 2007:126-137) het kort voor bogenoemde studie onder andere ook ondersoek gedoen na die stabiliteit van TAG vanaf 2005 tot 2007, ten einde uitsprake te kan maak oor óf studente se taalvermoë jaarliks agteruitgaan, al dan nie, en óf TAG wel stabiel bly en die moeilikheidsgraad nie van jaar tot jaar wissel, en studente daarom onregverdig behandel word nie. Vir hierdie ondersoek het hulle van 'Item Response Theory' (IRT) gebruik gemaak, omdat daar genoegsame oorvleueling in terme van taaktipes en items was tussen die drie toetse. Die analyses het uitgewys dat dit blyk dat die toetse oor die periode van drie jaar stabiel gebly het, en so ook die akademiese taalvermoë van eerstejaarstudente. Van der Slik en Weideman (2007:135-136) wys egter uit dat daar 'n hele aantal veranderlikes is wat die resultate kon beïnvloed het en dat daar dus nie net sonder die nodige verdere ondersoek geargumenteer kan word vir of teen die stabiliteit van die toetse of akademiese geletterdheidsvlakke nie. In die verband stel hulle dit dat "the TAG and TALL tests have proved to be highly reliable over years" (Van der Slik & Weideman, 2007:136), maar hulle waarsku op dieselfde bladsy ook dat

test developers and ... users [should] be careful ... [to assume] equivalence among different versions of tests [and that we should, when interpreting test results] refer to test difficulty, which implies a measure of comparison that we may not yet have or may not even have planned for. We cannot simply interpret such results at face value.

Alhoewel daar in hierdie studies blyke is van toets-hertoets-betroubaarheid, maar dit nie noodwendig onomwonde gestel kan word dat die toetse in hierdie opsig betroubaar is nie, het die genoemde outeurs wel weer die kwessie van deursigtige en verantwoordbare, teoreties regverdigbare, besluitneming ter tafel gelê en die weg

gebaan vir die 2008-studie (sien hierbo). Beide hierdie studies spreek egter ook ten dele die kwessie van ekwivalente-vorm-betroubaarheid aan (sien 5.2.2.2 hieronder waarin daar nie weer hierna verwys sal word nie, omdat bostaande uiteensetting ook daarvoor voldoende is).

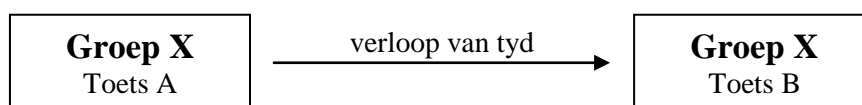
In die bostaande gedeelte is 'n verduideliking gebied van toets-hertoets-betroubaarheid en daar is drie voorbeelde bespreek, en na twee omvattende ondersoeke verwys, waaruit dit blyk dat TAG en TALL oor die potensiaal beskik om dit ten toon te stel, veral in die gevalle waar die steekproefgroottes voldoende was en ook as die veranderlikes (soos homogeniteit en intervensie) in ag geneem word. Die feit dat daar in die bogenoemde gevalle sprake van intervensies was, uitgesluit die ondersoeke van Weideman en Van der Slik, en uitsprake oor die toets-hertoets-betroubaarheid van TAG en TALL bevraagteken sou kon word, is wel waar. Die leser word egter weereens gewys op die argumente van Smit en Bachman in die inleidende deel tot hierdie afdeling oor toets-hertoets-betroubaarheid, waarin dit gestel word dat daar 'n redelike tydsverloop tussen die voor- en die natoets behoort te wees as daar sprake van 'n intervensie is uitsprake oor toets-hertoets-betroubaarheid uiteindelik gemaak wil word. Dit is veral gepas in 'n konteks waar daar etiese kwessies sou wees om sommige studente die intervensie te laat ondergaan en ander weer nie. Verder moet die uitspraak van Van der Slik en Weideman hierbo, naamlik dat daar versigtig omgegaan moet word met uitsprake oor onder andere stabiliteit ook in ag geneem word. In die volgende deel word ekwivalente-vorm-betroubaarheid bespreek.

5.2.2.2 Ekwivalente-vorm-betroubaarheid

Hierdie tipe betroubaarheid staan ook bekend as parallelle betroubaarheid en word gebruik wanneer daar bepaal moet word of twee vorme van 'n bepaalde toets (dus dieselfde konstruk, maar verskillende vrae) werklik ekwivalent aan mekaar is. Twee ekwivalente of parallelle vorme van dieselfde toets word dus op dieselfde groep, maar by verskillende geleenthede geadministreer (Smit, 1991:30). Ekwivalente-vorm-betroubaarheid is dus gemoeid met konstantheid van prestasie

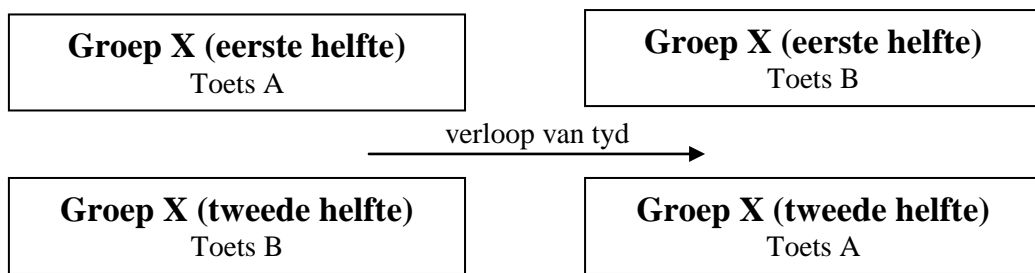
oor twee vorme van 'n bepaalde konstruk oor 'n tydperk heen. Die waarde van hierdie metode lê daarin dat dit kontrole uitoefen oor die fouttellings wat kan voorkom as gevolg van 'n wisseling in inhoud en die verloop van tyd.

Bachman (1990:183) wys daarop dat hierdie wyse van bepaling van betroubaarheid veral sinvol is in gevalle waar verskillende vorme van 'n toets vir sekuriteitsdoeleindes gebruik moet word, of waar die sogenaamde oefen-effek (leer met ander woorde hoe die toets saamgestel is en werk) teengewerk moet word ten einde uitsprake oor die vaardigheidsverbetering van toetslinge te kan maak wat nie aan toetsoefening toegeskryf kan word nie. Hierdie twee redes is natuurlik veral die geval by TAG, waar daar jaarliks van parallelle vorme vir sekuriteitsdoeleindes gebruik gemaak moet word en waar daar kwartaalliks van parallelle vorme gebruik gemaak moet word, aangesien die meerderheid van die studente wat TAG skryf ook die een of ander vorm van taalintervensie volg en hul vordering onder andere deur TAG gemonitor word. Figuur 5.5 is 'n grafiese voorstelling van die proses:



Figuur 5.5: Illustrasie van ekwivalente-vorm-betroubaarheid

Beide Smit (1991:30) en Bachman (1990:183) wys egter daarop dat die oefen-effek nie bloot geïgnoreer kan word wanneer daar uitsprake oor betroubaarheid gemaak word nie, aangesien dieselfde groep studente tog 'n soortgelyke vorm van die toets by 'n tweede, derde of selfs vierde geleentheid (soos die geval met TAG) aflê. Enige student kan dus by implikasie in die tweede vorm van die toets beter presteer as gevolg van die oefen-effek. 'n Metode wat gevolg kan word om dit uit te skakel is die van teenbalansering waar die helfte van die toetslinge byvoorbeeld vorm A by die eerste toetsgeleentheid aflê en die ander helfte vorm B. By die gebruik van ekwivalente vorms van die toets, sal oefeningseffekte nog voorkom, alhoewel in 'n mindere mate.



Figuur 5.6: Illustrasie van teenbalansering

Vir twee toetsvorme om parallel te wees is daar egter enkele voorwaardes waaraan dit moet voldoen, naamlik dieselfde groep toetslinge, dieselfde hoeveelheid vrae, uniformiteit (toetsinhoud, fasiliteitswaarde oftewel moeilikheidsgraad, en verteenwoordigendheid van die konstruk), dieselfde wyse van vraagstelling (formaat), dieselfde nasienprosedures, en 'n soortgelyke verspreiding van prestasie (Smit, 1991:30). Ten einde die ekwivalente-vorm-betroubaarheid van die toetse te meet, word die standaardafwykings vir elk van die twee vorme bereken en vergelyk om die ekwivalensie tussen die twee te bepaal, waarna 'n korrelasie tussen die twee datastelle (toets A en toets B) getrek word. Die korrelasie word dan geïnterpreteer as 'n indikasie van die ekwivalensie van die twee toetse, of anders gestel as 'n aanduiding van betroubaarheid. Indien daar 'n hoë korrelasie is, is dit 'n aanduiding daarvan dat die effek van wisseling (spesifiek wisseling in inhoud) en van tyd tot 'n groot mate gekontroleerd is en dat die fouttelling laag is, en daarom is die toets betroubaar. Hierdie metode van betroubaarheidsmeting is dus by implikasie ook 'n vorm van toets-hertoets-betroubaarheid, maar die fokus lê elders. Dit kan soos in Tabel 5.8 opgesom word:

Tabel 5.8: Opsomming van die ekwivalente-vorm-betroubaarheidskoëffisiënt

Vrae deur koëffisiënt beantwoord	Tipe fouttelling waarvoor voorsiening gemaak word
1. Hoe konsistent is resultate ongeag die vorm wat gebruik is?	Tyd en geleentheid Populasiesamestelling
2. Hoe stabiel is die resultate oor tyd heen?	Impak van intervensie Effek van inoefening Inhoud

Aangesien TAG tot dusver nog hoofsaaklik weens sekuriteitsredes in 'n ontwikkelingsfase was en dus van toets tot toets verskil in terme van 'n aantal voorwaardes (sien hieronder), is daar nog geen ondersoek van hierdie aard gedoen nie. Die vermoede is egter daar dat dit oor die volgende twee tot drie jaar wel moontlik sal wees, aangesien daar voldoende toetsmateriaal beskikbaar blyk te wees wat ook aan die genoemde voorwaardes sal voldoen. Daar sal egter op sistematiese wyse hiermee omgegaan moet word: daar sal eerstens toegesien moet word dat kwartaaltoetse nie meer loodstoetse is nie, maar wel geformaliseerde toetse, sodat die een kwartaaltoets vergelykbaar met die ander is in terme van die hoeveelheid vrae, uniformiteit (toetsinhoud, fasiliteitswaarde oftewel moeilikheidsgraad, en verteenwoordigendheid van die konstruk), dieselfde wyse van vraagstelling (formaat), dieselfde nasienprosedures, en 'n soortgelyke verspreiding van prestasie. Ten tweede sal daar na oorvleueling tussen afdelings gekyk moet word in terme van die voorwaardes hierbo gestel. In die derde plek sal die verskillende vorme van die toets op dieselfde populasie uitgetoets moet word. Dit kan gedoen word deur opsommings, soos geïllustreer in Tabel 5.9 hieronder te maak en dan afdelings met mekaar te vergelyk. 'n Verdere moontlikheid is om sogenaamde toetsekwivalensie te bepaal, waar ongeveer 30% van die items van een toets oorvleuel met die van 'n ander toets, soos gedoen in die ondersoek van Van der Slik en Weideman (2007) waarna hierbo verwys is – die fokus van die genoemde ondersoek was egter nie op ekwivalente-vorm-betroubaarheid nie, hoewel dit gemaklik uitgebrei sou kon word om ook dit te bepaal. Eers dan sal parallelle betroubaarheid bereken kan word.

Tabel 5.9: Voorbeeld van opsomming van twee afdelings van TAG 2005 en TAG 2006 vir doeleindes van bepaling van ekwivalente-vorm-betroubaarheid

	AFDELING 1 Skommelteks	AFDELING 3 Woordeskat
TAG 2005	Vrae 5	Vrae 10
	Uniformiteit FV = X Inhoud = X% oorvleueling Konstruk = Elemente X, Y, Z	Uniformiteit FV = X Inhoud = X% oorvleueling Konstruk = Elemente X, Y, Z
	Formaat Meervoudige keuse, 5 opsies	Formaat Meervoudige keuse, 4 opsies
	Nasien Objektief – rekenaarmatig	Nasien Objektief – rekenaarmatig
	Verspreiding Normaal met StdAfw van X	Verspreiding Normaal met StdAfw van X
		AFD 1: Skommelteks
TAG 2006	Vrae 5	Vrae 10
	Uniformiteit FV = X Inhoud = X% oorvleueling Konstruk = Elemente X, Y, Z	Uniformiteit FV = X Inhoud = X% oorvleueling Konstruk = Elemente X, Y, Z
	Formaat Meervoudige keuse, 5 opsies	Formaat Meervoudige keuse, 4 opsies
	Nasien Objektief – rekenaarmatig	Nasien Objektief – rekenaarmatig
	Verspreiding Normaal met StdAfw van X	Verspreiding Normaal met StdAfw van X

Alhoewel dit, soos reeds genoem, nog nie moontlik was om hierdie tipe betroubaarheid uitvoerig te ondersoek nie, blyk dit uit 'n voorlopige analise dat daar 'n korrelasie is van 0.48 met 'n p-waarde van $p \leq 0.0095$ vir drie afdelings (skommelteks, woordeskat en tekstipe) wat met mekaar vergelykbaar was in terme van die voorwaardes hierbo gestel. Op grond van hierdie voorlopige analise blyk

dit egter wel dat daar tekens van ekwivalente-vorm-betroubaarheid by TAG ter sprake is. Die korrelasiematriks volg in Tabel 5.10.

Tabel 5.10: Korrelasiematriks

	Skommel1	WSkat1	Tekstipe1	Totaal1	Skommel2	WSkat2	Tekstipe2	Totaal2
Skommel1	1							
WSkat1		0.21	0.16	0.59	0.26	0.19	0.20	0.27
Tekstipe1			1	0.88	0.30	0.39	0.23	0.43
Totaal1				1	0.20	0.20	0.20	0.26
Skommel2					1			
WSkat2						0.36	0.20	0.63
Tekstipe2							1	0.92
Totaal2								1

Hoewel die genoemde korrelasie nie besonder hoog is nie, is dit te verwagte, aangesien hierdie slegs drie afdelings van die volledige toets insluit en dus nie die volledige konstruk meet nie. Verder moet daar in ag geneem word dat daar 'n tydsverloop van minstens ses maande tussen die twee metings was, met 'n gefokusde intervensie – daar sou deur sommige hieroor geredeneer kon word dat as daar nie 'n intervensie was nie, die korrelasie nog hoër sou kon wees. Die feit dat dit egter beduidend is, veral as in ag geneem word dat dit slegs vir ongeveer 'n derde van die totale toets is, is wel noemenswaardig.

5.2.2.3 Interne konsistensie / Inhoudsbetroubaarheid

In teenstelling met die twee metodes hierbo bespreek, maak die berekening van interne konsistensie slegs gebruik van een toets se data om 'n betroubaarheidswaarde te bepaal. In werklikheid word die betroubaarheid van die instrument bereken deur te bepaal of daardie items wat die konstruk (behoort te) verteenwoordig, min of meer dieselfde resultate sal oplewer wanneer dit teen mekaar afgespeel word. Dit gaan dus oor hoe stabiel toetslinge se prestasie oor verskillende dele / afdelings van die toets bly – prestasie op een afdeling van 'n toets behoort teoreties in lyn te wees met 'n ander afdeling van die toets. Interne konsistensie word ook soms na verwys as die inhoudsbetroubaarheid.

Gesofistikeerde sagteware soos SPSS, Statistica, TiaPlus, en Iteman word deesdae vir hierdie tipe analyses gebruik. Hierdie pakkette bepaal die gemiddelde korrelasie tussen alle moontlike helftes van die toets (Robinson-Kurpius & Stafford, 2006:127). In kort gaan dit dus daaroor dat daar bepaal word of, en in welke mate, elke enkele item met elke ander item in dieselfde toets korreleer.

Smit (1991:32) stel dit dat inhoudsbetroubaarheid betrekking het op die homogeniteit van die toets en dat homogeniteit slegs bepaal kan word “deur ’n ontleding van die interne konsekwentheid van prestasie op items van die toets”. Hoe meer homogeen ’n toets dus in teorie is, hoe hoër is die betroubaarheid. Dieselfde outeur stel dit egter ook dat homogeniteit nie ’n vereiste vir betroubaarheid is nie, aangesien dit gaan oor dit wat die toets behoort te meet en dit kan wees dat ’n toetskonstruk byvoorbeeld meer heterogeen daar uitsien. Van der Slik en Weideman (2005:26) voer in hierdie verband aan dat TAG en TALL se konstruk juis multidimensioneel van aard is en dat daar dus nie nêr na die homogeniteit van toetsitems gekyk kan word in die bepaling van betroubaarheid nie. Dit is in lyn met Smit (1991:32) se argument dat toetse ook heterogeen kan wees as die kriterium wat die toets wil voorspel heterogeen van aard is, en, wat meer is, dat dit in sulke gevalle sinvol is om ’n toets in subtoetse of verkillende afdelings te verdeel, soos die geval met TAG en TALL.

In die literatuur word daar onder andere onderskei tussen die volgende mees algemene metodes om interne konsistensie te bepaal: die halfverdelingsmetode (‘split-half reliability’), die Kuder-Richardson-metode (inter-itemkorrelasie en / of ’n item-totaalkorrelasie), en die bepaling van Cronbach se alpha (α). Elkeen van hierdie metodes word vervolgens in meer besonderhede bespreek en daar sal waar van toepassing na geleenthede en voorbeelde verwys word waar hierdie metodes by TAG en TALL gebruik is.

5.2.2.3.1 Die halfverdelingsmetode

Die halfverdelingsmetode verdeel alle items wat veronderstel is om verteenwoordigend van dieselfde konstruk te wees lukraak in twee dele, waarna die

korrelasie tussen die prestasies op die twee helftes van die toets bereken word (Bachman, 1990:172). Die twee helftes word dus hanteer asof dit parallelle toetse is en daar moet met bepaalde aannames oor parallelle toetse gewerk word, naamlik dat die gemiddeldes en die variansie gelyk aan mekaar moet wees. Hierdie benadering is volgens Bachman (1990:173) veral gewens by toetse wat in meervoudige keuse formaat opgestel is, waar daar met die aanname gewerk kan word dat 'n lukrake seleksie van items uiteindelik twee toetse sal voortbring wat ekwivalent, maar ook onafhanklik van mekaar is. Wat die onafhanklikheidsaspek betref, voer Bachman (1990:172-173) aan dat

... we must assume that the two halves are independent of each other. That is, that an individual's performance on one half does not determine how he performs on the other. This assumption does not mean that the two halves will not be correlated with each other, but rather that the correlation will be due to the fact that performance on one half depends upon performance on the other. ... In order to interpret the relationship between the halves as an indication that they both measure the same ability, or that they both have the same true score variance, we must be able to rule out the second interpretation.

Smit (1991:33) en Bachman (1990:173) wys ook daarop dat daar omsigtig met hierdie metode omgegaan moet word en dat die toets nie bloot in twee helftes ('n eerste en 'n tweede helfte) opgedeel moet word nie, omdat toetse normaalweg met makliker items begin en dit opbou in moeilikheidsgraad na die einde van die toets – daar moet dus gepoog word om ekwivalente helftes, ook in terme van moeilikheidsgraad, met mekaar te korreleer. Die standaardprosedure is gevolglik om alternatiewe items aan die twee helftes toe te wys deur elke alternatiewe toetsitem aan 'n helfte toe te wys. Dit is egter ook nie noodwendig onproblematies nie, omdat so 'n verdeling kan meebring dat die parallelle vorme nie noodwendig ekwivalent is in terme van die inhoud wat gemeet word nie en daarom moet die verdeling eerder op basis van die inhoud gedoen word. In die geval van TAG en TALL is veral hierdie aspek van toepassing, aangesien die verskillende afdelings telkens ander aspekte van die konstruk meet. In hierdie geval blyk dit sinvol te wees om eers die toets, gebaseer op die afdelings, in subtoetse te verdeel en daarna 'n halfverdelingseksperiment uit te voer. So 'n benadering is egter ook nie sonder probleme nie, aangesien die subtoetse beperk is in terme van die aantal items en dus reeds tot 'n lae betroubaarheidsindeks sal lei. 'n Selfs nog groter probleem, soos

tipies die geval by TAG en TALL, is dat die items nie noodwendig onafhanklik van mekaar of konteks-vry is nie, soos die geval met skommeltekste, die leesbegrip- en die teksredigeringsafdelings. In gevalle soos hierdie waar daar nie sekerheid is dat die onderskeie items dieselfde aspek van byvoorbeeld akademiese geletterdheidsvermoëns meet nie, en waar items nie onafhanklik van mekaar is nie, is dit verkieslik dat toets-hertoets-betroubaarheid- en ekwivalente-vorm-betroubaarheidsmetings eerder gedoen word.

Bachman (1990:174) wys egter ook daarop dat die metode nie bloot net geïgnoreer kan word nie, omdat dit soms gewoon net nie moontlik is om hertoetse te doen of om ekwivalente vorme van 'n toets te administreer nie. Noodgedwonge moet daar dus soms die een of ander vorm van interne konsistensie bepaal word. Hy voer daarom aan dat "... we must make every attempt to split the test into halves in such a way as to maximize their equivalence and their independence."

Indien bogenoemde op TAG en TALL van toepassing gemaak word en daar byvoorbeeld net met die afdelings wat oor onafhanklike items beskik, soos die grafiese en visuele afdeling, die woordeskatafdeling en die afdeling oor styl en register, vir die halfverdelingsmetode gewerk word, is 'n verdere probleem dat daar uiteindelik met twee halwe of selfs kleiner toetse, en nie twee volle toetse nie, se prestasies gewerk word – die volle konstruk vir akademiese geletterdheid word dus nie getoets nie en die betroubaarheid gaan dus noodgedwonge daaronder ly.

'n Algemeen aanvaarde beginsel van betroubaarheid is dat die aantal vrae en die grootte van die steekproef 'n beduidende invloed op die betroubaarheid het. Om hierdie aspek teen te werk, moet daar dus 'n korreksie op die betroubaarheidskoëffisiënt aangebring word om te korrigeer vir die halvering van die toets, of soos Smit (1991:33) dit stel "om 'n indeks te verkry soos wanneer twee volle lengtes gebruik word." Alhoewel TAG en TALL dus uiteindelik in selfs kleiner dele as net twee helftes verdeel sal moet word om hierdie betroubaarheidsmeting te kan doen, geld dieselfde wiskundige beginsels as vir twee toetse wat ekwivalent en onafhanklik sal wees. As 'n korreksie verder nog

aangebring word, sal die betroubaarheidskoëffisiënt aansienlik daal en dit is daarom nie sinvol om hierdie roete vir die bepaling van betroubaarheid by TAG en TALL te volg nie. Die twee metodes waarvolgens 'n sogenaamd gekorrigeerde betroubaarheidsindeks bereken kan word, naamlik die Spearman-Brown-metode en die Guttman-metode, word gevolglik nie hier verder bespreek nie.

5.2.2.3.2 Die Kuder-Richardson-metode

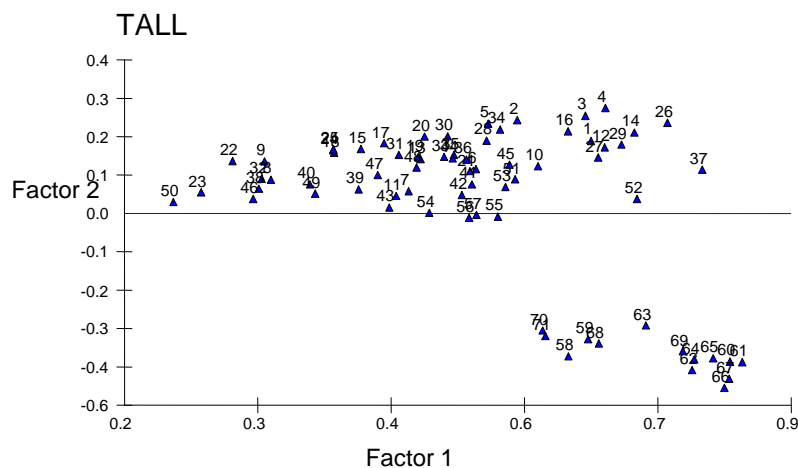
Die Kuder-Richardson-metode is net soos die halfverdelingsmetode 'n manier om by 'n enkele toetsgeleentheid met 'n enkele toets die betroubaarheid daarvan te bepaal deur ondersoek in te stel na die response van die toetslinge op die onderskeie items (Smit, 1991:35). Hierdie metode ondersoek dus die sogenaamde inter-itemkonsistensie en nie die resultate van die twee helftes van 'n toets, soos by die halfverdelingsmetode, nie. Hierdie metode word normaalweg gebruik by toetse wat digotomies van aard is (dus teoreties met waardes van 1 of 0) en gee 'n aanduiding van ekwivalensie deur inter-itemkonsistensie te bepaal. Indien die items redelik homogeen is, word daar 'n konsekwente en konstante prestasie van item tot item verwag en 'n hoë inter-itemkorrelasie sal dus verkry word. Thorndike en Hagen (1969:185) som dit as volg op:

The essential assumption in the procedure is that the items within one form of the test have as much in common with one another as do the items in that one form with the corresponding items in a parallel or equivalent form. This means that the items in a test are homogenous in the sense that every item measures the same general factors of ability or personality as do the others.

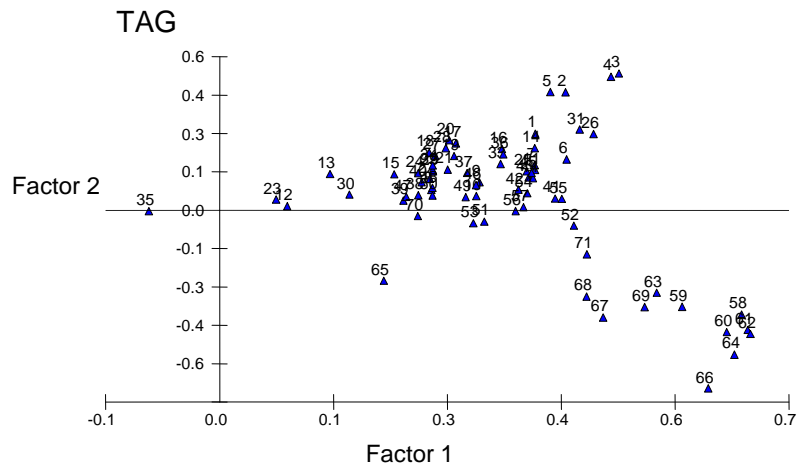
In die geval van TAG en TALL is dit egter nie so eenvoudig nie. Van der Slik en Weideman (2005:29-31) argumenteer in die verband dat die konstruk van TAG en TALL so ryk en omvattend is, dat 'n heterogene stel items wel aanvaarbaar is, en dat die kriterium vir dit wat die toets wil voorspel meer heterogeen as homogeen van aard is (Smit, 1991:32). Dit word juis verdiskonteer deur die toets in afdelings/subtoetse te verdeel deur homogeniteit in die subtoetse te probeer bewerkstellig terwyl heterogeniteit in die toets as geheel geakkommodeer word. Figuur 5.7a en 5.7b is 'n illustrasie hiervan wat uit Van der Slik en Weideman (2005:30) oorgeneem is. Uit die faktoranalises vir TALL en TAG onderskeidelik kan dit duidelik gesien word dat die meerderheid van die items kluster

(homogeniteit dus), maar dat daar uitskieters is, veral by die laaste afdeling van die toetse, naamlik teksredigering. Dieselfde outeurs het dit toe verder gaan ondersoek, omdat dit daarop sou kon dui dat die laaste afdeling van die toetse bloot swak is. Hulle het egter bevind dat die heterogeniteit wat hierdie items toon weglaatbaar klein is en dat elkeen van die items op hul eie goed funksioneer, met aanvaarbare moeilikheidsgraad en diskriminasiewaardes. Hoewel dit dus mag lyk asof daar twee faktore/domeine is wat gemeet word (die teksredigeringsafdeling naamlik die een faktor en die res van die toets die ander faktor), hang dit af van hoe streng die assosiasie is wat tussen die twee faktore behoort te bestaan – in 'n heterogene konstruk is 'n streng assosiasie vanselfsprekend nie 'n vereiste nie. Van der Slik en Weideman (2005:32) en artikuleer hul finale afleidings hieroor as volg

... the test designers are ... faced with a choice between the efficiency of having a completely homogeneous test and the appropriateness of the particular task type for measuring academic literacy. Such choices are often the basis of trade-offs that test developers have to make. In the present case, the arguments for a rich and varied construct are the most important consideration. The TALL/TAG designers claim to have designed an instrument that specifically acknowledges that academic discourse is an acquired, secondary discourse (cf. Gee 1998), with its own requirements and text types or genres, and have articulated the construct so as to capture this particular variety of language (Van Dyk & Weideman 2004a). If some degree of heterogeneity is what is required to ensure contextual appropriateness, it may have to be tolerated.



Figuur 5.7a: Faktoranalise vir TALL



Figuur 5.7b: Faktoranalise vir TAG

Uit die bostaande kan dit afgelei word dat daar hoë inter-itemkorrelasie (homogeniteit) by beide TAG en TALL ter sprake is, wat die betroubaarheid daarvan verhoog.

By hierdie afdeling is dit duidelikheidshalwe nodig om kennis te neem daarvan dat die Kuder-Richardson-metode en Cronbach se alpha (sien ook die volgende afdeling) in der waarheid identies aan mekaar is, in die geval van dichotome items, soos by TAG en TALL. Cronbach se alpha sluit dus in werklikheid reeds die bewerkings van die Kuder-Richardson-metode in.

5.2.2.3.3 Cronbach se alpha (α)

'n Ander metode om die interne konsistensie van 'n enkele toets te bepaal, en wat deurlopend vir TAG en TALL gebruik word, is die bepaling van Cronbach se alpha wat met die simbool α aangedui word. Hierdie metode vir die bepaling van betroubaarheid het uit die Kuder-Richardson-metode ontwikkel en vereis ook nie dat die toets letterlik verdeel word soos by die halfverdelingsmetode nie. Dit maak ook nie van 'n suiwer digotomiese benadering gebruik nie en kan waardes anders as 0 en 1 in ag neem by die bepaling daarvan. Cronbach se alpha gaan dus daarvoor dat die gemiddeld van alle moontlike halfverdelings bepaal word. Net soos by die bogenoemde metodes is homogeniteit by die bepaling van alpha dus van belang – daar moet bepaal word of die onderskeie items dit meet wat die konstruk van die

toets veronderstel. Nitko (2004:66) voer aan dat “[i]f assessment tasks are homogeneous, the KR20 and coefficient alpha procedures will give nearly the same results as the split-halves procedure”. Binne konteks van die bostaande verduideliking van ’n ryk en heterogene konstruk in die breë, maar wel ’n homogene konstruk in die klein, blyk dit gepas te wees om te volstaan by hierdie meting van betroubaarheid vir TAG en TALL. Nitko (2004:66) en Miller, Linn en Gronlund (2009:115) voer egter verder aan dat die KR20 en Cronbach se alpha by heterogene toetse normaalweg laer is as by homogene toetse.

Wat besonder interessant en selfs indrukwekkend van TAG en TALL is, is dat dit konsekwent hoë alpha-waardes toon, veral as dit aanvaar word dat toetse met laebelang ’n laer alpha kan hê en hoërbelangtoetse beslis ’n hoër alpha (Smit, 1991:43-44; Robinson Kurpius & Stafford, 2006:127-129). Miller, Linn en Gronlund (1999:131-132) stel dit dat die aanvaarbaarheid van ’n betroubaarheidswaarde in ’n verskeidenheid van oorwegings, in die vorm van vrae, lê: (i) hoe belangrik is die besluit wat op grond van die toetsresultate geneem moet word; (ii) hoe omkeerbaar of finaal is hierdie besluit; (iii) hoe verreikend is die gevolge van die besluit wat geneem moet; (iv) wie beïnvloed dit – die individu of die groep; en (v) hoe langdurig is die impak van die besluit wat geneem moet word? Dieselfde outeurs verduidelik dit as volg:

For irreversible decisions that are apt to have great influence on the lives of individual students, we should make stringent demands on the reliability of the assessments we use. For lesser decisions, especially for those that can be later confirmed or reversed without serious consequences, we should be willing to settle for less reliable measures. Teacher-made tests commonly have reliabilities between 0.60 and 0.85, but they are useful for the types of instructional decisions typically made by teachers. Thus, the degree of reliability required depends largely on how confident we need to be about the decision being made. Greater confidence requires higher reliability.

Binne hierdie konteks is daar besluit dat TAG en TALL altyd ’n betroubaarheidskoëffisiënt van minstens 0.80 behoort te hê. Hierdie koëffisiënt is aanvaarbaar omdat hierdie toetse hoofsaaklik vir plasingsdoeleindes gebruik word en dus laebelangtoetse is. Die feit dat die Universiteit Stellenbosch (US) egter hierdie toetse as deel van hul toegangstoetsbattery gebruik en dit onmiddellik toetse

met hoër belang word, verander die prentjie egter en daar is besluit om die parameters op te skuif na 0.85. Die rede waarom dit nie na 'n waarde van byvoorbeeld 0.90 verhoog is nie, is omdat die beveiliging daarin lê dat dit as deel van 'n battery gebruik word – die finale besluit word dus nie op TAG of TALL se resultaat alleen geneem nie, maar op 'n kombinasie van toetse in die battery (verwys Bylaag B). Boonop is die Universiteit Stellenbosch se toegangsmodel van so aard dat die toegangstoetse in 'n 40:60 verhouding met die nasionale senior sertifikaat vir doeleindes van toelating gebruik word, wat verder beveiliging vir 'n ietwat laer, hoewel steeds hoë, betroubaarheidskoëffisiënt bied. Tabel 5.11 bied 'n uiteensetting van die alpha-waardes van TAG en TALL vir die Universiteit van Pretoria (UP) se toetsadministrasies sedert hierdie toetse in gebruik geneem is. Die ander universiteite se alpha-waardes word nie hier weergegee nie, omdat hulle die toetse eers later in gebruik geneem het en aanvanklik net vir 'n loodsfase. Let egter daarop dat daar 'n hoogs beduidende korrelasie is tussen die alpha-waardes van die UP en die Universiteit van Stellenbosch, en so ook tussen die van UP en die Noordwes Universiteit, en die Universiteit Stellenbosch en die Noordwes Universiteit. Dieselfde prentjie ontvou dus by al die universiteite wat van TAG en TALL gebruikmaak.

Tabel 5.11: Uiteensetting van alpha-waardes van TAG en TALL sedert 2004

	TAG	TALL
2004	$\alpha = 0.86$	$\alpha = 0.95$
2005	$\alpha = 0.83$	$\alpha = 0.93$
2006	$\alpha = 0.87$	$\alpha = 0.94$
2007	$\alpha = 0.88$	$\alpha = 0.94$
2008	$\alpha = 0.83$	$\alpha = 0.94$
2009	$\alpha = 0.81$	$\alpha = 0.94$
2010	$\alpha = 0.84$	$\alpha = 0.93$

Dit is interessant om hier daarop te let dat die betroubaarheidskoëffisiënte vir TAG deurlopend laer is as vir TALL. Dit kan daaraan toegeskryf word dat heterogeniteit of homogeniteit in die toetspopulasie 'n invloed op die betroubaarheid van die meting het. So sien die populasie wat TAG skryf byvoorbeeld baie meer homogeen

daar uit (feitlik almal is eerstetaalsprekers van Afrikaans), en die wat TALL aflê meer heterogeen (daar is 'n beduidende aantal tweede en selfs addisionele taalsprekers, naas die eerstetaalsprekers, wat TALL jaarliks aflê – verwys Figuur 1.1 ter illustrasie). Lyman (1991:26) bevestig dit wanneer hy sê dat “[t]he variability of the group tested is also important in evaluating any reliability coefficient. If everything else is the same, higher reliability coefficients will be found for groups which vary more in ability.”

'n Ander aspek wat volgens Bachman (1990:187) uit die bespreking van Cronbach se alpha voortvloei het, en wat 'n invloed op die betroubaarheid van 'n meetinstrument het, is die teorie van veralgemeenbaarheid ('generalizability theory'), ook G-teorie genoem. In kort gaan dit daaroor dat veralgemenings slegs gemaak kan word as daar sekerheid is oor watter gevolgtrekkings deur toetsdata ondersteun word, watter gevolgtrekkings die toetsdata ongeldig maak en watter gevolgtrekkings die toetsdata nóg ondersteun nóg ongeldig maak. Dit gaan dus oor, soos Nitko (2004:53) dit stel, hoe wyd toetsdata geïnterpreteer kan word, oftewel dat daar 'n veelheid van faktore is wat 'n toetsresultaat kan beïnvloed (hierbo onder fluktuasie bespreek). Hierdie faktore (wat normaalweg aan die regulatiewe kant van die teoretiese raamwerk van hierdie studie lê) het gevolglik 'n beperkende effek op betroubaarheidskoeffisiënte (wat weer aan die konstitutiewe kant van die teoretiese raamwerk lê). In praktyk beteken dit dat die toetsontwikkelaars van TAG en TALL byvoorbeeld 'n toetsresultaat moet kan interpreteer en daarvolgens kan veralgemeen deur uitsprake te maak soos dat Student X se toetsresultaat 'n aanduiding is van sy / haar vermoëns in die domein as geheel (in hierdie geval letterlik alles van wat die domein van akademiese geletterdheid moontlik kan veronderstel). Bachman (1990:187) verduidelik dit dat

G-theory treats a given measure or score as a sample from a hypothetical universe of possible measures. When we interpret a test score we generalize from a single measure to a universe of measures. In other words, on the basis of an individual's performance on a test we generalize to her performance in other contexts. The more reliable the sample of performance, or test score, is, the more generalizable it is. Reliability, then, is a matter of generalizability, and the extent to which we can generalize from a given score is a function of how we define the universe of measures. And the way we define a given universe of

measures will depend upon the universe of generalization – the decisions or inferences we expect to make on the basis of the test results.

Daar word wiskundig voorsiening gemaak vir toetse van veralgemeenbaarheid in die formule vir die ‘analysis of variance’ (ANOVA) waar verskillende faktore soos ras, geslag, ouderdom, opvoedingsagtergrond, sosio-ekonomiese omstandighede, ensovoorts verreken kan word. Vir doeleindes van die bepaling van ’n betroubaarheidskoëffisiënt vir TAG en TALL is dit wel al aan die hand van die sogenaamde ‘Differential Item Functioning’ (DIF)-analises gedoen, maar dit word uiters genuanseerd en selfs versigtig aangebied, omdat die huidige politieke bestel in Suid-Afrika gelyke beregtiging vir alle groeperinge, op watter vlak ook al, voorstaan. Daar word dus polities korrek opgetree deur dit nie in ag te neem by enige besluitneming nie. Indien ander aspekte, soos sosiale verantwoordbaarheid en regverdiging, wat aan die regulatiewe kant van die teoretiese raamwerk van hierdie studie lê, egter in ag geneem word, sou dit wel nodig wees om hierdie aspek in meer besonderhede te ontrafel. Vir doeleindes van hierdie studie, waar die fokus op die konstitutiewe kant van die teoretiese raamwerk is, is dit voldoende om bloot kennis te neem van G-teorie en die invloed wat dit kan hê op betroubaarheid. Dit kan wel sinvol wees om vir in verdere navorsing ’n meer volledige studie hiervan te maak.

5.3 Die standaardmetingsfout

Reeds in paragraaf 5.2 hierbo is daar aangetoon dat daar in die berekening van die betroubaarheidskoëffisiënt voorsiening gemaak moet word vir fluktuasie wat tot metingsfoute bydra as gevolg van die verskille tussen die waargenome tellings en die ware tellings. Geen bespreking van betroubaarheid is dus volledig sonder ’n bespreking van metingsfoute nie – die standaardmetingsfout (‘standard error of measurement: SEM’) is trouens direk verwant aan betroubaarheid, veral ook as G-teorie in ag geneem word, omdat G-teorie in wese op fouttellings fokus. In kort kom dit neer op die volgende: hoe kleiner die SEM, hoe hoër die betroubaarheid. Robinson-Kurpius en Stafford (2006:132) stel dit dan ook as volg: “[a] standard error of measurement is a deviation score and reflects the area around an obtained

score where you would expect to find the true score. This area is called the confidence interval.”

Dit kom daarop neer dat die waargenome telling ’n akkurate weergawe (dus ’n hoë vertrouensinterval) van die ware telling moet wees met ’n baie lae fouttelling. Wanneer toetsresultate dus geïnterpreteer word, moet dit binne konteks van die SEM gedoen word – die SEM maak voorsiening vir fluktuasie wat tot fouttellings lei tussen die waargenome en ware tellings. Smit (1991:27) definieer die metingsfoute as “veranderlikes wat nie relevant tot die doel van die toets is nie, en/of van tyd tot tyd fluktuasies in toetstellings veroorsaak”.

’n Teoretiese uiteensetting van die gebruik van die standaardmetingsfout kom op die volgende neer (Miller, Linn & Gronlund, 2009:120):

- Enige toetsling het ’n waargenome telling en ’n ware telling – laasgenoemde is ’n hipotetiese waarde.
- Indien toetslinge herhaaldelik getoets sou kon word, sonder dat enige sistematiese of onsistematiese fluktuasie ’n effek sou hê, sou die waargenome telling en die ware telling baie na aan mekaar gelê het en die data sou normaal rondom die ware telling versprei gewees het – daar word dus met normaalverspreidings gewerk as basiese vertrekpunt by die interpretasie van resultate.
- Ongeveer 68% van die waargenome tellings behoort binne een standaardafwyking weerskante van die ware telling te val, ongeveer 95% binne twee standaardafwykings, en ongeveer 99.7% binne drie standaardafwykings. Let daarop dat die SEM ’n funksie van die standaardafwyking is, aangesien dit in die formule vir die bepaling daarvan verreken word.
- Alhoewel die ware telling nooit bereken kan word nie, kan die SEM gebruik word om grense rondom die waargenome telling te stel waarbinne die ware telling behoort te lê.

- Hierdie grense is die vertrouensintervalle. Wanneer die SEM klein is, is die skatting akkuraat en die vertrouensinterval dus klein, en wanneer die SEM hoog is, is die skatting minder akkuraat en vertrouensinterval vergroot.

Bogenoemde uiteensetting dui aan waarom toetsresultate eerder in 'n band geïnterpreteer moet word, as om bloot na 'n enkele resultaat te verwys. Daar is dus ook 'n teoreties regverdigbare basis daarvoor om die resultate van TAG en TALL in bande bekend te maak en dit daarvolgens te interpreteer:

Band 1	Uiters hoë risiko
Band 2	Hoë risiko
Band 3	Risiko (word ook gebruik om grensgevalle te identifiseer)
Band 4	Lae risiko
Band 5	Lae tot geen risiko nie

Die afsnyppunte vir die onderskeie bande moes egter gestandaardiseer word en dit is met die inagneming van Davies *et al.* (1999:41) se aanwysers hiervoor gedoen. In die eerste instansie is die subjektiewe oordeel van toets- en vakkenners gebruik vir die bepaling van growwe afsnyppunte vir aanvaarbare of nie-aanvaarbare vlakke van akademiese taalvermoë. Tweedens is daar van die konsep van sogenaamde absolute standarde gebruik gemaak waar die vermoë van 'n kandidaat teen die een of ander eksterne kriterium gemeet word – in die geval van TAG, die konstruk van die toets waarin die spesifieke elemente van akademiese geletterdheid wat gemeet word gespesifiseer word; dit is verreken teen historiese data uit die ELSA Plus waar ongeveer een derde van alle kandidate as kandidate met risiko geag is. Volgens Davies *et al.* (1999:41) is daar drie metodes wat gebruik kan word om afsnyppunte te bepaal wanneer absolute standarde gebruik word:

- metodes wat op item-/taaktipes fokus waar die verskillende toetsitems saam gegroepeer word volgens die taak, of die moeilikheidsgraad, of die belangrikheid/gewig van die spesifieke element wat gemeet word;
- metodes wat die adekwaatheid van toetslinge se vermoëns in ag neem en waar daar 'n oordeel gevel word oor of 'n toetsling aan die kriterium voldoen

het, al dan nie – die mediaan word soms hier in ag geneem om te onderskei tussen aanvaarbare en onaanvaarbare vaardigheidsvlakke; en

- metodes waarin ’n oordeel gevel word oor die persentasie toetslinge wat aanvaarbare vaardigheidsvlakke het.

Dit gaan dus daaroor dat “methods for setting standards all ultimately rely on the subjectivity of human judgement [and] this judgement is typically exercised by experts in relation to test content (...), performance of test takers (...), or a combination of the two (Davies, *et al.* 1999:186).

Dit was egter ingewikkeld om op óf die een óf die ander metode vir absolute standarde te besluit, omdat die konstruk van TAG eerstens besonder ryk is, en ten tweede omdat die onderskeie instellings wat TAG gebruik elk unieke behoeftes en eise het waaraan voldoen moet word. Daar word dus gewoonlik ’n kombinasie van die tweede en die derde metodes, hierbo uiteengesit, gebruik vir die bepaling van TAG se afsnyppunte. In praktyk kom dit daarop neer dat Bande 1 en 2 ongeveer die onderste 16% van ’n spesifieke toets vir ’n spesifieke jaar van inname sal verteenwoordig, terwyl Band 5 die boonste 16% verteenwoordig. Band 3 verteenwoordig dus die onderste 16% minus 50% van die groep. Net so verteenwoordig Band 4 weer die boonste 16% minus 50%.

Vir die bepaling van risikovlakke is daar besluit om die vertrekpunt van normaalverspreidings te gebruik waar ongeveer 68% van die data binne die gemiddelde punt en een standaardafwyking behoort te val. Dit hou in dat die risikovlakke as volg beskryf kan word:

- Band 1: minder as (gemiddeld – 1 standaardafwyking)
- Band 2: minder as gemiddeld, maar meer as (gemiddeld – 1 standaardafwyking)
- Band 4: meer as gemiddeld, maar minder as (gemiddeld + 1 standaardafwyking)
- Band 5: meer as (gemiddeld + 1 standaardafwyking)

Let daarop dat daar hoofsaaklik vir sekuriteitsredes vir elke kohort of inname 'n ander akademiese geletterdheidstoets gebruik word, en dat, soos reeds genoem, elke instelling unieke behoeftes het. Die punteverspreiding het derhalwe van toets tot toets / jaar tot jaar verskil en die afsnypunte vir die risikobande word dus afsonderlik vir elke toets en instelling bereken en nie vir al die innames gesamentlik nie, maar dit wat hierbo uiteengesit is, is die benaderde afsnypunte wat oor 'n tydperk van ses jaar met verskillende toetse en met die inagneming van die verskillende instellings se behoeftes vorendag gekom is.

Neem ook in ag dat daar hierbo nie van Band 3 melding gemaak word nie, omdat die sogenaamde grensgevalle as Band 3-kandidate geag word en daar addisionele toetse met hierdie toetslinge gedoen word ten einde te bepaal of hulle uiteindelik in Band 2 of Band 4 hoort – die metingsfout word onder andere in berekening gebring wanneer daar bepaal word hoe groot of klein Band 3 aanvanklik behoort te wees (sien ook die uiteensetting vroeër in hierdie hoofstuk oor metingsfout en die ware telling en waargenome telling). Wat van belang is, is nie soseer die metode wat gebruik word vir die bepaling van afsnypunte nie, maar dat dit volgens Davies *et al.* (1999:41)

needs to ensure success on a critical number of test items on a specified range of language skills and at an appropriate level of difficulty. The cut-score also needs to be monitored against real-world evidence to ensure that qualified candidates are not unnecessarily excluded and/or unqualified candidates admitted. Whether the cut-score errs on the side of leniency or severity is influenced by a consideration of the harm caused by excluding qualified candidates as opposed to harm caused by passing unqualified candidates.

Ten einde die Band 3-kandidate so ver moontlik ter wille te wees, word die volgende sisteem gebruik wanneer afsnypunte vir hierdie band bepaal word (cf. Weideman & Van der Slik, 2008:168-171): TiaPlus (Cito:2005) word gebruik om die sogenaamde misklassifikasies te bepaal – misklassifikasies verwys na 'n beraming van daardie studente wat, statisties gesproke, moontlik onregverdig behandel kon gewees het (tot hul persoonlike voordeel of nadeel), omdat geen toets ooit 100% betroubaar is nie. 'n Toetsling kan dus verkeerdelik die toets slaag of drui. TiaPlus maak gebruik van Cronbach se alpha en die GLB-koëffisiënt ('greatest lower bound') wat verreken word teen die sogenaamde RXX- en die RXT-

waardes. Eersgenoemde korreleer toetsresultate met 'n hipotetiese parallelle toets, en laasgenoemde korreleer waargenome tellings met ware tellings. Daar is altyd meer misklassifikasies waar Cronbach se alpha gebruik word en minder waar GLB gebruik word, omdat alpha se parameters strenger is. Anders gestel: Cronbach se alpha gee 'n onderskatting van die werklike betroubaarheid, met die gevolg dat die moontlike getal misklassifikasies oorskot word; die GLB bied aan die ander kant 'n meer akkurate skatting van betroubaarheid en gevolglik is die misklassifikasies iets minder. Die gemiddeld tussen die hoogste en laagste waarde word dan telkens gebruik om te bepaal hoeveel misklassifikasies daar kon gewees het. Hoewel misklassifikasies bo én onder die afsnypunt kan voorkom en net die helfte van die gemiddelde waarde hierbo genoem dus gebruik hoef te word, word meer as 50% as waarde gebruik en slegs vir studente wat onder die afsnypunt val – die venster word dus so wyd moontlik oopgemaak. Studente wat onder die afsnypunt, maar binne die gemiddelde waarde, soos hierbo verduidelik, lê, het dan 'n keuse om 'n tweede sogenaamde valskermtoets te kom skryf, waarvan die punt gebruik word om 'n finale oordeel te vel. Die valskermtoets is 'n ekwivalente toets.

So byvoorbeeld is die alpha-waarde vir TAG 2010 as geheel 0.87, en die GLB 0.91. Die alpha-gebaseerde RXX-waarde is 17, en die GLB-gebaseerde RXX-waarde is 15. Die alpha-gebaseerde RXT-waarde is 13, en die GLB-gebaseerde RXT-waarde is 11. Die gemiddeld van 17 en 11 is 14 en as dit as 'n persentasie van die aantal studente wat die toets afgelê het, uitgedruk word, beteken dit 0.5% van die populasie se resultate kan aan misklassifikasie toegeskryf word. Die helfte van hierdie persentasie lê hipoteties onder die afsnypunt en die ander helfte bo die afsnypunt en slegs 0.25% hoef dus tot 'n sogenaamde valskermtoets toegelaat te word. Ten einde die studente egter ter wille te wees, word tussen 0.25% en 0.50% onder die afsnypunt as venster oopgemaak, wat beteken dat die eerste ongeveer 14 studente onder die afsnypunt vir 'n hertoetsgeleentheid kwalifiseer. Van der Slik en Weideman (2005:29) en Weideman en Van der Slik (2008:168-171) kan vir nog spesifieke voorbeelde geraadpleeg word.

Miller, Linn en Gronlund (2009:123) wys egter ook daarop dat die gebruik van normaalverspreidings en die interpretasie van die waargenome telling teenoor die werklike telling problematies kan wees by kandidate met besondere hoë of besondere lae resultate. In laasgenoemde geval kan 'n toetsling wat 'n toets byvoorbeeld besonder moeilik ervaar het se punt eerder aan die sogenaamde raafaktor toegeskryf word, as aan 'n waargenome telling wat naby 'n ware telling behoort te lê – in sulke gevalle sal die SEM altyd hoog wees, sou dit apart geanaliseer word. Dit word dus soms gevind dat die SEM apart gerapporteer word vir die swakste en sterkste groepe, maar dit word nie in die geval van TAG en TALL so gedoen nie, omdat die SEM slegs ter sprake kom by grensgevalkandidate (soos hierbo uiteengesit).

Die metingsfout bied ten slotte vir toetsadministreerders die geleentheid om betroubaarheidskoëffisiënte van toepassing te maak op individue se waargenome toetstellings. Die waarde daarvan lê in die feit dat dit vir toetsadministreerders bevestig dat daar altyd variasie en fluktuasie in toetsresultate sal wees en dat dit op die een of ander manier verreken móét word. 'n Goeie voorbeeld van so 'n verdiskontering is die feit dat TAG en TALL se resultate op 'n verantwoordbare wyse gebruik word om studente in 'n band te plaas, en hulle dan daarvolgens op gepaste steunkursusse te plaas. Binne konteks van die teoretiese raamwerk wat vir hierdie studie gebruik word, naamlik die van Weideman (2006), is dit dus van kardinale belang dat die regulatiewe aspek van billikheid moet artikuleer met die konstitutiewe aspek van betroubaarheid – die een kan nie sonder die ander nie.

5.4 Faktore wat betroubaarheidskoëffisiënte beïnvloed

Uit die aard van die saak, en soos reeds ook in paragraaf 5.2 hierbo genoem, is daar altyd faktore wat die betroubaarheidskoëffisiënt beïnvloed. Hierdie faktore moet in ag geneem word wanneer die betroubaarheidskoëffisiënt van 'n toets geïnterpreteer word. Die waarde hiervan is dat dit eerstens vir toetsadministreerders van nut sal wees by die interpretasie van die betroubaarheidswaarde, en tweedens dat dit

waarde sal toevoeg tot die herontwerp van bestaande toetse of die skep van nuwes – dus weereens die kwessie van verantwoordbaarheid.

Die literatuur (Smit, 1991; Bachman, 1990; Nitko, 2004; Miller, Linn & Gronlund, 2009; en Alderson, Clapham & Wall, 1995) wys die volgende uit as faktore wat betroubaarheidskoëffisiënte beïnvloed:

- Lengte van die toets (aantal vrae)
- Toetsprosedure (gestandaardiseerde toetsadministrasie-praktyke)
- Tydsbeperking (hoe minder druk, hoe hoër die betroubaarheid en andersom)
- Omvang van die verspreiding van resultate en die variansie in die groep (dus met die inagneming van die moeilikheidsgraad)
- Objektiviteit/subjektiviteit by nasienprosedures (dit is veral belangrik waar toetse met die hand nagesien word)
- Metode van bepaling van betroubaarheidskoëffisiënt (toets-hertoets-betroubaarheid, ekwivalente-vorm-betroubaarheid, interne konsistensie, en betroubaarheid tussen beoordelaars)
- Konstruksie van die toets (aspekte in die toets self wat tot fluktuasie en dus 'n verhoging in die fouttellings bydra)
- Die invloed van die raafaktor (veral by die sterkste en swakste groepe)

Vir TAG en TALL word, met die uitsondering van die laaste faktor, al die bogenoemde verreken. Wat wel interessant is, is die feit dat TAG en TALL relatief kort toetse is (ongeveer 62 vrae, jaar op jaar) en altyd met 'n tydsbeperking (ongeveer 55 minute) gepaardgaan. Teoreties gesproke sou dit dus die betroubaarheidswaardes negatief kon beïnvloed. As die sorgsaamheid (loodstoetse, analise van individuele items binne konteks van gestelde parameters vir diskriminasiewaardes en moeilikheidsgraad van items, sifting van items ten opsigte van oor- of onderverteenwoordiging van konstruk, ensovoorts) waarmee hierdie toetse gemaak en deurlopend aangepas word en die teoreties verdedigbare konstruk waarop dit gegrond is egter in ag geneem word, alles juis ter wille van groter betroubaarheid, is dit te verstane dat die betroubaarheidskoëffisiënte so hoog is.

Verder moet dit ook in ag geneem word dat die vermoë om onder tydsdruk te kan funksioneer ook 'n teken van akademiese geletterdheid is – studente word deurlopend daarmee gekonfronteer om groot hoeveelhede studiemateriaal in 'n kort tydjie te prosesseer en aan die hand daarvan weer nuwe produkte te skep en TAG en TALL kan dus ook geag word as 'n goeie aanduiding van hierdie vaardigheid. Daar sou egter ook geargumenteer kon word dat dit 'n teken van intelligensie is, maar hierdie kwessie val nie binne die bestek van hierdie studie nie. Omdat die betroubaarheidswaardes van TAG en TALL deurlopend op dieselfde wyse bereken is, het die metode van bepaling ook nie 'n invloed nie – dit bly van jaar tot jaar en oor taalgrense heen dieselfde.

5.5 Die interpretasie van betroubaarheidskoëffisiënte

Dit word algemeen aanvaar dat geen toets ten volle betroubaar is nie. Smit (1991:42) stel dit ook dat toetsbetroubaarheid 'n graadverskynsel is en nie 'n alles-of-niks-beginsel nie. Hy voer verder aan dat daar nie arbitrêre beginsels vasgelê kan word waarvolgens 'n toets byvoorbeeld as onbevredigend of bruikbaar geklassifiseer kan word nie. Die graad van betroubaarheid wat toetsadministrateurs en vakkundiges verwag, hang tot 'n groot mate af van die besluit wat geneem moet word en is op die populasie as geheel gebaseer en nie op individue nie, omdat daar veralgemeen moet word (Bachman, 1990:187-202). Die kwessie van laebelang- of hoëbelangtoetse is dus weer ter sprake: by eersgenoemde kan 'n lae betroubaarheidswaarde aanvaar word, omdat die impak of gevolge daarvan nie so hoog is nie, maar by laasgenoemde is 'n hoë betroubaarheidskoëffisiënt wel van kardinale belang.

In die geval van TAG en TALL kon lae betroubaarheidswaardes vroeër toe dit slegs as plasingsinstrument gebruik is nog aanvaar word, hoewel dit nooit nodig was nie, omdat die betroubaarheidswaardes van hierdie toets van meet af aan hoog was. Sedert hierdie toetse egter ook as deel van 'n toelatingsmeganisme gebruik word, is die toleransievlak besonder laag.

Ten slotte kan dit gestel word dat wanneer daar 'n besluit oor 'n aanvaarbare graad van betroubaarheid geneem moet word, 'n aantal riglyne in oorweging geneem moet word. Dit word in Tabel 5.12, wat oorgeneem en aangepas is uit Miller, Linn en Gronlund (2009:132), opgesom:

Tabel 5.12: Riglyne vir die oorweging van toleransievlakke by betroubaarheidskoëffisiënte

Hoë betroubaarheidskoëffisiënte word verwag wanneer die besluit	Lae betroubaarheidskoëffisiënte kan getolereer word wanneer die besluit
wat geneem moet word belangrik is.	wat geneem moet word minder belangrik is.
finaal is.	nog in 'n voorlopige fase is.
onomkeerbaar is.	omkeerbaar is.
onbevestigbaar is.	bevestig kan word deur ander data.
individue raak.	groepe raak.
langdurige gevolge het.	slegs tydelike of min gevolge het.

5.6 Samevatting

In Hoofstuk 2 is daar aangevoer dat die oplossings vir taalprobleme meermale die teoretiese regverdiging daarvoor voorafgaan. Dit is tipies wat met TAG en TALL gebeur het en wat ook uit hierdie hoofstuk oor betroubaarheid behoort te blyk: die probleem is geïdentifiseer, tegniese kennis en verbeelding is toe ingespan om met 'n voorlopige oplossing vorendag te kom en eers daarna is 'n teoretiese regverdiging daarvoor gekry wat tot 'n finalisering van die teoretiese basis gelei het. Hoewel die begrip betroubaarheid aan die konstitutiewe kant van die teoretiese raamwerk wat in Hoofstuk 2 bespreek is lê, is dit deurlopend in interaksie met kwessies soos bruikbaarheid, sorgsaamheid en die regverdiging daarvoor, wat aan die regulatiewe kant van die raamwerk lê. Daar behoort dus voortdurend artikulasie tussen die twee kante van die model te wees, maar aspekte wat bydra tot fluktuasie moet deurlopend in ag geneem word. Hierdie punt word goed geïllustreer deur die volgende woorde van Bachman (1990:226):

Fundamental to the development and use of language tests is being able to identify and estimate the effect of various factors on language test scores. In order to interpret test scores as indicators of a given language ability, we must be sure that they are influenced as much as possible by that ability. Any factors other than the ability being tested that affect test scores are potential sources of error that decrease both the reliability of scores and the validity of their interpretations. Therefore it is essential that we be able to identify these sources

of error and estimate the magnitude of their effect on test scores. Our ability to do this depends upon how we define the various influences on test scores.

Naas geldigheid, wat in Hoofstuk 6 en ten dele in Hoofstuk 7 bespreek sal word, is betroubaarheid seker dié belangrikste eienskap waaroor 'n goeie toets behoort te beskik. In hierdie hoofstuk oor betroubaarheid is aangetoon dat betroubaarheid verwys na die konsistensie en stabiliteit van resultate van een meting na 'n volgende en dat dit 'n konsep is wat deur statistiese bewerkings bepaal, en syfermatig uitgedruk kan word. In die konseptuele terme van die teoretiese raamwerk wat in Hoofstuk 2 voorgehou is, druk die tegniese betroubaarheid van die toets wat as toegepaste linguistiese meetinstrument gebruik word, die verhouding uit tussen die tegniese, leidinggewende funksie daarvan en die kinematiese dimensie van ons ervaring. In die kinematiese sfeer kry 'n mens oorspronklik met konsistente of reëlmatige beweging te doen, en die begrip tegniese konsistensie bring dit juis tot uitdrukking.

Implisiet tot hierdie hoofstuk lê die waarneming dat betroubaarheidswaardes binne verskillende kontekste verskillend gemeet en verskillend geïnterpreteer moet word, afhangend van die doel daarmee. Daar is daarom na verskeie metodes en statistiese tegnieke verwys vir hoe betroubaarheid bepaal kan word, sowel as na aspekte wat dit beïnvloed. Sommige van hierdie metodes en tegnieke was bruikbaar vir analises van TAG en TALL en ander weer nie, as gevolg van die wyse waarop TAG en TALL saamgestel, of gekonstrueer, is en die doel waarvoor dit gebruik word. Daar is vir hierdie studie uitsluitlik van die beginsels van klassieke toetsteorie ('classical test theory') gebruik gemaak, hoewel dit beperkinge het. Die keuse vir die gebruik van klassieke toetsteorie is dat TAG en TALL in wese 'n kriteriumgebaseerde toets is (die konstruk is die kriteria waarteen toetslinge gemeet word) en klassieke toetsteorie hom goed leen tot analises van hierdie tipe toetse; 'item response theory' kon eweneens ook gebruik gewees het vir hierdie tipe analises. Bachman (1990:227) stel dit dat "[i]n the context of [criterion-referenced] testing, reliability is concerned with both the dependability of test scores as indicators of an individual's level of mastery in a given domain of abilities and the dependability of

decisions that are based on [criterion referenced] test scores.” Hierdie woorde van Bachman plaas weer die kwessie van artikulasie tussen die konstitutiewe en die regulatiewe op die voorgrond, naamlik dat daar artikulasie en konjunksie tussen die twee kante van die raamwerk behoort te wees, omdat bepaalde toetsresultate bepaalde gevolge het.

Die standaardmetingsfout is verder ook bespreek en die nut daarvan vir die bepaling van (konstante) vertrouensintervalle is uitgelig en op TAG en TALL van toepassing gemaak.

Daar moet egter ook opgelet word dat ’n hoë betroubaarheidswaarde nie noodwendig impliseer dat ’n toets geldig is nie en dat dit afhanklik is van aspekte soos die lengte van die toets, die moeilikheidsgraad, die objektiviteit waarmee die toets nagesien word, die metode/tegniek wat gebruik is, ensovoorts – alles aspekte wat onder andere in G-teorie verreken word. Die hoofstuk het verder ook konteks geskep vir aanvaarbare vlakke van betroubaarheid, weereens met spesifieke verwysing na TAG en TALL.

In die volgende hoofstuk word die kwessie van geldigheid (ook ’n konstitutiewe begrip) onder die loep geneem. Dit bou voort op hierdie hoofstuk oor betroubaarheid waaruit dit duidelik behoort te blyk dat betroubaarheid ’n invloed op geldigheid het en dat betroubaarheid ’n nodige, maar nie voldoende, karaktertrek van ’n geldige toets behoort te wees.

Geldigheid: drie moontlike perspektiewe

6.1 Inleidend

Hoofstuk 1 van hierdie studie voer aan dat dit noodsaaklik is vir universiteite om die voorbereidingsvlakke van veral (voornemende) eerstejaarstudente in terme van akademiese taalvermoëns te meet alvorens hulle vir verdere studie registreer (dus vir toelatingsdoeleindes), of nadat hulle toegelaat is (vir doeleindes van plasing). Yeld (2001:226) kwalifiseer dit deur te argumenteer dat taal eerstens gebruik word om te leer ('n heuristiese funksie dus) en tweedens om feite en kennis oor te dra/te produseer ('n produktiewe funksie). Cliff, Ramaboa en Pearce (2007:34) artikuleer dit nog meer spesifiek deur vier redes hiervoor aan te bied wat aansluit by dit wat breedvoerig in Hoofstuk 1 van hierdie studie uiteengesit is. Hulle som die Suid-Afrikaanse situasie daarna as volg op:

There is now more than anecdotal evidence that Higher Education institutions and admissions committees or panels are taking seriously the need for responsible, ethical and equitable approaches to admissions decisions, and a parallel need to make use of the multiple sources of information collected about applicants for the placement of those eventually registered into appropriate curricula. (Cliff, Ramaboa en Pearce (2007:34)

In hierdie aanhaling van Cliff, Ramaboa en Pearce word begrippe soos 'responsible', 'ethical', 'equitable', 'multiple sources', en 'appropriate' gebruik, wat artikuleer met begrippe wat in die konseptuele raamwerk wat in Hoofstuk 2 as

teoretiese vertrekpunt van hierdie studie voorgehou is, gebruik word. In dieselfde hoofstuk is daar aangevoer dat toegepaste linguïste (die tipiese ontwerpers en ontwikkelaars van taaltoetse) selde tyd maak om terug te tree en te reflekteer, en teoretiese regverdiging te soek vir dit wat hul praktyk rig, aangesien daar altyd die een of ander dringendheid aan die werk van diesulkes is – oplossings vir taalprobleme (tegniese ontwerpe) moet vinnig bedink en eintlik sonder uitsondering dringend geïmplementeer word (Weideman, 2006:71). 'n Afleiding wat hieruit gemaak sou kon word, is dat daar onder andere om bostaande redes nie altyd met die nodige verantwoordelikheid, omsigtigheid, verantwoordbaarheid, ensovoorts met tegniese ontwerpe omgegaan word nie. In lyn met Cliff, Ramaboa en Pearce, en Weideman se argumente hierbo is die doel van hierdie studie, en meer spesifiek nog hierdie en die volgende twee hoofstukke, onder andere om juis terug te staan en te reflekteer oor 'n tegniese ontwerp en ontwikkeling van 'n linguïstiese artefak (die toetse TAG en/of TALL). Dit word aan die hand van die konseptuele raamwerk waarop hierdie studie berus, gedoen. Om verantwoordbaar oor 'n veelheid van veranderlikes diskussie te kan voer, ondersoek hierdie hoofstuk spesifiek konsepte wat verband hou met die krag of effek van toetse soos TAG en TALL. Geldigheid het in terme van die konseptuele raamwerk wat hier aangewend word in die eerste plek te doen met die tegniese krag of effek wat van die meetinstrument vereis word - 'n instrument beskik oor geldigheid as dit die nodige effek tot gevolg kan hê.

6.2 Geldigheid gedefinieer

Soos reeds in die vorige hoofstuk genoem, is geldigheid naas betroubaarheid seker dié belangrikste eienskap van enige toets / toetsresultaat, of soos Van der Walt en Steyn (2007:138) dit stel: “validity is the central concept in language assessment.” Geldigheid blyk egter 'n komplekse begrip te wees (Brown, 2004:22) wat met verloop van tyd verskeie interpretasies ondergaan het, maar dit word veral gekompliseer deur die invloedrykheid van Samuel Messick (1989; 1994) en sy interpretasie daarvan, naamlik dat alles, selfs betroubaarheid, onder die begrip geldigheid ingetrek kan word en dat geldigheid nie 'n kenmerk of 'n attribuut van 'n toets op sy eie kan wees nie, maar dat dit alles omvat en dus insluit. Hierdie

spesifieke kwessie word hieronder en so ook ten dele in Hoofstukke 7 en 8, waar onder andere 'n tegniese-teoretiese regverdiging vir TAG en TALL aangebied sal word, meer breedvoerig behandel. Vir doeleindes van hierdie hoofstuk en die twee wat daarop volg, sowel as die vorige een, is daar egter besluit om die begrippe betroubaarheid en geldigheid apart te behandel ter wille van konseptuele skerpheid (Weideman, 2009a:24) - dit is belangrik om die konsepte op hul eie eers deeglik uit te pak en te begryp alvorens daar vir of teen die huidige ortodoksie geargumenteer kan word. Ten einde die kompleksiteit van die begrip geldigheid meer verstaanbaar te maak, is daar besluit om die definisies van geldigheid sistematies te benader en in drie paradigmas op te deel, te wete die tradisionele paradigma, die hedendaagse paradigma, en 'n alternatief tot die huidige ortodoksie.

Let egter intussen daarop dat 'n toets volgens Alderson *et al.* (1995:187) nie geldig kan wees as dit nie betroubaar is nie, wat beteken dat 'n toets konsistent in metings moet wees, anders kan dit nie as 'n akkurate meting geag word nie. Aan die ander kant kan 'n toets egter betroubaar wees, sonder om geldig te wees. Dit hou in dat alhoewel konsistente toetsresultate opgelewer word, die toets moontlik nie dit meet waarvoor dit aanvanklik ontwerp is nie. Daar kan dus afgelei word dat betroubaarheid 'n invloed op geldigheid het en dat dit 'n nodige, dog nie 'n voldoende, kenmerk of attribuut van 'n toets behoort te wees. Hierdie twee konsepte is so verweef met mekaar dat 'n toets se betroubaarheid nie beoordeel kan word sonder om ook op die geldigheid daarvan te let nie, en omgekeerd. Chapelle (1994; 1999) argumenteer byvoorbeeld dat betroubaarheid een van die bewyse is wat vir die geldigheid van 'n toets aangevoer kan word en sy beveel gevolglik aan dat betroubaarheid deel van geldigheidsargumente behoort uit te maak.

Die verweefdheid van die tegniese konsistensie en geldigheid van 'n toets is inderdaad ook sonder meer duidelik as dit beskou word vanuit die hoek van die tegniese sy van ons ervaring, wat die leidende funksie uitmaak van 'n ontwerpte taaltoets se samehang met die kinematiese en fisies-energetiese ervaringsmodusse. As die verwantskap tussen die tegniese en die kinematiese teoreties blootgelê word,

dan stel dit die toegepaste taalkundige in staat om te werk met die begrip tegniese konsistensie, of betroubaarheid, soos in die vorige hoofstuk uiteengesit. As die tegniese met die fisies-energetiese konnekteer, genereer dit weer die konsep van tegniese krag, of tegniese effek, want in die fisiese dimensie word oorspronklik die gedagte van oorsaak en gevolg, die dinamiese werking van kragte, aangetref. Dat betroubaarheid en geldigheid twee kante van dieselfde saak is, is duidelik; ook dat tegniese betroubaarheid in 'n funderende (konstitutiewe) rigting geldigheid onderlê en voorafgaan. Toetse moet dus betroubaar en geldig wees alvorens dit as regverdig vir 'n spesifieke doel geag kan word en empiriese ondersoek (waarvan Hoofstukke 5 en 7 vir 'n deel 'n voorbeeld is) word gevolglik gereeld onderneem om die betroubaarheid en geldigheid van toetse te ondersoek. Hierdie tipe ondersoek is egter nie noodwendig altyd 'n adekwate of 'n toepaslike regverdiging vir 'n toets nie en belanghebbendes, soos universiteitsowerhede wat oor kwessies soos toelating en plasing moet besin, moet versigtig wees om nie te verval in regverdigingsargumente wat verbloem word deur oënskynlik wetenskaplike nomenklatuur en deugsamheid nie (Weideman, 2006:73).

Hetsy betroubaarheid of geldigheid gemeet word, is die belangrikste oorweging of 'n toetsresultaat regverdig en 'n akkurate refleksie is van 'n kandidaat se vermoëns (Alderson *et al.*, 1995:187-188; Hughes, 2003:50; Miller, Linn, Gronlund, 2009:71). Die res van hierdie hoofstuk bied daarom 'n teoretiese uiteensetting van die begrip geldigheid in aanloop tot Hoofstuk 7 wat 'n empiriese ondersoek na die geldigheid van TAG is. In beide hierdie hoofstukke, sowel as in Hoofstuk 8, sal daar deurlopend gepoog word om verantwoordbaar oor 'n veelheid van veranderlikes wat onder andere tot die geldigheid van 'n toets bydra, diskussie te voer.

6.2.1 Die eerste paradigma: 'n tradisionele benadering tot geldigheid

Die begrip geldigheid word reeds dekades gelede, binne konteks van die psigometrika, deur Magnussen (1966:123) gedefinieer as "... the accuracy with which meaningful and relevant measurements can be made ... in the sense that it

actually measures the traits it was intended to measure.” In dieselfde jaar verklaar French en Michael (1966:165) ook dat verskillende toetse vir verskillende doeleindes ontwikkel is en dat die geldigheid van ’n toets afhanklik is van die mate waartoe ’n spesifieke toets gemeet het wat dit ten doel gehad het om te meet: “[t]ests are used for several types of judgements and for each type of judgement, a different type of investigation is required to establish validity.” Beide bostaande definisies bou voort op die werk van Kelley (1927), Cattell (1946) en Cronbach en Meehl (1955), maar daar word doelbewus gekies om eerder die definisies van Magnussen en French en Michael hierbo aan te bied, omdat dit eerstens ’n goeie opsomming van bostaande werk is, en tweedens omdat dit reeds ’n suggestie is van wat in die volgende paradigma sou volg. Let intussen daarop dat die klem in hierdie definisies implisiet, dog sterk, val op die empiriese – verskillende ondersoek is nodig om bewyse te kan verskaf vir geldigheid. Daar is egter ook, soos reeds genoem, suggesties van die feit dat relevante interpretasies van toetspunte gemaak moet word ten einde oordele oor toetslinge te kan vel.

Lado (1961b:30) wys in aansluiting by die bostaande daarop dat geldigheid nie algemeen nie, maar wel spesifiek is. Dit vind steeds weerklank in hedendaagse beskouings (die tweede én derde paradigmas wat hieronder bespreek word) van geldigheid, naamlik dat “validity is contextual, local and specific” (Van der Walt & Steyn, 2007:138).

Die tradisionele beskouing van geldigheid kom dus daarop neer dat dit nie net konsistent moet wees in die resultate wat dit oplewer nie (betroubaar dus), maar dat dit veral moet meet wat dit veronderstel is om te meet. ’n Toets is daarom net geldig vir ’n spesifieke doel in ’n spesifieke konteks, met ander woorde ’n toets kan opsigself nie oor geldigheid beskik nie, maar dit kan slegs geldigheid vir ’n spesifieke intensie hê. Geldigheid word derhalwe geag as ’n inherente deel van ’n toets, ’n eienskap waaroor ’n toets sal beskik, mits dit meet wat dit veronderstel is om te meet (Kelley, 1927; Cattell, 1946; Lado, 1961a). Sluifster (1998:69) argumenteer egter dat die beantwoording van die vraag of ’n toets meet wat dit

veronderstel is om te meet, nooit die einddoel van 'n geldigheidsondersoek kan wees nie. Hy stel dit dat

[e]en positief antwoord op deze vraag is een noodzakelijke, maar geen voldoende, voorwaarde voor een test om een rol te kunnen spelen bij het nemen van een bepaalde beslissing. Indien een valideringsonderzoek zich zou beperken tot de vraag of een test meet wat deze beoogt te meten, is het niet mogelijk een uitspraak te doen over de geschiktheid van deze test voor het nemen van een bepaalde beslissing.

Interessant genoeg word daar in die tradisionele benadering tot geldigheid, soos reeds hierbo gesien kan word, gesinspeel op wat logieserwys uit hierdie paradigma na die volgende, wat 'n meer hedendaagse benadering tot geldigheid is, sou voortvloei. Dit is naamlik dat geldigheid daardie eienskap van 'n toets is wat toetsresultate funksioneel of bruikbaar maak – 'n toets op sigself is nie geldig nie; dit is slegs geldig vir spesifieke doeleindes (Sundberg, 1977:43; Lado, 1961b:30). Dit word bevestig in die volgende woorde van Van der Walt en Steyn (2007:138): “[validity pertains] to a specific use of a test, i.e. one asks whether the test is valid for *this* situation.”

Binne konteks van die teoretiese raamwerk wat vir hierdie studie gebruik word, en in Hoofstuk 2 uiteengesit is, is dit hier reeds duidelik dat konstitutiewe begrippe soos betroubaarheid en geldigheid inderdaad die grondbegrippe is waarop tegniese ontwerpe berus, maar daar is ook reeds aanduidings daarvan dat hierdie grondbegrippe slegs ontsluit kan word deur die leidende funksies van die ontwerp - in hierdie geval spesifiek die juridiese en etiese dimensies van ons bestaan, wat opvattinge oor regverdigheid en billikheid antisipeer. In die volgende gedeelte word daar na die hedendaagse beskouing tot geldigheid gekyk wat as't ware 'n teenreaksie op die oordrewe klem van die bewysbare was. Daar sal duidelik gesien kan word hoe die verskillende dimensies van linguistiese artefakte in wisselwerking met mekaar behoort te tree, maar dat daar in die tweede paradigma 'n oordrewe klemverskuiwing na die regulatiewe kant van die teoretiese raamwerk plaasgevind het, wat aanleiding gegee het tot die ontstaan van die derde paradigma wat aan die hand doen dat die konstitutiewe en die regulatiewe momente in tegniese ontwerpe in harmonie met mekaar moet wees.

6.2.2 Die tweede paradigma: die huidige ortodoksie

Die tradisionele benadering tot geldigheid is geag as beperkend van aard, omdat dit vanuit sy perspektief volgens teoretici nie genuanseerd genoeg nie, en selfs simplisties, na geldigheid gekyk het. So is daar byvoorbeeld n t drie sogenaamde tipes geldigheid vir hierdie paradigma geidentifiseer, naamlik kriterium-gerigte geldigheid (wat beide konkurrente en voorspellingsgeldigheid insluit), inhoudsgeldigheid, en konstrugeldigheid – die klem is duidelik op dit wat gemeet en bewys kan word. Geldigheid, so word daar vandag oor die algemeen geredeneer, is egter baie meer kompleks as die definisies hierbo en daarom word ’n fokus op slegs die drie genoemdes tipes geldigheid nie meer as voldoende geag nie. Die hedendaagse benadering tot geldigheid tref dus nie meer ’n onderskeid tussen slegs die drie tradisionele tipes geldigheid nie, maar verenig dit in ’n raamwerk waar konstrugeldigheid as die kern geag word en die ander twee as aspekte daarvan. Dit kry reeds in 1955 beslag in die werk van Cronbach en Meehl (1955:283) waar hulle konstrugeldigheid as volg definieer:

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.

Hoewel die eerste paradigma se definisie, naamlik dat ’n toets moet meet wat dit veronderstel is om te meet, en dat dit bewys moet kan word, steeds as die basiese vetrekpunt vir geldigheid geag word (cf. Alderson *et al.*, 1995:170; Davies, 1990:21; Hughes, 2003:50; Jordan, 1997:88; Weir, 1993:19), en die tweede paradigma in lyn daarmee is, word dit deesdae tog meer genuanseerd en selfs versigtiger omskryf as wat Cronbach en Meel hierbo gedoen het, en daar vind ’n klemverskuiwing na die interpretatiewe kant van toetsresultate plaas, waarop in Hoofstuk 8 teruggekom sal word.

So wys McNamara en Roever (2006:11) byvoorbeeld daarop dat geldigheid in essensie te make het met die maak van afleidings oor die vermoens van toetslinge, vir ’n spesifieke domein, en dat hierdie afleidings gemaak word op grond van toetsresultate wat ge nterpreteer moet word. Dieselfde outeurs voer aan dat hoewel

Cronbach aangedui het dat die nodige bewyse vir die geldigheid van 'n toets ingesamel moet word ten einde toetsresultate geregverdig te kan interpreteer, hy nie in sy aanvanklike werk die kwessie van sosiale verantwoordbaarheid in ag geneem het nie en eers veel later op sake soos konsekwensiële geldigheid gefokus het (McNamara & Roever, 2006:10).

Etlke jare na Cronbach se interpretasie van die begrip geldigheid wys Bachman (1990:289) steeds daarop dat geldigheid, naas die feit dat 'n toets moet meet wat dit veronderstel is om te meet, alleen bepaal kan word as dit 'n proses is wat aan deurlopende analise en empiriese ondersoek van toetsresultate onderwerp word, wat 'n aanduiding daarvan is dat die interpretasie van toetsresultate belangriker begin raak in die gedagtes van taaltoetsers. Miller, Linn en Gronlund (2009:70) voer dieselfde argument as Bachman hierbo, maar selfs in meer besonderhede, naamlik dat geldigheid slegs bereik kan word as toetsresultate toepaslik, betekenisvol en bruikbaar geïnterpreteer kan word, en as die gevolge van besluitneming gebaseer op hierdie interpretasies billik en regverdig sal wees. Dit is sonder meer duidelik dat beide hierdie bronne voortbou op die perspektief van Samuel Messick, wat hieronder in meer besonderhede bespreek sal word.

Bachman en Palmer (1996:9) som die bostaande egter goed op as hulle daarop wys dat daar twee basiese beginsels is waaraan enige taaltoets moet voldoen. Die eerste hiervan dikteer die nodigheid vir 'n noue verband tussen 'n toetsling se prestasie op 'n toets en sy / haar toekomstige taalgebruik in 'n spesifieke situasie (soos in die eerste, psigometriese paradigma). Dit gaan dus oor “a correspondence between language test performance and language use [in the target language use situation]” (Bachman & Palmer, 1996:9). Ten einde hierdie ‘correspondence’ te bewerkstellig, behoort daar 'n agtergrond of 'n konteks te wees waarteen taaltoetsresultate in terme van werklike taalgebruiksituasies geposisioneer en beoordeel kan word. Dit is duidelik dat hierdie soort siening, waar konteks en die effek van toetsresultate op toetslinge aandag begin geniet het, tot die tweede paradigma behoort, hoewel daar reeds blyke daarvan in die eerste paradigma was. Die punt is uiteindelik dat

empiriese bewyse alleen nie meer voldoende geag word vir uitsprake oor geldigheid nie. Toetsontwikkelaars behoort daarom altyd die taalgebruiksituasie in ag te neem wanneer hulle toetse ontwerp en ontwikkel, en terselfdertyd die spesifieke doelgroep waarvoor die toets ontwikkel word.

McNamara (2000:8) wys egter daarop dat daar ook 'n onderskeid tussen die kriterium (korrekte taalgebruik in die taalgebruiksituasie) en die toets self getref moet word, omdat toetse veral te make het met die maak van afleidings uit hul resultate. Toetsresultate, of prestasie op 'n toets, kan slegs gebruik word om afleidings oor prestasie op die kriterium (die taalgebruiksituasie), en niks anders nie, te maak. Let daarop dat daar verskeie wyses is waarvolgens toetsresultate geïnterpreteer kan word, maar die keuse vir 'n bepaalde metodiek hang af van die doel waarvoor die toets ontwerp is (Bachman, 1990:226). Uit die voorafgaande is dit duidelik dat die tradisionele benadering en die huidige ortodoksie hier noue aansluiting by mekaar vind, maar dat in laasgenoemde die element van interpretasie al hoe meer prominent raak.

Die tweede basiese beginsel waaraan taaltoetse moet voldoen (Bachman en Palmer, 1996:9), het te make met die bruikbaarheid van 'n toets, en bruikbare toetse is by implikasie met 'n spesifieke doel voor oë ontwerp vir 'n spesifieke doelgroep en moet dus interpreteerbaar wees. Hierdie beskouing van geldigheid kom daarop neer dat die konsep altyd te make het met die benutting van toetsresultate vir die een of ander doel (dit artikuleer dus met die tradisionele beskouing), maar dat dit aan die voorwaardes van bewysbaarheid en, soos wat hieronder gesien sal kan word, 'n geïntegreerde veelheid van bewyse, en billikheid (wat die juridiese dimensie van ons bestaan eggo) moet voldoen, wat weereens in lyn is met die teoretiese vertrekpunt van hierdie studie, naamlik dat konstitutiewe en regulatiewe momente van tegniese instrumente in wisselwerking met mekaar is, en moet wees.

'n Meer hedendaagse benadering tot geldigheid kom dus daarop neer dat dit, in aansluiting by die eerste paradigma, daaroor gaan of 'n toets meet wat dit ten doel het om te meet, sowel as die adekwaatheid en korrektheid / toepaslikheid waarmee

toetsresultate geïnterpreteer word. Geldigheid gaan verder in die kontemporêre siening dikwels oor graadverskille – dit is nie ’n alles-of-niks-situasie nie, en ’n toets, of die interpretasie van die toetsresultaat, kan dus binne die tweede paradigma nie bloot as geldig of ongeldig beskryf word nie. Die rede vir laasgenoemde is dat die geldigheid van ’n instrument of die interpretasie van die resultate daarvan slegs van toepassing kan wees op ’n spesifieke doel. Messick (1989:13) stel dit dat

[v]alidity is an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores ...

In die ‘integrated evaluative judgement’ vind die moment van interpretasie weereens neerslag. In die tweede paradigma is daar, soos reeds genoem, derhalwe wegbeweeg van die redenasie dat daar drie tipes (bewysbare) geldigheid is en dit is vervang met ’n verenigde benadering tot geldigheid waarin konstruktgeldigheid sentraal staan, en kriterium-gerigte geldigheid en inhoudsgeldigheid as deel van konstruktgeldigheid geag word, omdat dit deel is van die verdedigbaarheid (’n konstitutiewe begrip) van die konstruk (Messick, 1989:20; 1996:248). In aansluiting hierby streek konstruktgeldigheid volgens Bachman (1990:290) veel verder as kriterium-gerigte geldigheid of inhoudsgeldigheid, aangesien dit “empirically verifies (or falsifies) hypotheses derived from a theory of factors that affect performance on tests – constructs, or abilities, and characteristics of the test method.” Messick (1980:1015) het dieselfde punt egter al ’n dekade vroeër geopper toe hy verklaar het dat kriterium-gerigte geldigheid en inhoudsgeldigheid nie geskik geag kan word “[to] bear the name validity and to wear the mantle of all that name implies.” Sy regverdiging hiervoor in sy seminale werk oor geldigheid (Messick, 1989:17) is dat inhoudsgeldigheid bewyse oor die domein en die verteenwoordigende aard van die toetsinhoud van die spesifieke domein verskaf, terwyl dit behoort te gaan oor ’n regverdiging vir die interpretasie van toetsresultate, wat ’n regulatiewe moment is. Dit is interessant om alreeds hier op die paradoks in sy redenasie te let, naamlik dat hy die tradisionele beskouings van kriterium-gerigte geldigheid en inhoudsgeldigheid afmaak as onvoldoende en ontoepaslik, maar dat hy nie ’n probleem het met die feit dat dit wel bewyse lewer wat bydraend is tot die regverdiging van die konstruk nie.

Benewens die kwessie van teoretiese verdedigbaarheid gaan hy in dieselfde hoofstuk selfs nog verder en voeg twee komponente by tot die verenigde benadering tot geldigheid, naamlik die sosiale aspek van toetsing, oftewel die impak of effek daarvan op toetslinge, en die bruikbaarheid daarvan (Messick, 1989:13). In die verband voer hy aan dat geldigheid 'n samevoeging behoort te wees van bestaande bewyse vir, en die moontlike gevolge van, die interpretasie van toetsresultate en die bruikbaarheid daarvan, terwyl hy 'n dekade vroeër aangedui het dat 'n fokus op verskillende tipes geldigheid bydra tot “confusion and, in the face of confusion, oversimplification” (Messick, 1980:1014). Die waarskynlikste rede hiervoor is dat Messick 'n ongemak getoon het met die feit dat 'n enkele ondersoek of 'n stel bewyse van 'n sekere aspek van geldigheid daartoe aanleiding kon gee dat toetsontwikkelaars / navorsers / owerhede 'n toets as geldig ag, gebaseer op 'n beperkte ondersoek en bewyse. Die teendeel is egter ook waar, naamlik dat 'n verenigde benadering eerder tot vaagheid en 'n oorvereenvoudiging kon bydra as wat die geval sou wees met 'n meer gefokusde benadering. Hierdie kwessie word in meer besonderhede hieronder bespreek.

Messick (1989:10) se benadering tot geldigheid word in Tabel 6.1 aangehaal, en in die paragraaf daarna geïllustreer, waar daar gesien kan word dat verskeie aspekte van toetsing saamgevoeg, en selfs vervaag word onder 'n enkele, oftewel 'n verenigde, begrip naamlik geldigheid.

Tabel 6.1: Messick se benadering tot geldigheid

	Test interpretation	Test use
Evidential basis [empirical]	Construct validity (CV)	CV + Relevance / Utility (R/U)
Consequential basis [impact]	CV + Value implications (VI)	CV + R/U + VI + Social consequences

Messick (1989) se benadering tot geldigheid kom dus eerstens daarop neer dat geldigheid 'n verenigende begrip is wat verskillende aspekte byeenbring, en dat konstruktorgeldigheid ten grondslag van alles lê. Die tradisioneel-onderskeibare konsepte van geldigheid, wat hierbo bespreek is, vervaag dus en word van gelyke waarde geag – die een is nie van 'n hoër prioriteit as 'n ander nie, behalwe waar dit

by konstrugeldigheid kom. Dit is interessant om daarop te let dat Messick in die tweede plek, soos reeds genoem, tog terugbeweeg na dit wat in die tradisionele benadering sy beslag gekry het, naamlik dat geldigheid gaan oor die argument wat gevoer word (gegrond op die bewyse), maar dan verklaar hy dat die verskillende bewyse eintlik behoort saam te vloei in 'n verenigde argument wat handel daaroor dat die toets sy doel gedien het / meet wat dit veronderstel is om te meet – dit gaan dus nie meer oor verskillende tipes geldigheid nie, maar oor verskillende tipes bewyse. Messick argumenteer ten derde dat geldigheid nie 'n eienskap van 'n toets kan wees nie, maar slegs van die afleidings wat oor die toetsresultaat gemaak kan word. Die rede hiervoor is dat die aanvanklike doel van toetse soms vervaag en derhalwe in situasies gebruik, of liever misbruik, word en dat die resultate dan geïnterpreteer word asof dit vir daardie spesifieke doel ontwerp is. Ten slotte dra Messick se argument by tot die ondersoek van die gevolge van toetsresultate op individue, gemeenskappe, ensovoorts. In Messick (1989:10) se eie woorde is konstrugeldigheid die “integrating force that unifies validity issues into a unitary concept... [that] ...binds the validity of test use to the validity of test interpretation... [and that] ...binds social consequences of testing to the evidential basis of test interpretation and use.”

Borsboom, Mellenbergh en Van Heerden (2004:1061) stel dit egter dat “most of the validity literature either fails to articulate the validity problem clearly or misses the point entirely”, wat weerklank vind in Van der Walt en Steyn (2007:139) se opmerking dat hoewel Messick se werk oor geldigheid seminaal van aard was, en steeds is, dit ook betreklik vaag is, en selfs moeilik interpreteerbaar. Bostaande artikuleer met McNamara en Roever (2006:33) se opinie dat toetsontwikkelaars nie noodwendig kundiges op die gebied van toetsvalidering is nie, wat aansluiting vind by dit wat in Hoofstuk 2 aangevoer is, maar dat daar tog 'n dringendheid is vir diesulkes om hulle, onder andere, daarmee te bemoei. Hulle voer daarom aan dat veral die kwessie van validering meer hanteerbaar en verstaanbaar vir onervare (en selfs ervare) toetsontwikkelaars gemaak moet word. Dieselfde outeurs se aanbieding van Messick se raamwerk hierbo is daarom miskien 'n beter vertolking

daarvan en dit word in Tabel 6.2 hieronder aangehaal (McNamara en Roever, 2006:14):

Tabel 6.2: McNamara en Roever se herinterpretasie van Messick se benadering tot geldigheid

	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?

Die poging van McNamara en Roever hierbo het primêr ten doel gehad om op 'n eenvoudiger manier aan te toon hoe Messick se raamwerk op grond van die nodige bewyse ook kennis neem van die sosiale dimensie van toetsing en in sekere kontekste geregverdig is om te gebruik (McNamara & Roever, 2006:13). Met hierdie herinterpretasie het hulle egter ook 'n swak plek in Messick se matriks blootgelê, naamlik dat die sosiale dimensie van toetsing slegs in die onderste deel van die matriks na vore tree en die regverdigingsdimensie ('fairness') in die boonste deel (McNamara & Roever, 2006:13). Dit het onmiddellik 'n vraag oor die werklike geïntegreerdheid van die Messick-raamwerk na vore laat tree.

Messick spreek homself byvoorbeeld teen 'n sogenaamde kompartementalisering van geldigheid uit, omdat dit volgens hom tot 'n oorvereenvoudiging van die kwessie bydra. Terselfdertyd bied hy egter onintensieel en op geïmpliseerde wyse 'n gefragmenteerde raamwerk daarvoor aan waar waardes en die gevolge wat uit besluite oor toetsresultate voortvloei as afsonderlik van wetenskaplike bewyse en besluitneming geag word. Wat vreemd opval, is dat, om sy eie woorde te gebruik: "validity as a unified concept is not to imply that validity cannot be differentiated into facets... [and that] ...the distinctions introduced may seem fuzzy because the facets of validity are not only intertwined but overlapping" (Messick, 1989:9). Dieselfde outeur (1995:744) se verklaring hiervoor is dat die onderskeibare fasette van 'n toets juis belangrik is, omdat dit die regverdigbare en

funksionele aspekte van geldigheid aanspreek wat die kompleksiteite inherent aan enige toets blootlê. Hy voer in die verband aan dat “by distinguishing facets reflecting the justification and function of testing, it becomes clear that the distinct features of construct validity need to be emphasized, in addition to the general mosaic of evidence, as one moves from the focal issue of one cell to that of the others” (Messick, 1995:748).

Hierdie vaaghede van Messick, in die vorm van kategoriese stellings, het die potensiaal om groot verwarring te veroorsaak, omdat geldigheid dus nou nie meer net ’n verenigde konsep nie, maar ewe skielik ook uit verskillende fasette bestaan, wat bydra tot ’n komplekse netwerk van argumente, waarin hy en ander teoretici verval. Dit word bevestig deur Borsboom *et al.* (2004:1062) wat verklaar dat

... in the past century, the question of validity has evolved from the question of whether one measures what one intends to measure ... to the question of whether interpretations and actions based on test scores are justified – not only in the light of scientific evidence but with respect to social and ethical consequences of test use... Thus, validity theory has gradually come to treat every important test-related issue as relevant to the validity concept and aims to integrate all these issues under a single header. In doing so, however, the theory fails to serve either the theoretically oriented psychologist or the practically inclined tester: The theoretically oriented are likely to get lost in the intricate subtleties of validity theory, whereas the practically oriented are unlikely to derive a workable conceptual scheme with practical implications from it. A theory of validity that leaves one with the feeling that every single concern about psychological testing is relevant, important, and should be addressed in psychological testing cannot offer a sense of direction to the working researcher. ... Validity is not complex, faceted, or dependent on nomological networks and social consequences of testing. It is a very basic concept and was correctly formulated, for instance, by Kelley (1927, p. 14) when he stated that a test is valid if it measures what it purports to measure.

Hoewel die saak moontlik nie heeltemal so eenvoudig is as wat Borsboom *et al.* dit aanbied nie, soos hieronder sal blyk, is dit tog ironies dat hierdie kwessie nog nie meermale deur teoretici ondersoek is nie en dat daar bloot net indirekte en geïmpliseerde uitsprake daarvoor gemaak word. ’n Voorbeeld hiervan is Bachman en Palmer (1996:17-36) wat, sonder om direkte kritiek op Messick uit te spreek, tog subtiele klemverskuiwings van ’n vereniging van onderskeibare konsepte onder die begrip geldigheid na ’n vereniging van konsepte onder die begrip nuttigheid teweegbring. Hierdie konsepte sluit betroubaarheid, konstrukgeldigheid,

outentisiteit, interaktiwiteit, impak en bruikbaarheid in. Wat Bachman en Palmer (1996) se klemverskuiwing wel prysenswaardig maak, is dat dit die genoemde ses konsepte as gelykwaardig ag en nie een as 'n superordinaat daarstel soos die geval by Messick nie. Die probleem bly egter steeds 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.

'n Moontlike rede vir die gebrek aan deeglike ondersoek is dat eers Messick en later ander noemenswaardige teoretici se invloedrykheid daartoe bygedra het dat hierdie benadering tot geldigheid, waar geldigheid as 'n kenmerk of attribuut van 'n toets verskuif het na die interpretasie van toetsresultate, hedendaags steeds as die norm aanvaar word. Van der Walt en Steyn (2007:139) som dit as volg op: “[v]alidity is [nowadays still] an integrated evaluative judgement based on theoretical and empirical evidence, which supports test score interpretation and use. It is seen as a unitary but multifaceted concept.” Artikulasie met die eerste paradigma is dus steeds sigbaar, maar die fokus verskuif na die billike interpreteerbaarheid van toetsresultate en dit kan slegs bewerkstellig word deur die proses van validasie. Volgens Davies en Elder (2005:799) gaan die konseptuele skerpheid waarop Messick in sy benadering aanspraak maak juis verlore, omdat hy die oplossing vir problematiese kwessies rondom geldigheid heel gerieflik op die proses van validering afskuif, met die gevolg dat geldigheid steeds 'n abstrakte en ingewikkelde konsep is.

Dit is juis hierdie vereniging van toetskenmerke of –attribute onder 'n enkele sambreelterm, geldigheid, wat kritici genoodsaak het om daarop te reageer en wat aanleiding tot 'n alternatiewe paradigma gegee het, wat hieronder, in aansluiting by die bostaande voorlopige artikulasie van die probleem, bespreek sal word. Vir nou is dit egter voldoende om hierdie afdeling af te sluit deur wel erkenning te gee aan Messick dat hy die kwessie van impak (‘consequential validity’) op die voorgrond geplaas het. Die belang hiervan is dat daar verskeie faktore is wat tot sogenaamde

konstruk-irrelevante variansie kan bydra en dus 'n impak op toetsresultate kan hê. Konstruk-irrelevante variansie dra by tot die ongeldige interpretasie van toetsresultate en verwys na daardie aspekte in 'n toets wat veroorsaak dat die interpretasie van die resultaat ongeldig is. Dit gaan dus daaroor dat die toets te wyd meet / aspekte meet wat nie deel van die domein vorm wat gemeet moet word nie.

Interessant genoeg word daar in 'n onlangse werk soos die van Chapelle, Enright en Jamieson (2009:319) wegbeweeg van die benadering van Messick wat alles onder 'n enkele konsep wil verenig, maar dan kan hierdie outeurs hulle tog nie verhelp om weereens te verval in 'n soortgelyke argument as Messick hierbo nie. Hulle redeneer dat

the backing for warrants in an interpretive argument for a test would become comprehensible by constructing a narrative that points to a plausible conclusion ... however, the data that help to construct [this] narrative are generated from a variety of diverse scenarios, and therefore they need to be woven into a single account of [an] interpretive argument.

'n Vereniging van konsepte vind in hierdie voorstel van hulle dus weer uiting, maar dié keer binne die konteks van die interpretatiewe. Hulle voer as rede hiervoor aan dat die versameling van bewyse vir 'n geldigheidsargument problematies is, omdat dit eerstens ingewikkeld is om te bepaal watter tipe bewyse om te versamel, en tweedens omdat daar nie noodwendig altyd 'n aanduiding is van die hoeveelheid bewyse wat benodig word nie, en daarom is dit belangriker om eers resultate te interpreteer en dan vir die nodige bewyse te gaan soek. Chapelle *et al.* (2009:321) regverdig hierdie siening met die volgende woorde:

An interpretive argument consisting of different types of inferences provides guidance as to the types of research needed. Research produces the backing required to support assumptions associated with each inference. The backing is expressed through statements that summarize findings that support inferences, and the design validity argument is composed of these statements within an overall argument leading to the intended conclusion.

Die teoretiese raamwerk van hierdie studie voer aan dat tegniese-linguistiese artefakte ontwerp behoort te word met 'n antisipatoriese en 'n terugskouende ingesteldheid en daar is dus nie fout te vinde met hul soektog na bewyse of teoretiese begroning nadat 'n toets afgelê is nie, maar wat hulle uit die oog verloor,

is dat die beste moontlike interpretasie nie die probleem van konstruk-irrelevante variansie oplos of selfs aanspreek nie. Die resultate sal verdag bly, omdat die instrument (nie die interpretasie nie) ongeldige elemente (konstruk-irrelevante elemente) kan bevat. Waar Messick se teorie nie net alles onder geldigheid wou verenig nie, maar dit ook afhanklik van die interpretasie van die toetsresultate gemaak het, kom Chapelle *et al.* (2009) en probeer dit vereenvoudig deur alles weer onder interpretasie ('n subjektiewe beoordeling) te verenig. In die proses om dit te probeer bewerkstellig, word daar waarskynlik op net so 'n gevaarlike terrein beweeg as wat die geval met Messick was. Interpreteerbaarheid kan nie probleme uit 'n objektiewe instrument verwyder nie en dit is juis hier waar teoretici wat tot hierdie paradigma behoort se argument skuiwergate begin toon, omdat konseptuele skerpheid verlore gaan.

6.2.3 Die derde paradigma: 'n alternatief tot die huidige ortodoksie

Soos reeds in die vorige afdeling genoem, het die vereniging van dit wat tradisioneel as kenmerke / attribute van toetse gesien is, byvoorbeeld betroubaarheid, geldigheid, outentisiteit, interaktiwiteit, impak en die praktiese sy daarvan (implementeerbaarheid / bruikbaarheid), tot die derde paradigma wat 'n alternatief tot die huidige ortodoksie bied, aanleiding gegee. Die kern hiervan lê in die woorde van Van der Walt en Steyn (2007:140), naamlik dat

... the psychometric tradition of language testing has obscured the role and effect of language testing in society, especially its sorting and gate-keeping roles, which ultimately depend on the policies and values that underlie any test. The practice of decision-based interpretations has now become part of validity, although as yet there is no coherent theory of the social context in current validity theory.

Hieruit blyk dit dat die eerste paradigma se oordrewe fokus op akkuraatheid en die bewysbare aanleiding gegee het tot Messick se teenreaksie en 'n vereniging van kenmerke of attribute onder 'n enkele term ten einde 'n meer interpretatiewe perspektief te bewerkstellig met 'n sterker fokus op die impak wat toetse op mense kan hê. Messick se ongeartikuleerde kritiek op die modernisme, naamlik dat syfers as't ware vanself spreek en altyd die basis van die waarheid vorm, het bes moontlik aanleiding tot sy verenigde model gegee. Sy ongemak was daarom geregverdig in

die opsig dat 'n toetsresultaat opsigself nie 'n voldoende of 'n toepaslike aanduiding kan wees van 'n toets se geldigheid nie. Dit is presies die rede waarom hy die nosies van adekwaatheid (voldoende / goed genoeg) en toepaslikheid (die bruikbaarheid van 'n toetsresultaat vir 'n gegewe doel op 'n gegewe moment) in sy verenigde model ter sprake gebring het (Messick, 1988:33-34). Hy definieer geldigheid daarom as “an overall evaluative judgement, founded on empirical evidence and theoretical rationales, of the adequacy and appropriateness of inferences and actions based on test scores” (Messick, 1988:33). Uit hierdie definisie is dit duidelik dat hy nie die waarde van die bewysbare onderskat nie, maar dat hy van oordeel is dat dit nie al is wat tot geldigheid bydra nie; daar is ook ander elemente wat tot geldigheid bydra, maar wat impak het op daardie persone waarvoor oordele gevel moet word na aanleiding van hul toetsresultate.

'n Toets, of 'n toetsresultaat, moet volgens hom dus adekwaat en toepaslik wees en as dit nie is nie, sal dit nie teoreties (gebaseer op empiriese data) of sosiaal regverdigbaar wees wanneer oordele gevel word nie. Hierdie siening het egter weer aanleiding daartoe gegee dat die pendulum oorgehel het na die anderkant van die spektrum, wat sekere kritici se ongemak met Messick se benadering tot gevolg gehad het. Binne konteks van die teoretiese raamwerk wat vir hierdie studie gevolg word en wat volledig in Hoofstukke 2 en 3, en ten dele in Hoofstuk 4 uiteengesit is, kan dit weereens beredeneer word dat die grondliggende funksies waarop toegepaste linguisties-tegniese ontwerpe berus, ontsluit word deur regulatiewe aspekte en dat die twee kante van die raamwerk ideaal-gesproke 'n balans moet wees tussen hierdie kwaliteite / attribute, eerder as wat daar te sterk op die een of ander gefokus word. Soms sal daar wel toegewings gemaak moet word na die een of die ander kant, afhangend van die konteks, maar dit is myns insiens krities dat al (of minstens soveel moontlik van) die genoemde kwaliteite / attribute sentraal staan tot kwaliteitskontrole deur die loop van die toetsontwikkelingsproses. Nie een daarvan kan geïgnoreer / weggelaat / vervaag word nie en juis daarom is Messick se benadering problematies. Brown (2004:22) stel dit byvoorbeeld dat “[t]here is no final, absolute measure of validity, but several different kinds of evidence may be

invoked in support.” Die konsep ‘geldigheid’ sou dus, binne konteks van die teoretiese raamwerk van hierdie studie, eintlik beperk word deur dit slegs vanuit die Messick-perspektief te benader.

Hoewel die Messick- en Bachman en Palmer-benaderings hierbo oënskynlik ruimte laat vir ’n verskeidenheid van perspektiewe en bewyse om geldigheid te bepaal (wat opsigself alreeds problematies is), gaan geldigheid volgens Van der Walt en Steyn (2007:140) in die eerste plek altyd oor die waarheid. Die vraag ontstaan egter of die waarheid absoluut of relatief is wanneer dieselfde outeurs sê dat “[t]ruth remains a relative concept, a question of judgement, a matter of degree, subject to new or more relevant evidence. There is no such thing as an absolute answer to the validity question...” (Van der Walt en Steyn, 2007:140). Die kwessie van waarheid word dus vervaag in bogenoemde benaderings deur ’n aantal veranderlikes te verenig onder ’n oorkoepelende begrip, hetsy dit geldigheid of nut genoem word. Sluitjer (1998:71-72) verwoord die problematiek hiermee as volg:

Het aantal eisen ten aanzien van validiteit, zoals met name door Messick ... verwoord zijn, is uitgebreid. De relevantie van de verschillende eisen varieert met het specifieke doel van een test. Het werk van Messick biedt echter nauwelijks criteria met behulp waarvan de mate van relevantie van al deze eisen, gegeven een specifiek gebruiksdoel, valt te bepalen ... De afwezigheid van concrete richtlijnen voor het opzetten van een valideringsonderzoek maakt het moeilijk om onderzoeksvragen te prioriteren. Een bijkomend probleem is dat er veel bronnen van informatie zijn waar onderzoekers uit kunnen putten om na te gaan of een bepaalde interpretatie van een score op een test of toets juist is, of om een bepaalde vorm van gebruik te rechtvaardigen. In de meest recente versie van de ‘Standards for Educational and Psychological Testing’ staat dat *‘Resources should be invested in obtaining the combination of evidence that optimally reflects the value of a test for an intended purpose.’* ... De twee voornoemde probleme tezamen zorgen er echter voor dat nie expliciet vast te stellen is of in een valideringsonderzoek sprake is van een optimale of zelfs goede ‘combination of evidence’.

Van der Walt & Steyn (2007:140) voer verder aan dat die bepaling van geldigheid, soos ook hierbo genoem, deurlopend behoort te geskied. Empiriese data, versamel oor tyd heen, is inderdaad nodig om uitsprake oor die geldigheid van ’n toets te kan maak. Konteks moet ook altyd in ag geneem word, soos reeds in paragraaf 6.2.1 genoem. Uiteindelik is elke toets en die omstandighede daaromheen uniek en dit moet daarvolgens beoordeel word.

Dit bring die kwessie van validering ('n subjektiewe proses) wat staan naas geldigheid ('n objektiewe, meetbare eienskap) weer ter sprake en wat in meer besonderhede in Hoofstuk 7 aangespreek sal word. Vir nou is dit voldoende om daarop te wys dat die probleem met hierdie twee begrippe, volgens die Messick-perspektief, is dat dit ineenvloei en daarom vervaag. Weideman (2009a:243) verduidelik dit as volg:

... the distinction between the subjective process of validation and the objective validity of a test is an essential one, which seems to be forgotten in some ... discussions ... 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 have lacked not only validity, but also interpretability.

Lank voor Van der Walt en Steyn se opmerking hierbo word dit reeds in die werk van bekende taaltoetskundiges soos Bachman en Palmer (1996:38), Davies (1990:6) en Alderson *et al.* (1995:180) geïmpliseer dat geldigheid liever gesien moet word net vir wat dit is, naamlik 'n soeke na juistheid deur 'n verskeidenheid van toetsattribute te ondersoek. Die presiese probleem met 'n verenigde benadering tot geldigheid is dat dit alles begripsmatig onder geldigheid wil integreer ten einde teoreties regverdigbare interpretasies van toetsresultate te kan maak. Dit is bes moontlik as gevolg van sy (Messick se) poging om weg te beweeg van 'n oordrewe klem op die aritmeties-bewysbare, en om sosiale verantwoordbaarheid tot sy reg te laat kom. In die proses trek hy toe egter ook terme soos adekwaatheid (wat in analoë sin waarskynlik net 'n eufemisme vir krag of geldigheid is) en toepaslikheid ('n sosiale analogie) by sy verenigde raamwerk in. Die dilemma hiermee is dat hy eers alles onder geldigheid intrek en dan weer geldigheid gelykstel met interpretasie wat daartoe bygedra het dat konseptuele verwarring ontstaan het.

Desnieteenstaande sy pogings, is Messick aanvanklik hieroor gekritiseer deur Moss (1994; 1996) wat aangevoer het dat daar steeds 'n oordrewe klem op die empiriese (die bewysbare) en nie die hermeneutiese (die verklarende dus) is nie.

Met verloop van tyd het dit egter duidelik geword dat in stede daarvan dat Messick en ander soos Bachman en Palmer dit kon bewerkstellig om aan te toon hoe die kwaliteite / attribute van 'n toets in wisselwerking tot mekaar staan, dit eerder veroorsaak het dat konseptuele skerpeheid verlore gegaan het, wat ons terugbring by die tradisioneel-onderskeibare tipes geldigheid. Binne die konteks van die tweede paradigma is daar dan weinig ander keuse as dat onderskeibare tipes geldigheid bloot weer bewys moet word aan die hand van sogenaamde hipoteses, sodat die resultate van die toets minstens sosiaal verantwoordbaar sal wees. Davies en Elder (2005:798) beskryf hierdie situasie as volg:

... in spite of the unitary view now taken of validity, it has to be operationalized through the usual suspects of content and construct validity, concurrent and predictive validity ... To the usual suspects we need to add reliability; some will want to include face validity. However, the joker is validity ... called consequential validity.

Weideman (2009c:241) sluit hierby aan en argumenteer dat “[there is a] need to be critical of notions that subsume others. Conceptual clarity is essential – indeed, a necessary condition for test design ...” Ten einde dit te bewerkstellig bied dieselfde outeur (Weideman, 2009a:240) gevolglik nog 'n interpretasie van Messick se matriks oor geldigheid aan wat toon dat Messick se stellings nie meer oor geldigheid *per se* gaan nie, maar eerder oor die interaksie wat verskillende aspekte van toetse met mekaar het en fundamenteel is tot die interpretasie van toetsresultate. Dit word in Tabel 6.3 en die deel wat daarop volg aangehaal.

Tabel 6.3: Weideman se herinterpretasie van Messick se benadering tot geldigheid met 'n fokus op die verband tussen fundamentele oorwegings in taaltoetsing

	<i>Adequacy of ...</i>	<i>Appropriateness of ...</i>
inferences made from test scores	depends on multiple sources of 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

The matrix ... can be read as a number of claims about or requirements for language testing, as follows (left to right, top to bottom):

- ... The technical adequacy of inferences made from test scores depends on multiple sources of empirical evidence.
- ... The appropriateness of inferences made on from test scores relates to the detrimental or beneficial impact or consequences that the use of a test will have.
- ... 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.
- ... 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.

Uit bostaande is dit duidelik dat die Messick-konsepte, soos reeds genoem, nie noodwendig in 'n geldigheidsmatriks as sulks ingedwing hoef te word nie, en dat konstrugeldigheid nie meer in elkeen van die selle hoef voor te kom nie, hoewel die konsepte steeds van Messick se teorie as basiese vertrekpunt gebruik maak. Die matriks in Tabel 6.3 het eerder ten doel om aan te toon dat daar koherensie en passing is tussen 'n aantal onderskeibare konsepte wat fundamenteel tot enige toets behoort te wees – let egter daarop dat sommige van hierdie konsepte steeds te make het met die krag van 'n toets en dus geldigheidsaspekte kan insluit. Weideman beweeg met hierdie herinterpretasie van Messick dus weg van die oordrewe klem op konstrugeldigheid en verskuif die aandag na 'n konseptueel meer gefokusde benadering. Hy voer in die verband aan dat “... if one does not deliberately distinguish what is conceptually distinct, the distinction so avoided subsequently obtrudes itself upon the conceptual analysis.”

Hierdie meer gefokusde benadering waarop Weideman klem lê, sluit onder andere aspekte van toetse soos betroubaarheid, geldigheid, bruikbaarheid, impak, verantwoordbaarheid, en regverdiging in. Die verenigde benadering tot geldigheid van Messick skaad myns insiens, en in aansluiting by Weideman hierbo, die belang van ander fundamentele aspekte van toetse. Weideman se herinterpretasie bied 'n kritiese blik op Messick wat broodnodig is in die huidige konteks waar teoretici hom vandag steeds slaafs navolg. Weideman (2009a:240-241) som dit die beste op met die woorde “[s]urely 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...” Waarna hier verwys word in die begrippe ‘technical adequacy’, ‘appropriateness’, ‘technical meaningfulness’, ‘utility’, ‘relevance’, en ‘defensibility’ is niks anders as die elementêre begrippe wat die ruggraat vorm van die teoretiese raamwerk van hierdie studie nie. Dit is tegniese krag (adekwaatheid) – die verwysing binne die tegniese sy van ons bestaan na die fisies-energetiese, tegniese toepaslikheid (’n sosiale analogie), tegniese sinvolheid (linguale antisipasie), tegniese bruikbaarheid (ekonomiese verwysing), relevansie (weereens ’n sosiale analogie), en openbare regverdiging van die meetinstrument (juridiese analogie). Dit is duidelik dat Messick al hierdie begrippe gebruik, maar dit is tog jammer dat hy dit as onderdele van een (geldigheid) beskou.

In teenstelling met Weideman se oorwoë uitsprake hieroor, gaan nog ander soos Fulcher en Davidson (2007:279), en Borsboom, Mellenburgh en Van Heerden (2004:3) selfs verder en stel dit eksplisiet, naamlik dat die Messick-perspektief foutief en selfs onsinnig is. Fulcher en Davidson (2007:279) vra, sonder enige omhaal van woorde

[h]as this validity-as-interpretation mantra perhaps become over-used? If a test is typically used for the same inferential decisions, over and over again, and if there is no evidence that it is being used for the wrong decisions, could we not speak to the validity of that particular test – as a characteristic of it?

6.3 Samevatting

Uit die bostaande bespreking van die drie paradigmas behoort dit duidelik te wees dat geldigheid vanuit verskillende perspektiewe benader kan word. Binne konteks van hierdie studie blyk dit wel die beste te wees om die benadering van die derde paradigma te volg, omdat dit ’n meer gebalanseerde, en gefokusde, kyk op geldigheid bied as enige van die ander paradigmas. Dit kom in wese daarop neer dat geldigheid inderdaad ’n attribuut van ’n toets kan wees. Hoewel die tweede paradigma, naamlik dat geldigheid in wese oor die interpretasie van toetsresultate gaan, wyd aanvaar word, moet daar in ag geneem word dat dit afhanklik is van die doel waarvoor die toets ontwerp is. Verder is daar verskeie faktore (intern en ekstern) wat toetsresultate kan beïnvloed wat weer die eerste paradigma se denke

sterk op die voorgrond laat tree. Myns insiens behoort, soos wat Borsboom, Mellenburgh en Van Heerden (2004) voorstel, 'n volledige terugbeweeg na die tradisionele benadering tot geldigheid egter nie nodig te wees nie; dit is trouens nie geregverdig nie, omdat daar wel 'n saak vir die inagneming van sosiale aspekte uit te maak is.

Uit die bostaande bespreking, waar geldigheid gedefinieer is vanuit verskillende perspektiewe, is dit belangrik om daarop te let dat daar inderdaad 'n saak uit te maak is vir onderskeibare konsepte soos nut, regverdiging, betroubaarheid en geldigheid. Al hierdie konsepte sluit egter ook by mekaar aan in die opsig dat dit die belang van konstrugeldigheid beklemtoon, wat Messick ter ere nagegee moet word. Teen die konteks van die teoretiese raamwerk wat vir hierdie studie gebruik word, is dit egter nodig om versigtig te wees om konstrugeldigheid as die begin en einde van enige toets te ag, omdat, soos Weideman (2009a:249) dit stel,

... the constitutive concept of technical validity can be enriched by articulation of the theoretical idea or rationale for a test. It can also be further enriched and opened up when one considers the notions of its social results or impact. But such enrichment does not constitute either a basis for privileging the concept, or for subsuming everything under it. It merely points to the unfolding or opening up of the design to the regulative conditions for language testing.

Dit blyk dus sinvol te wees om geldigheid binne die derde paradigma te ondersoek waar 'n oper en 'n meer buigbare benadering tot toetsontwikkeling en –interpretasie aan die orde is. Die rede hiervoor is dat hierdie paradigma juis daardie elemente uit die eerste en tweede paradigmas byeenbring wat objektief sal bydra tot geldigheid, maar in balans met mekaar en as deel van die subjektiewe proses van validering.

Hierdie hoofstuk het ten slotte aangetoon waarom dit belangrik is dat konsepte wat onderskeibaar is liefst apart van mekaar gehou behoort te word, naamlik om konseptuele skerpheid na te streef wat sal bydra tot 'n beter begrip van die funderende en leidende kante van die bestaan van toetse. Slegs wanneer die konstitutiewe en die regulatiewe momente in wisselwerking met mekaar kom en daar gepoog word om 'n balans te bewerkstellig, met bepaalde toegewings na die

een of die ander kant, kan tegniese ontwerpe soos taaltoetse met vrug en sekerheid gebruik word.

Die konsep geldigheid as deel van die proses van validering sal daarom verder en op gefokusde wyse in die volgende hoofstuk uitgepak en ondersoek word, met die doel om ter wille van konseptuele skerpheid voldoende bewyse (ook sosiaal van aard) te versamel vir elke oordeel wat gevel moet word. Vanselfsprekend sal dit alles bydraend wees tot die formulering van 'n geldigheidsargument vir TAG (en TALL).

Validering: 'n eenheid in 'n verskeidenheid van argumente

7.1 Inleidend

In die vorige hoofstuk is die begrip geldigheid vanuit drie paradigmas gedefinieer, te wete 'n tradisionele paradigma, die huidige ortodoksie en 'n verruimde standpunt daaroor, naamlik dat dit terselfdertyd 'n attribuut van 'n toets kan wees en as konstitutiewe begrip ontsluit kan word deur verdere (regulatiewe) idees met betrekking tot taaltoetsing. Hoofstuk 6 was voorts teoreties-inleidend tot hierdie hoofstuk waar die fokus sal wees op toetsbekragtiging of validering en die ondersteuning wat 'n verskeidenheid gegewens en data daarvoor sal kan bied.

Die konsep geldigheid word gevolglik as deel van die proses van validering verder op sistematiese wyse in hierdie hoofstuk ondersoek. Verskillende soorte geldigheid en gelyklopende bekragtigingsprosesse word ter wille van konseptuele skerpheid ondersoek ten einde vas te stel of TAG werklik meet wat dit voorgee om te meet, naamlik akademiese geletterdheid. 'n Aantal statistiese tegnieke word gebruik om toetsdata te orden en te vertolk, op grond waarvan sekere gevolgtrekkings oor die geldigheid van TAG gemaak kan word.

Hierdie hoofstuk is in sy geheel, binne konteks van veral Hoofstukke 5, 6, en 8) die artikulasie van 'n geldigheidsargument vir TAG (en TALL) waartoe 'n veelheid van

perspektiewe bydra en wat in verband gebring sal word met die teoretiese raamwerk waarteen hierdie studie hom afspeel, naamlik dat dit geregverdig is dat toegepaste linguistiese of tegniese ontwerpe (soos TAG en TALL) gebruik kan word by die plasing van Suid-Afrikaanse eerstejaarstudente.

7.2 Toetsvalidering gedefinieer

Soos reeds hierbo vermeld, word geldigheid in die gangbare mening bewerkstellig deur 'n proses van validering. Dit sluit volgens Van der Walt en Steyn (2007:140) “a chain of inferences” in wat die abstraktheid van die konsep geldigheid operasionaliseerbaar, en gevolglik meer verstaanbaar, maak. Dit is dus die versameling van 'n veelheid van toetsverwante argumente en datastelle - 'n verskeidenheid van bronne - ten einde 'n koherente argument vir die geldigheid van 'n toets te kan opbou.

Deesdae, veral in 'n hedendaagse Suid-Afrika, is dit egter belangrik dat dit deursigtig en verantwoordbaar gedoen moet word, of soos Cronbach (1988:3) dit heel komies stel, “[v]alidation was once a priestly mystery, a ritual performed behind the scenes, with the professional elite as witness and judge. Today it is a public spectacle combining the attractions of chess and mud wrestling.” Yeld (2001:191) druk dit as volg uit: “rigorous validation [needs to be] conducted, not only so that candidates and the public can be confident that fair judgements are being made, but also so that developers and users can defend their tests and the decisions made on the basis of performance on these tests”, wat weerklank vind in Bachman en Palmer (1996:21) se bewering dat validering gedoen word om te verseker dat toetsontwikkelaars daartoe in staat gestel word om voldoende regverdiging te verskaf vir besluite wat op grond van toetsresultate geneem word. Kane (1992) is selfs pedanties daaroor wanneer hy in sy artikel hieroor die kwessie van interpreteerbaarheid as die enigste fokus van validering beskryf. Hy suggereer wel dat validering sistematies en weldeurdag aangepak moet word, maar dan kan hy hom nie verhelp nie om weereens, in lyn met Messick wat in die vorige hoofstuk

bespreek is, te sê dat die enigste doel daarvan is om 'n interpretatiewe argument daar te stel wat gebruik kan word om besluite oor toetslinge te regverdig.

Nitko (2004:34), Bachman (2005:267), Robinson-Kurpius en Stafford (2006:141), en Miller, Linn en Gronlund (2009:73) vind byna twee dekades later steeds artikulasie met die bogenoemde, wat die invloedrykheid van Messick en sy verenigde benadering tot geldigheid met 'n klem op die interpretatiewe bevestig. In hierdie sienings, wat in die heersende ortodoksie (die tweede paradigma – verwys Hoofstuk 6) uitdrukking vind, word toetsresultate die objektiewe basis vir geldige afleidings daaroor / die interpretasie daarvan (Bachman & Palmer, 1996:19-20). Die internasionaal aanvaarde en mees gesaghebbende bron hieroor, naamlik die *Standards for Educational and Psychological Testing*, stel dit dat “the process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations” (American Educational Research Association, American Psychological Association, National Council for Measurement in Education, 1999:9). Die vaagheidsprobleem wat Sluijter (1998:72) uitwys, en wat hierbo in die vorige hoofstuk in meer besonderhede aan bod gekom het, is egter steeds van toepassing, naamlik dat die definisie vir 'n ‘sound scientific basis for the proposed score interpretations’ nie duidelik is nie – wat word byvoorbeeld geag as adekwate bewyse en in watter kontekste is dit toepaslik om te gebruik?

Vanuit 'n derde paradigma-perspektief beskou (verwys weereens Hoofstuk 6), bly die vraag dus steeds of 'n subjektiewe proses van validering waarin objektiewe bewyse / resultate vir geldigheid versamel word, haalbaar is – die een is tog immers 'n funksie van die ander. Is hierdie twee kante van die saak in disjunksie met mekaar, of vul dit mekaar dalk juis aan? Weideman (2009a:242-243) argumenteer dat daar juis 'n onderskeid tussen die subjektiewe proses van validering en die objektiewe geldigheid van 'n toets getref moet word ten einde adekwate en toepaslike afleidings oor toetsresultate te kan maak, wat in 'n tweede paradigma-perspektief vervaag, omdat alles met mekaar verenig word (sien ook die verwysing na Sluijter in die vorige paragraaf). Davies en Elder (2005:797) voer in die verband

aan dat enige toets met verloop van tyd, en wel aan die hand van deurlopende valideringsprosesse, 'n goeie reputasie kan opbou, en dat hierdie reputasie dan mettertyd aan niks anders toeskryfbaar is as aan die verworwe (objektiewe) geldigheid daarvan nie. Om te sê dat 'n toets in objektiewe sin geldigheid kan verwerf, is volgens dieselfde outeurs dus veel meer as net “a trick of semantics” (Davies & Elder, 2005:798).

Davies *et al.* (1999:220-221) se definisie vir validering is waarskynlik die mees gebalanseerde, omdat dit die belangrikste aspekte van al drie paradigmas byeenbring. Die krag van die tweede paradigma skemer egter steeds deur, maar tussen die lyne is dit duidelik dat vertrekpunte van die eerste en derde paradigmas nie uit die oog verloor is met die skryf van hierdie definisie nie (verwys ook die bespreking direk na die onderstaande aanhaling). Die genoemde outeurs voer aan dat

[v]alidation involves gathering and evaluating the evidence for the ... validity of a test when used for a given purpose. If the test is described as appropriate for a particular context or group of learners, for example, this claim should be evaluated on the basis of evidence derived from test scores, which should be produced by a representative sample of this population. ... Perhaps the most important question, however, to be considered in validation of a test is that of its construct validity. This involves analysing how test scores may be understood in relation to the theoretical framework underlying the construct the test is designed to measure. It thus aims to answer the fundamental (and fundamentally difficult) questions: ‘What exactly does the test claim to measure?’ ‘What inferences may be made about test takers?’ ‘On what basis may these inferences be made?’ ... Other issues to be considered in test validation relate to [among others] ... ethical questions of how tests are used, including consequential validity (Davies *et al.*, 1999:220-221).

Woordkeuses soos ‘a given purpose’ en ‘appropriate for a particular context or group’ skakel met sienings van die eerste paradigma wat in die vorige hoofstuk bespreek is. Die fokus op konstrugeldigheid en 'n woordkeuse soos ‘inferences’ dui op die interpreteerbaarheid van toetsresultate ('n tweede paradigmaperspektief). 'n Onderbeklemtoning van begrippe soos ‘ethical’ en ‘consequential validity’ wat in 'n sin voorkom wat met die woorde ‘Other issues’ amper terloops ingelei word, wys moontlik daarop dat hulle juis nie 'n oordrewe klem wil plaas op die gevolge van die interpretasie van toetsresultate nie, maar dit word tog genoem, wat 'n aanduiding daarvan is dat ons wel daarvan kennis moet neem.

Validering is ten slotte volgens Van der Walt en Steyn (2007:142) 'n proses waardeur 'n samehangende geldigheidsargument vir of teen 'n toets, interpretasies van resultate, of die implementeerbaarheid / bruikbaarheid van die instrument gemaak word. Dit gaan dus inderdaad oor die versameling van voldoende ('adequate') bewyse, sowel as 'n aanvaarbare / toepaslike ('appropriate') interpretasie van toetsresultate en die bruikbaarheid daarvan. Maar, omdat dit veral ook gaan oor wat 'n toets nie kan bied nie, oftewel waar die tegniese krag daarvan juis nie lê nie, is dit nodig om dit konseptueel skerp en gefokus aan te bied. Dit moet met ander woorde liefs vanuit 'n derde paradigma-perspektief benader word, omdat die werklike sterkte van die geldigheidsargument uiteindelik tog maar sal berus op die swakste deel daarvan (Kane, Crooks & Cohen, 1999:15). 'n Geldigheidsargument moet daarom volgens Fulcher en Davidson (2007:20) aan die volgende eise voldoen: (i) die eis van eenvoud (konseptuele skerpheid / fokus), (ii) die eis van koherensie (byeenbring van verskillende aspekte tot 'n samehangende geheel), (iii) die eis van bewysbaarheid (voorspellings kan slegs gemaak word op veranderlikes wat toetsbaar is), en (iv) die eis van omvattendheid (volledigheid van die argument).

Daar sal verder in hierdie hoofstuk gepoog word om 'n geldigheidsargument vir TAG te konstrueer, wat volgens Davies en Elder (2005:802-803) jammerlik genoeg nie gereeld gebeur nie. Bogenoemde eise vir 'n geldigheidsargument sal ook nagestreef word. Die argument sal daarom deur 'n proses van validering opgebou word deur 'n *a posteriori* benadering te volg, en wel op 'n pragmatiese wyse soos voorgestel deur Van der Walt en Steyn (2007:142). Ter wille van konseptuele skerpheid, en in lyn met die derde paradigma, sal dit sistematies gedoen word deur hipoteses te formuleer vir onderskeidelik die interne en die eksterne geldigheid van die toets, wat die fokus volgens Davies en Elder (2005:798) net maar weer terugskuif na die "usual suspects": interne geldigheid sluit gesigs-, inhouds- en konstruktorgeldigheid in, terwyl eksterne geldigheid kriterium-gerigte geldigheid behels wat samevallende geldigheid, voorspellingsgeldigheid, en soms konsekwensiële geldigheid insluit.

7.3 'n Geldigheidsargument vir TAG

Daar word in hierdie afdeling gepoog om op 'n sistematiese wyse 'n geldigheidsargument vir TAG te konstrueer wat aan 'n aantal eise sal voldoen. 'n Aantal hipoteses sal ondersoek word ten einde 'n verskeidenheid bewyse te versamel en te sistematiseer vir die teoretiese regverdiging van TAG, sonder 'n oordrewe klem op die interpretatiewe funksie van toetsresultate, soos wat te verwagte is in 'n benadering wat geskoei is op die derde paradigma wat in die vorige hoofstuk bespreek is. Daar sal dus nie weer op die kwessies van adekwaatheid, toepaslikheid en bruikbaarheid van toetsresultate *per se* gefokus word nie (cf. afdelings 6.1 en 6.2 hierbo vir 'n uiteensetting). 'n Meer pragmatiese benadering soos voorgestel deur Fulcher en Davidson (2007:18) sal eerder gevolg word ten einde die argument op te bou. Die rede hiervoor is dat dit waarskynlik die beste tot konseptuele skerpheid sal bydra. Hierdie gefokusde benadering sal ten doel hê om uiteindelik 'n teoreties regverdigbare, deursigtige en verantwoordbare bydrae tot die ondersoek van die geldigheid van TAG lewer.

7.3.1 Interne geldigheid

Interne geldigheid verwys na kwaliteite / attribute van 'n toets wat intern daartoe lê. Dit gaan dus onder andere oor sake soos betroubaarheid (interne konsistensie), toetsinhoud, die eienskappe van toetsitems, die tipes response wat items tot gevolg het, en hoe dit op sigwaarde vertoon (Davies *et al.* 1999:220-222). Interne geldigheid sluit dus gesigsgeldigheid, inhoudsgeldigheid en konstrukgeldigheid in. Interne geldigheid is oorwegend konseptueel van aard, maar kan ook met statistiese bewerkings bevestig word, veral waar dit by konstrukgeldigheid kom.

7.3.1.1 Hipotese 1: TAG beskik oor gesigsgeldigheid

Hierdie aspek van 'n toets verwys nie na wat die toets werklik meet nie, maar wel na dit wat op sigwaarde gemeet word (Smit, 1991:49). Hier gaan dit dus oor hoe 'n toets daar uitsien en wat dit blyk te meet, en nie net na die oordeel van vakkundiges nie. Cronbach (1984:182-183) en Davies *et al.* (1999:59) wys daarop dat gesigsgeldigheid 'n eienskap / attribuut van 'n toets is wat ook vir 'n leek dit sal laat

lyk na 'n sinvolle meting van 'n bepaalde vaardigheid, kennis of domein wat relevant is vir 'n spesifieke doel. Hoewel gesigsgeldigheid normaalweg afgemaak word as 'n minder belangrike kenmerk van 'n toets, is dit volgens Davies *et al.* (1999:59) en Yeld (2001:243) wel belangrik en selfs kragtig, omdat die openbare beeld of legitimiteit ('n regulatiewe konsep) van 'n toets daarsonder erge skade kan ly. Dit kan weer lei tot die aanname dat die teoretiese onderbou van die toets (konstruk dus) onvoldoende is vir wat dit veronderstel is om te meet; die werklikheid moet dus so ver moontlik deur 'n taaltoets soos TAG gereflekteer word om dit op sigwaarde vir belanghebbendes soos universiteitsowerhede, ouers, studente self, ensovoorts aanvaarbaar te maak.

Gesigsgeldigheid word op een van twee maniere bepaal. Die eerste is om deursigtig oor die toets te wees en byvoorbeeld gereeld in die publieke domein daaroor te kommunikeer. In die geval van TAG en TALL is daar byvoorbeeld, sedert die ontstaan daarvan ses jaar gelede, reeds meer as 20 kongresreferate hieroor gelewer en 14 artikels in geakkrediteerde, eweknie-geëvalueerde, tydskrifte hieroor gepubliseer. Verder is daar universiteitsintern by die Universiteite van Pretoria en Stellenbosch, sowel as by die Noordwes Universiteit by verskeie fakulteitsraadsvergaderings, programkomiteevergaderings en toelatingskomiteevergaderings ekstensiewe terugvoer gegee oor toetsresultate en die interpretasies daarvan, besprekings oor die konstruk onderliggend aan TAG en TALL is ook by verskeie geleenthede gedoen en beide hierdie pogings tot deursigtigheid het besonder goeie aanklank gevind (cf. Hoofstuk 3 van hierdie studie en Van Dyk en Kistner, 2008). Voorts is daar voldoende inligting oor die toetse op van die universiteite se webwerwe, sowel as in verwelkomingsgidse / aansoekdokumentasie / eerstejaarsgidse, ensovoorts beskikbaar gestel. Benewens dit het daar ook al verskeie berigte in die populêre media oor TAG en TALL verskyn en van die toetsontwikkelaars het selfs al in radioprogramme oor die toetse gepraat.

Die tweede wyse waarop inligting oor gesigsgeldigheid versamel kan word, is om na afloop van 'n toetsadministrasie onderhoude met studente oor die toets te voer,

of om aan die hand van 'n vraelys inligting daaroor te versamel. Yeld (2001:244) beweer in die verband dat

[i]f the gathered data are sufficiently detailed, it is sometimes possible to relate information to test performance – for example, when a particular group of candidates reports finding certain items alienating, or novel, it is possible to establish through correlational analysis, whether these perceptions are related to performance on these items.

Van der Walt en Steyn (2007:149-151) het na afloop van die 2007-toetsgeleentheid aan die hand van 'n vraelys inligting oor TAG, en studente op die Potchefstroom-kampus van die Noordwes Universiteit se persepsies daaroor, ingesamel. Hieruit het dit geblyk dat deursigtigheid tog deurlopend aandag behoort te geniet en daar derhalwe nie net aanvaar kan word dat studente sal weet wat van hulle in die toetse verwag sal word nie. Verder is daar in hierdie ondersoek bevind dat 'n beduidende persentasie van toetslinge die toets as relevant tot hul studies ervaar het, maar dat ongeveer 20% van die studente wat deur die toets as risikokandidate uitgewys is nie die gepaardgaande, verpligte akademiese geletterdheidskursus uit eie keuse sou volg nie. Voorts is bevind dat tydsdruk 'n invloed op resultate gehad het en toetslinge het geoordeel dat daar meer tyd toegestaan kon gewees het.

Uit die bostaande kan daar afgelei word dat die toets op sigwaarde meet wat dit veronderstel is om te meet, naamlik akademiese geletterdheid, maar dat dit wat van studente in die toetse verwag word nog duideliker aan hulle geartikuleer en gekommunikeer kan word.

Verder is die feit dat ongeveer 20% van die studente nie die nut van 'n akademiese geletterdheidskursus (en dus by implikasie ook 'n akademiese geletterdheidstoets) kan insien nie ook kommerwekkend. Daar moet egter in ag geneem word dat hierdie ongeveer 20% uit die onderste 30% van die populasie as geheel kom en dit 'n invloed op response kon hê. Dit sal interessant wees om te sien hoe die populasie as geheel op so 'n vraag sou reageer, en ook hoe die verskillende populasies by die onderskeie instellings wat van TAG gebruikmaak daarop sou reageer, omdat die demografie van die studentepopulasie vir die onderskeie instellings verskil. Die insameling van longitudinale data sou ook 'n bydrae kon lewer tot indrukke oor hoe

gesigsgeldigheid verhoog kan word. Nog 'n moontlikheid vir verdere ondersoek is hoe studente se persepsies oor die toets, en die gepaardgaande intervensie, oor 'n jaar verander - Van Dyk, Zybrands, Cillié en Coetzee (2009) wys byvoorbeeld daarop dat studente se persepsies oor TAG en TALL en die gepaardgaande intervensies deur die loop van die eerstejaar blyk te verander, en wel in 'n positiewe rigting.

Hoewel Van der Walt en Steyn (2007:151) die kwessie van tydsdruk wat resultate negatief sou kon beïnvloed as ietwat onthutsend beskryf, is dit nie noodwendig rede tot kommer nie, aangesien dit die akademiese omgewing reël voorstel – die argument is juis dat eerstejaarstudente nie die mas kan opkom met die hoeveelheid lees- en skryfwerk waarmee hulle in die eerstejaar gekonfronteer word nie as gevolg van tydsbeperkinge (cf. Hoofstuk 1). Daar sou dus geargumenteer kon word dat die vermoë van 'n eerstejaarstudent om onder tydsdruk die lees- en skryfwerk wat van hom / haar verwag word te kan afhandel, opsigself 'n aanduiding van akademiese geletterdheid is.

Uit die bostaande behoort dit duidelik te wees dat TAG (en by implikasie dus ook TALL) oor gesigsgeldigheid beskik.

7.3.1.2 Hipotese 2: TAG beskik oor inhoudsgeldigheid

Inhoudsgeldigheid het te make met die inhoudelike van 'n toets, oftewel die substantiewe elemente soos Smit (1991:48) dit noem. Dit gaan oor die wyse waarop 'n toets daardie kennis of vaardighede meet wat dit veronderstel is om te meet en dit word verkry deur ondersoek in te stel na die mate waartoe die toetstake verteenwoordigend is van die domein wat dit veronderstel is om te meet, in hierdie geval die domein van akademiese geletterdheid. Davies *et al.* (1999:34) beskryf dit as 'n tipe geldigheid wat deur middel van 'n sistematiese analise van die toetsinhoud 'n aanduiding daarvan sal kan gee of dit verteenwoordigend van die domein is, al dan nie. Inhoudsgeldigheid is daarom ook, soos trouens alle ander tipes geldigheid, spesifiek tot die doel van die toets en nie algemeen van aard nie.

Hoe duideliker elke item in 'n toets dus die domein karteer, hoe groter is die kans op inhoudsgeldigheid. Yeld (2001:214) verduidelik inhoudsgeldigheid as

the operationalisation of the test construct. The challenge for content validity is thus to ensure that the behaviours and performances (the tasks and items) in the test, as well as the way they are scored, actually elicit and reflect the knowledge, skill, and other attributes [to be measured].

Die rede vir al die bogenoemde is dat daar volgens Miller, Linn en Gronlund, (2009:75) na afloop van 'n toets, en op grond van die resultate daarvan, veralgemenings daarvoor gemaak moet kan word – verwys afdeling 5.2.2.3.3 vir meer oor veralgemeenbaarheid. Inhoudsgeldigheid toon dus op die oog af 'n verband met gesigsgeldigheid, maar die verskil is dat dit hier gaan oor wat 'n toets werklik meet.

Ten einde 'n oordeel oor die inhoudsgeldigheid van 'n toets te kan vel, word die kundigheid van vakspesialiste normaalweg ingeroep, wat ook die geval met TAG en TALL is, waar vakkundiges op 'n jaarlikse basis vir ongeveer twee weke ('n week elke ses maande) byeenkom om in werkswinkelformaat bestaande toetse te beoordeel, aan te pas, en om nuwe toetse te ontwikkel. Die eerste fase van so 'n byeenkoms behels 'n uiteensetting en bespreking van die toetskonstruk en die toetsspesifikasies (verwys afdeling 3.4.1.3 vir die konstruk van TAG en TALL, sowel as Tabel 4.7 en Bylaag B vir 'n uiteensetting van die toetsspesifikasies). Daarna word 'n spesifieke toets aan die werkswinkelpersoneel / beoordelaars uitgedeel met die doel dat letterlik elke item teen die konstruk getakseer moet word. Dit word gedoen aan die hand van 'n matriks waarvan 'n uittreksel ter illustrasie in Tabel 7.1 hieronder voorgehou word. In die matriks word daar letterlik onder elke item (of kluster van items) aangetoon watter aspek / aspekte van die konstruk dit meet, deur merke in die ooreenstemmende kolom(me) te maak. Die basiese vertrekpunt is dat die konstruk in sy geheel deur die toets verteenwoordig moet word. Daar sal uit die aard van die saak sekere aspekte van die konstruk wees wat meer in items verteenwoordig word as ander, omdat bepaalde tekste hulle meer tot sekere items leen. Solank die toets egter in geheel verteenwoordigend is van die konstruk, is dit in orde.

Tabel 7.1: Matriks vir die bepaling van die konstrukverteenvoerdigende aard van die toets

Aspek van konstruk	Item 1	Item 2	Item 3	Item 4	Ensovoorts
begrip binne konteks van 'n omvattende akademiese woordeskatreeks					
die vertolking en gebruik van metafore en idiome, en begrip van bybetekenis, woordspeling en dubbelsinnigheid					
begrip van verhoudinge tussen verskillende teksdele, bewustheid van die logiese ontwikkeling van 'n (akademiese) teks vanaf inleiding tot gevolgtrekking, en kennis van taalgebruik wat die verskillende teksdele samehangend aanmekeer bind					
die vertolking van verskillende teksoorte (genre), en sensitiwiteit vir die betekenis wat daardeur oorgedra word en die gehoor op wie dit gemik is					
die vertolking, gebruik en skep van inligting in grafiese of visuele formaat					
<i>Ensovoorts</i>					

Naas die bostaande is daar ook ruimte vir kommentaar van óf die toetsontwikkelaar óf die beoordelaars op elke item, soos geïllustreer in Tabel 7.2. Dit sluit tipies kommentaar van die toetsontwikkelaars se kant af in oor die fasiliteitswaarde (die moeilikheidsgraad van 'n item) of diskriminasiewaarde (hoe goed die spesifieke item diskrimineer tussen die boonste en die onderste kwartiel van die populasie). Let in hierdie verband daarop dat dit as 'n vereiste gestel word dat geen item in 'n amptelike TAG of TALL opgeneem mag word as dit nie deur minstens een loodsfasie is nie. As parameters is 'n ideale fasiliteitswaarde van tussen 0.50 en 0.60 gestel (die item moet dus deur ongeveer 55% van die totale populasie reggekry word, maar uitsonderings en toegewings word in sekere kontekste tot by 0.30 en 0.80 gemaak). Wat diskriminasiewaardes betref, is parameters van 0.3 en hoër gestel. Beide hierdie stelle parameters is in lyn met internasionale norme. Van die kant van die beoordelaars kan dit weer kommentaar insluit oor byvoorbeeld die bevooroordeeldheid van die item, die duidelikheid van instruksies, die relevansie van die item, en die produktiwiteit daarvan ('n kombinasie van die belyning van die

item met die konstruk en die voldoening daarvan aan die vereistes van graad van moeilikheid en diskriminasie).

Tabel 7.2: Matriks vir beoordelaars / vakkundiges se kommentaar

Kommentaar op Item 1						
Moeilikhedsgraad	Diskriminasiewaarde	Bevoordeeldheid	Duidelikheid van instruksie	Relevansie van item	Produktiwiteit van item	Ander kommentaar

Na afloop van bostaande oefening word daar normaalweg ’n oordeel gevel oor konstrukondervteenwoordiging en aspekte wat tot konstruk-irrelevante variansie bydra – dit word aanvaar dat hierdie twee aspekte ’n bedreiging vir geldigheid inhou. Eersgenoemde is makliker om te bepaal, veral as prosesse soos wat hierbo vir TAG beskryf, gevolg word, en dui daarop dat die meting by geleentheid miskien nie wyd genoeg was nie en dus nie verteenwoordigend van die konstruk in sy geheel is nie. Volgens Miller, Linn en Gronlund (2009:81) dra konstrukondervteenwoordiging by daartoe dat “validity is reduced to the degree that important aspects of the construct are underrepresented in the assessment.” Laasgenoemde is egter ’n ingewikkelder konsep, wat daarop dui dat die meting weer te wyd was en aspekte meet wat nie relevant tot die domein is nie. Konstruk-irrelevante variansie verwys hier dus na faktore wat irrelevant is tot die konstruk en bydra tot ’n verlaging in geldigheid, of soos Miller, Linn en Gronlund (2009:81) dit stel: konstrukgeldigheid word negatief beïnvloed deur ’n gebrek aan vaardigheid in aspek X wat ’n invloed het op ’n toets wat Y meet. Die belang hiervan is benewens die feit dat konstruk-irrelevante faktore tot groter variansie in die toetsresultate sal

bydra, dit ook tot onregverdigheid teenoor die toetsling kan bydra. Daar moet dus gepoog word om konstruk-irrelevante faktore altyd so laag as moontlik te hou.

Messick (1995:742) wys daarop dat “[b]oth threats are operative in all assessments. Hence a primary validation concern is the extent to which the same assessment might underrepresent the focal construct while simultaneously contaminating the scores with construct-irrelevant variance.”

Twee tipes foute wat tot konstruk-irrelevante variansie bydra word geïdentifiseer, naamlik ‘construct-irrelevant difficulty’ en ‘construct-irrelevant easiness’. In eersgenoemde geval dra aspekte wat buite die domein wat gemeet word lê daartoe by dat ’n taaktipe of ’n item te moeilik is vir bepaalde individue of groepe. Dit moet egter nie verwar word met items wat juis goed behoort te diskrimineer tussen individue of groepe nie. Messick (1995:743) maak in hierdie verband die volgende stelling:

Indeed, construct-irrelevant difficulty for individuals and groups is a major source of bias in test scoring and interpretation and of unfairness in test use. Difference in construct-irrelevant difficulty for groups, as distinct from construct-relevant group differences, is the major culprit sought in analyses of differential item functioning.

Hierdie aspek vind ook uitdrukking in die analises en kommentaar in Van der Slik en Weideman (2010) se artikel wat tans ter perse en getiteld is: “Examining bias in a test of academic literacy: Does the Test of Academic Literacy Levels (TALL) treat students from English and African language backgrounds differently?”

Die teenoorgestelde is ook waar wanneer dit by ‘construct-irrelevant easiness’ kom: aspekte wat buite die domein wat gemeet word lê, dra daartoe by dat ’n taaktipe of ’n item besonder maklik is vir bepaalde individue of groepe. Insgelyks kan temas en subtemas in toetse bekend wees aan bepaalde individue en hul agtergrondkennis kan hul dus onregverdiglik bevoordeel.

Ten einde konstruk-irrelevante variansie so ver moontlik te vermy word daar relatief robuus omgegaan met die parameters wat vir TAG gestel is in terme van die

fasiliteitswaarde en die diskriminasiewaardes hierbo uiteengesit. Soos reeds genoem, word geen item in 'n amptelike TAG of TALL opgeneem as dit nie deur minstens een loodstoets is én aan die gestelde vereistes voldoen nie. Binne hierdie konteks bied Miller, Linn en Gronlund (2009:85) die volgende definisie vir konstrugeldigheid, wat 'n goeie sinopsis van die bostaande is:

Broadly conceived, construct validation is an attempt to account for the differences in assessment results. During the development of an assessment, an attempt is made to rule out extraneous factors that might distort the meaning of scores, and follow-up studies are conducted to verify the success of these attempts. The aim is to clarify the meaning of student performance by identifying the nature and strength of all factors influencing the scores on the assessment.

Ten slotte en ter wille van volledigheid is dit ook belangrik om versigtig daarvoor te wees om doelbewus te kapitaliseer op konstrug-irrelevante variansie, wat nie die geval by TAG en TALL is nie, omdat dit oor die algemeen as laebelangtoetse of hoogstens mediumbelangtoetse geag word. Messick (1995:743) stel dit dat 'n

complication arises when construct-irrelevant variance is deliberately capitalized upon to produce desired social consequences, as in score adjustments for minority groups, within-group norming, or sliding band procedures... However, recognizing that these adjustments distort the meaning of the construct as originally assessed, psychologists should distinguish such controversial procedures in applied testing practice ... from the valid assessment of focal constructs and from any score uses based on that construct meaning.

'n Derde stap, wat as deel van die bepaling van inhoudsgeldigheid gevolg word, is om die leestekste te beoordeel. Dit word ook aan die hand van 'n matriks gedoen, soos in Tabel 7.3 geïllustreer. Hierin verskyn weereens kommentaar van óf die toetsontwikkelaar óf die beoordelaars op elke teks. Met teks word bedoel die langer leestekste in die toets by tipies die leesbegripafdeling, sowel as die grafika-afdeling en die afdeling waar tekste by wyse van spreke geredigeer moet word om aan te toon dat daar begrip is vir kwessies oor grammatika en teksverhoudings. Kommentaar oor die leesbaarheid (wanneer 'n misindeks bepaal word) van die teks, die lengte van die teks en die graad van leesbaarheid word hier gegee. As parameters is dit byvoorbeeld vooraf gestel dat tekste 'n misindeks (soms ook 'n leesbaarheidsindeks genoem, wat 'n aanduiding van die gemak is waarmee die teks gelees kan word) van tussen 45% en 55% moet hê en dat dit op die vlak van 'n Graad 10 - 12 leerder moet wees, terwyl tekslengte tussen 450 en 600 woorde

behoort te wees. Verskillende tekste in 'n enkele toets behoort verder tematies so ver moontlik by mekaar aan te sluit, hoewel dit nie 'n vereiste is nie. Dit moet ook so algemeen as moontlik wees, met ander woorde nie dissipline-spesifiek nie. Kontroversiële onderwerpe, wat kwessies oor geloof, ras, geslag, seksuele voorkeure, politiek, misdaad, ensovoorts ter sprake bring moet liefs vermy word, aangesien dit emosiebelaaide temas is, wat kandidate kan ontstel, omdat dit vooraf hul lewens kon geraak het of steeds raak, en dus met hul resultate kan inmeng (verwys Figuur 5.2 en die bespreking daaromheen, sowel as bostaande uiteensetting oor konstruk-irrelevante variansie).

Tabel 7.3: Matriks vir beoordelaars / vakkundiges se kommentaar oor teks

Kommentaar op Teks 1			
Misindeks	Graadvlak	Lengte van die teks	Kommentaar ter verbetering (indien toepaslik)
Bv. 55%	Bv. Graad 11	Bv. 500 woorde	Bv. Te min diskoersmerkers

Die doel van hierdie matrikse is in die eerste instansie om inligting te versamel wat die hooftoetsontwikkelaar in staat sal stel om, in samewerking met die beoordelaars wat ook deel van die toetsontwikkelingspanne is, die toetse sodanig aan te pas dat dit die konstruk en die toetsspesifikasies verteenwoordig. Let egter daarop dat hoewel daar teoreties na letterlik elke item gekyk word, daar ook soms na klusters van items gekyk word, omdat een item afhanklik van die ander is. Afdeling 1 (Skommelteks) is 'n goeie voorbeeld hiervan waar die een item nie los van die ander in die afdeling beoordeel kan word nie.

Hierdie matrikse het die voordeel dat die reikwydte daarvan verder strek as net vir die bepaling van inhoudsgeldigheid. Dit kom byvoorbeeld ook handig te pas wanneer uitsprake oor die bruikbaarheid van items en die invloed daarvan op die konsekwensiële geldigheid van 'n toets gemaak moet word en daar onder andere op

die bevooroordeelende aard van items en tekste gelet word – sake wat tot konstruk-irrelevante variansie kan bydra. Verder gee dit ook 'n aanduiding van die gewigte wat tydens die ontwikkelingsproses onbewustelik aan sekere aspekte van die konstruk toegeken is in terme van die aantal vrae – so sou die matriks in Tabel 7.1 byvoorbeeld heel effektief kan uitwys dat woordeskat potensieel oorverteenswoordig is in die toets en iets soos metaforiek onderverteenswoordig is. Dit is veral ook belangrik om, ter wille van stabiliteit, toetse oor verloop van tyd met mekaar te vergelyk en te sien hoe verteenwoordigend die 2006 toets teenoor die 2007 toets is, ensovoorts. So word 'n geheelbeeld van die toetse se inhoudsgeldigheid opgebou.

Hoewel bogenoemde aangewend kan word as argumentasie vir die inhoudsgeldigheid van TAG, wys Yeld (2001:219) daarop dat 'n meer objektiewe beoordeling daarvan ook soms nodig is, aangesien beoordelaars konsensus oor 'n item kan bereik en dan is dit nie verteenwoordigend van die konstruk nie. Sy voer aan dat dit gebeur

on the basis of the dynamic the group has built up, where the opinions of certain members might carry more weight than others by force, for example, of seniority or personality, rather than by force of principle (Yeld, 2001:219).

Dit is daarom belangrik dat die hooftoetsontwikkelaar, wat gewoonlik ook die sameroeper van die genoemde werkswinkels is, sal oplet dat hierdie saak goed bestuur word, én dat hy / sy self nie idees op die groep sal afdwing nie, sonder om op sy / haar kundigheid en ervaring in te boet.

Smit (1991:50) voer aan dat inhoudsgeldigheid ook bepaal kan word deur die verband te bereken tussen twee vorme van dieselfde toets, wat niks anders is nie as ekwivalente-vorm-betroubaarheid (let hier weer op die onlosmaaklike verband tussen betroubaarheid en geldigheid). Soos egter reeds in afdeling 5.2.2.2 aangevoer, was dit nog nie moontlik om 'n uitvoerige ondersoek in die verband te doen nie, maar uit 'n voorlopige en informele analise het dit geblyk dat daar 'n korrelasie van 0.48 is met 'n p-waarde van $p \leq 0.001$ vir drie afdelings (skommeltekst, woordeskat en tekstipe) van TAG wat met mekaar vergelykbaar was. Hoe hoër die korrelasie, hoe hoër die inhoudsgeldigheid van die toets. Die feit dat hier 'n relatief

hoë korrelasie sigbaar is en dat dit hoogs beduidend is, gee wel indrukke van inhoudsgeldigheid, veral as in ag geneem word dat slegs 'n deel van die twee toetse met mekaar vergelyk kon word.

Verder kan daar ook weer verwys word na die relatiewe homogeniteit van die konstruk soos uiteengesit in afdeling 5.2.2.3.2 (cf. ook Van der Slik en Weideman, 2005:30). As alle items in 'n toets 'n enkele domein meet, dra dit nie net tot betroubaarheid en konstrukgeldigheid by nie, maar ook tot inhoudsgeldigheid.

Op hierdie punt sou daar dus geargumenteer kon word dat TAG, gegewe die fases waardeur daar gegaan word in die proses van toetsbeoordeling en toetsontwikkeling, oor inhoudsgeldigheid beskik. Die rede is dat die sorgsaamheid waarmee die toetse ontwikkel, beoordeel en verfyn word, bygedra het daartoe dat TAG sedert die ontstaan daarvan verteenwoordigend van die konstruk was, en dat sistematiese en onsistematiese fluktuasie so ver moontlik tot die minimum beperk is. Boonop beskik dit oor 'n mate van deskundige beoordeling in terme van die inhoudsgeldigheid daarvan, soos hierbo gesien. Hier is dit ook belangrik om daarop te let dat items wat nie aan die basiese vereistes en parameters gestel, voldoen nie, en deur die beoordelaars as irrelevant tot die konstruk geag word, geskrap word en dit kan bydra daartoe dat inhoudsgeldigheid gemakliker verkry kan word.

In aansluiting by bogenoemde bespreking is dit die moeite werd om 'n minder bekende aspek van inhoudsgeldigheid, en wat nog by TAG tekortskiet, te vermeld. Die kwessie van konstrukrelevansie, naas konstrukverteenvoording (soos hierbo uiteengesit) sou daarom in verdere ondersoeke nagevors kon word. Yeld (2001:222-243) bied 'n teoretiese uiteensetting van konstrukrelevansie waarin sy verduidelik dat taaktipes oor bepaalde kenmerke moet beskik en dat hierdie kenmerke in drie breë kategorieë verdeel kan word, naamlik situasie, materiaal en rubrieke. Onder elkeen van hierdie kategorieë is daar weer 'n verskeidenheid van sub-kategorieë (cf. Yeld, 2001:229). Hoewel die genoemde outeur hierdie kategorieë bespreek het, was dit meer toevallig van aard, eerder as wat dit in die bespreking aan deeglike sistematiek en ondersoek onderwerp is. Die waarde van

haar bespreking lê egter daarin dat dit wel aantoon dat 'n volledige argument oor die inhoudsgeldigheid nie gevoer kan word sonder uitsprake oor konstrukrelevansie nie. Van hierdie kategorieë en subkategorieë sou dus inderdaad ook vir TAG bespreek kon word, maar omdat hierdie studie nie op inhoudsgeldigheid *per se* fokus nie sou 'n bloot terloopse bespreking onbevredigend wees en daarom is besluit om dit nie hier verder te voer nie. Die meriete van so 'n ondersoek kan egter glad nie onderspeel word nie, maar dit sal deeglike beplanning en sistematiek vereis.

7.3.1.3 Hipotese 3: TAG beskik oor konstrukgeldigheid

Die konsep konstrukgeldigheid het, soos reeds in die vorige hoofstuk genoem, in 1955 prominent geraak met die publikasie van Cronbach en Meehl (1955) se seminale werk daaroor. Hoewel die konsep enkele jare tevore reeds benoem en ondersoek is, het bogenoemde werk die weg gebaan vir vele ander publikasies daaroor. Die basiese definisie en beginsels van konstrukgeldigheid, soos deur Cronbach en Meehl (1955) beskryf, word vandag steeds aanvaar en met groot vrug gebruik, hoewel dit sterker op die teoretiese as die praktiese fokus. Bachman (1990:6) stel dit byvoorbeeld dat konstrukvalidering hedendaags steeds sentraal staan tot navorsing in taaltoetsing, omdat dit ondersoek instel na die verband tussen prestasie op die toets en die vaardighede waarop dit gebaseer is. Dit is trouens belangrik om te bepaal of 'n toetsresultaat 'n ware voorspeller is van toekomstige prestasie en of dit bloot toevallig is (Bachman, 1990:290; McNamara, 2000:104).

Volgens Bachman en Palmer (1996:21) en Messick (1996:243) is genoegsame empiriese bewyse, ironies genoeg as hul standpunt oor die interpretatiewe in ag geneem word, nodig om 'n toetsresultaat regverdiglik te interpreteer. Konstrukgeldigheid verwys volgens hulle dus na die mate waartoe 'n toetsresultaat interpreteerbaar is as 'n indikator van daardie vermoëns (of konstruk) van 'n spesifieke domein wat gemeet word (Bachman & Palmer, 1996:22), maar daar moet altyd in ag geneem word dat geen toetsresultaat ooit 100% geldig geïnterpreteer kan word nie. Konstrukvalidering in hierdie opvatting gaan dus oor die proses wat

gevolg word om te bepaal na watter mate 'n toetsresultaat teen die konstruk geïnterpreteer kan word.

Hoewel interpretasie en resultate vanselfsprekend saak maak, moet die konteks van dit wat in Hoofstuk 6 bespreek is egter hier weer uitgelig word, naamlik dat konstrukgeldigheid in die eerste plek nie gaan oor die toetsresultate (die tegniese effekte of uitkomst van die toets) of hulle interpretasie (die tegniese artikulasie van hul betekenis) nie, maar oor die teoretiese regverdigbaarheid (die tegnies-analitiese dimensie) van die onderliggende teorie van wat taalvermoë behels. Om begripsmatig dus skerper te werk te gaan, moet daar dus liefers 'n onderskeid getref word tussen tegniese effek, tegniese interpretasie, en tegnies-teoretiese verdedigbaarheid. Hierdie drie elemente van die raamwerk eggo begrippe wat afkomstig is van drie stelle relasies tussen die tegniese en die afsonderlike bestaansmodusse, te wete die energetiese, die linguale en die analitiese dimensies van die werklikheid. Die konsep van konstrukgeldigheid wat uit hierdie raamwerk voortvloei staan dus veel nader aan die van Davies *et al.* (1999:33) wat dit stel dat “[t]he construct validity of a language test is an indication of how representative it is of an underlying theory of language ...” Daar word egter ook soms na konstrukgeldigheid as interne strukturele geldigheid verwys, wat 'n aanduiding is van die mate waartoe alle items in 'n toets verband hou met mekaar, sowel as 'n meting van die teoretiese onderbou van die toets, met ander woorde 'n bepaling van wat dit konseptueel veronderstel is om te meet (Robinson-Kurpius & Stafford, 2006:150). Hierdie definisie stel dit dus ietwat breër as wat Davies *et al.* hierbo gedoen het, maar daar moet opgelet word dat die uitgebreide definisie wel ook geïmpliseerd in die werk van Davies *et al.* voorkom.

Beide die teorie onderliggend tot die konstruk en die eienskappe van die toetstake moet dus in ag geneem word wanneer toetsresultate geïnterpreteer word. Daar word twee redes vir die analise van toetstake aangevoer, waarvan die eerste is dat daar bepaal moet word hoe na aan, of hoe ver van die toetstaak die werklike

taalgebruikssituasie is, en die tweede die mate waartoe die toetstaak 'n verband toon met die toetsling se taalvermoë (Bachman & Palmer, 1996:21).

Volgens Messick (1995:742-743; 1996:244-245) is daar twee bedreigings vir konstrugeldigheid. Die eerste is konstrug-onderverteenwoordiging en die tweede konstrug-irrelevante variansie. Beide hierdie konsepte is hierbo reeds bespreek en word nie hier weer in besonderhede behandel nie.

Ter afsluiting van hierdie inleidende deel tot konstrugeldigheid is dit belangrik om daarop te let dat hierdie konsep inderdaad blyk wyer te strek as gesigs- en inhoudsgeldigheid, aangesien, soos ook in die vorige hoofstuk genoem, dit “empirically verifies (or falsifies) hypotheses derived from a theory of factors that affect performance on tests – constructs, or abilities, and characteristics of the test method” (Bachman, 1990:290). Dit is om hierdie rede wel te verstane dat Messick (1980; 1981; 1988; 1989; 1994; 1995; 1996) aanhou om 'n verenigde benadering tot geldigheid voor te staan, en weens sy statuur en so ook die van ander invloedrykes in die taaltoetsgemeenskap, word dit slaafs nagevolg. Uit die voorafgaande behoort dit duidelik te wees dat daar wel oorvleueling is tussen aspekte van konstrugeldigheid, en gesigs- en inhoudsgeldigheid en dit moet Messick *c.s.* wel toegegee word. Die risiko met die verenigde benadering tot geldigheid, soos uitvoerig in die vorige hoofstuk bespreek en ook reeds in hierdie hoofstuk aangedui, is dat dit bydra tot konseptuele vaagheid, of selfs konseptuele verwarring, soos in Borsboom *et al.* (2004) en Weideman (2009a) uiteengesit. Die probleem met konseptuele vaagheid lê daarin dat taaltoetsers binne die hedendaagse ortodoksie geneig is om een of twee valideringseksperimente uit te voer en dit dan as bewyse vir konstrugeldigheid aan te bied, en siende dat konstrugeldigheid ‘an overarching concept’ is, om Messick, Bachman en Kane se terme te gebruik, word die toets dan geag om geldig te wees, terwyl daar soveel aspekte van bewyslewering is wat glad nie verder ondersoek word nie. So gebeur dit byvoorbeeld gereeld dat toetsers, juis omdat die konsep geldigheid en die proses van validering nie gefokus genoeg benader word nie, toetsresultate interpreteer as absoluut en onfeilbaar. Die

adekwaatheid en toepaslikheid daarvan kom nie sterk genoeg in fokus nie as gevolg van 'n gebrek aan konseptuele skerpheid (Popham, 1997:11). Nog 'n rede hiervoor, soos ook in die vorige hoofstuk aangedui, is dat die verenigde benadering tot geldigheid besonder verwarrend is en toetspraktisyns gewoonlik nie weet watter roete om te volg nie. Dit word, soos ook reeds genoem, bevestig deur Davies en Elder (2005:802–803) se bekommernis dat daar te min omvattende valideringstudies gedoen word – dus te min in kwantiteit én in omvang.

7.3.1.3.1 Sub-hipotese 3a: TAG beskik oor interne konsekwentheid

Soos reeds in Hoofstuk 5 aangedui is betroubaarheid 'n nodige dog nie 'n voldoende kenmerk van 'n toets nie en dat 'n toets betroubaar kan wees sonder om geldig te wees, maar dat die omgekeerde nie waar is nie. Verder het dit ook uit die voorafgaande bespreking geblyk dat daar 'n verband tussen betroubaarheid en geldigheid is. Die verband word bevestig in die woorde van Bachman (1990:238-239) dat betroubaarheid en geldigheid komplementêrend ten opsigte van mekaar is, en dieselfde outeur som dit as volg op: “Reliability is the agreement between two efforts to measure the same trait through maximally similar methods. Validity is represented in the agreement between two attempts to measure the same trait through maximally different methods (Bachman, 1990:240).” Betroubaarheid gaan in wese dus daarvoor of 'n toets konsistente resultate oplewer, en indien 'n toets nie konsistente resultate oplewer nie, is die interpretasie daarvan problematies wat daartoe bydra dat die geldigheid daarvan ingeperk word, of soos Miller, Linn en Gronlund (2009:108) dit stel: lae betroubaarheidswaardes dui aan dat daar 'n lae graad van geldigheid sal wees, terwyl hoë betroubaarheidswaardes nie noodwendig 'n hoë graad van geldigheid sal verseker nie. Betroubaarheid verskaf dus bloot die konsistensie wat geldigheid moontlik maak. In terme van die raamwerk wat hier gebruik word, berus tegniese effek of krag (wat 'n verwysing binne die tegniese na die fisies-energetiese is) op tegniese konsistensie ('n kinematiese analogie).

Aangesien daar reeds verskeie kere na die konsistente hoë betroubaarheidswaardes van TAG en TALL in hierdie studie verwys is (sien byvoorbeeld Tabel 5.11), sowel as na die relatiewe homogeniteit (verwys bespreking by paragraaf 5.2.2.3.2 en

Figure 5.7a en 5.7b) van die konstruk, word dit nie weer hier aangetoon nie, maar die volgende opmerking blyk wel geregverdig te wees, naamlik dat die konsistente hoë alpha-waardes en relatiewe homogeniteit van TAG en TALL 'n basis verskaf vir 'n argument oor die konstrukgeldigheid van hierdie toetse. In die ondersoek van Van der Walt en Steyn (2007) na die validering van TAG word die kwessie van homogeniteit ook bespreek en die bevinding is dat die interne korrelasies van die subtoetse met mekaar en ook met die toets as geheel so goed is as wat dit kan wees en dat daar kongruensie is met hierdie bevindings en uitsprake wat in die literatuur daarvoor gemaak word (Van der Walt en Steyn, 2007:147-148; 151).

7.3.1.3.2 Sub-hipotese 3b: TAG beskik oor intratoetsgeldigheid

Intratoetsgeldigheid verwys na die interne struktuur van die toets, met ander woorde dit wat inherent tot die toets self lê. Hier gaan dit, benewens aspekte wat met inhoudsgeldigheid te make het, ook oor die verband tussen die items en / of die subtoetse. Die verband word in die vorm van 'n betroubaarheidskoëffisiënt uitgedruk, waar die interne konsistensie vir elke afdeling van die toets afsonderlik, en vir die toets as geheel bepaal word. Vir doeleindes hiervan is die 2007-, 2008- en 2009-resultate van TAG vir die Universiteit van Pretoria en die Noordwes Universiteit geneem en die alpha-waardes vir elk van die afdelings van die toetse bepaal, soos in Tabel 7.5 hieronder uiteengesit. Daar is besluit om nie die Universiteit Stellenbosch se resultate ook in ag te neem nie, omdat hierdie universiteit nie dieselfde toets as bogenoemde universiteite afgeleë het nie, en ook die feit dat 'n meer heterogene populasie die betroubaarheidswaardes vir hierdie instelling kunsmatig hoër gemaak het. Let daarop dat die alpha-waarde 'n funksie is van die aantal items waarop dit gebaseer is, en dit is daarom te verwagte dat die betroubaarheidskoëffisiënt van sekere afdelings laer as die internasionaal aanvaarde norm van 0.80 (cf. Weir, 2005:29) sal wees. Die betroubaarheidswaarde vir die toets as geheel is egter deurlopend hoër as die norm. Dit word bevestig deur Van der Walt en Steyn (2007:143) se bevindinge hieroor. Dit is ook interessant om daarop te let dat dieselfde patroon vir beide die Universiteit van Pretoria en die Noordwes Universiteit vorm. Op grond van die data in Tabel 7.5, kan daar afgelei

word dat TAG ten dele oor konstrugeldigheid beskik wat intratoetsaspekte betref (die beskrywende statistiek onderliggend aan hierdie analise word in Tabel 7.4 aangebied).

Tabel 7.4: Beskrywende statistiek onderliggend aan die resultate vir Tabel 7.5

	Universiteit van Pretoria			Noordwes Universiteit		
	2007	2008	2009	2007	2008	2009
Mean	56.66	55.83	59.44	51.14	48.62	54.83
Standard Error	0.303	0.295	0.274	0.299	0.287	0.267
Median	56	56	59	50	48	55
Mode	47	51	59	41	51	57
Standard Deviation	15.41	14.23	14.52	15.56	14.97	14.50
Sample Variance	237.56	202.60	210.95	242.14	223.97	210.17
Kurtosis	-0.55	-0.17	-0.31	-0.41	-0.30	-0.33
Skewness	0.04	-0.06	-0.03	0.21	0.15	0.12
Range	88	97	87	89	91	86
Minimum	0	0	0	0	0	0
Maximum	100	100	100	100	100	100
Count	2582	2333	2811	2707	2711	2951
Coefficient Alpha	0.83	0.81	0.84	0.83	0.82	0.83
Std. Error of Measurement	6.37	6.23	5.87	6.41	6.27	5.96
Number of items	63	62	67	63	62	67

Tabel 7.5: Betroubaarheidskoëffisiënte van die TAG-subtoetse oor ses administrasies heen by die Universiteit van Pretoria en die Noordwes Universiteit

	Universiteit van Pretoria			Noordwes Universiteit		
	2007	2008	2009	2007	2008	2009
Afdeling 1	$\alpha = 0.83$	$\alpha = 0.62$	$\alpha = 0.64$	$\alpha = 0.84$	$\alpha = 0.55$	$\alpha = 0.60$
Skommelteks	(5 vrae)	(5 vrae)	(5 vrae)	(5 vrae)	(5 vrae)	(5 vrae)
Afdeling 2	$\alpha = 0.63$	$\alpha = 0.49$	$\alpha = 0.65$	$\alpha = 0.63$	$\alpha = 0.45$	$\alpha = 0.66$
Interpretasie van grafiese/visuele	(7 vrae)	(6 vrae)	(7 vrae)	(7 vrae)	(6 vrae)	(7 vrae)
Afdeling 3	$\alpha = 0.66$	$\alpha = 0.72$	$\alpha = 0.75$	$\alpha = 0.67$	$\alpha = 0.72$	$\alpha = 0.72$
Tipes taalgebruik	(5 vrae)	(5 vrae)	(5 vrae)	(5 vrae)	(5 vrae)	(5 vrae)
Afdeling 4	$\alpha = 0.63$	$\alpha = 0.60$	$\alpha = 0.64$	$\alpha = 0.64$	$\alpha = 0.60$	$\alpha = 0.60$
Teksbegrip	(22 vrae)	(22 vrae)	(22 vrae)	(22 vrae)	(22 vrae)	(22 vrae)
Afdeling 5	$\alpha = 0.62$	$\alpha = 0.53$	$\alpha = 0.57$	$\alpha = 0.61$	$\alpha = 0.58$	$\alpha = 0.59$
Akademie se woordeskate	(9 vrae)	(10 vrae)	(12 vrae)	(9 vrae)	(10 vrae)	(12 vrae)
Afdeling 6	$\alpha = 0.89$	$\alpha = 0.92$	$\alpha = 0.86$	$\alpha = 0.89$	$\alpha = 0.95$	$\alpha = 0.89$
Grammatika en teksverhoudings	(15 vrae)	(14 vrae)	(16 vrae)	(15 vrae)	(14 vrae)	(16 vrae)
Toets as geheel	$\alpha = 0.83$	$\alpha = 0.81$	$\alpha = 0.84$	$\alpha = 0.83$	$\alpha = 0.82$	$\alpha = 0.83$
	(63 vrae)	(62 vrae)	(67 vrae)	(63 vrae)	(62 vrae)	(67 vrae)

'n Verdere tegniek waarmee geëksperimenteer kan word en wat volgens Bachman (1990:258) voorgelê kan word as ondersteunend tot die konstrueerbaarheid van 'n toets, is om korrelasies tussen die onderskeie afdelings van 'n toets, en met die toets as geheel te trek. Die rede hiervoor is dat elke afdeling veronderstel is om 'n ander aspek van die toets te meet en ietwat laer korrelasies sou hier dus verkieslik wees. Net so behoort hoë korrelasies tussen elke afdeling en die toets as geheel weer aan die orde te wees, aangesien die finale toetsresultaat 'n meer algemene aanduiding van bemeestering van die domein is. Let daarop dat hoewel akademiese geletterdheid as een konsep geag kan word en daar om hierdie rede teoreties gesproke hoër korrelasies tussen die onderskeie afdelings te verwagte sal wees, is die konstrueer van TAG en TALL egter, en soos reeds geargumenteer, besonder ryk en daarom behoort korrelasies tussen die onderskeie afdelings verkieslik nie te hoog te wees nie. Tabelle 7.6 en 7.7 hieronder is onderskeidelik die beskrywende statistiek en die korrelasiematriks vir die 2007-data van die Universiteit Stellenbosch, die Noordwes Universiteit en die Universiteit van Pretoria saam gegroepeer.

Tabel 7.6: Beskrywende statistiek vir die bepaling van inter-subtoetskorrelasies vir die 2007-inname aan die Universiteit Stellenbosch, die Noordwes Universiteit en die Universiteit van Pretoria

Veranderlike	Totaal	Afd 1	Afd 2	Afd 3	Afd 4	Afd 5	Afd 6
N	9374	9374	9374	9374	9374	9374	9374
# of items	63	5	7	22	9	5	15
Mean	52.16	2.33	4.36	23.93	10.85	2.23	8.45
Std Dev	16.43	1.63	1.81	9.27	4.15	1.27	4.40
SEM	6.37	0.72	1.11	5.20	2.55	0.86	1.52
Average P-value	52.16	46.66	62.32	47.86	60.29	44.59	56.35
Coefficient Alpha	0.85	0.81	0.63	0.69	0.62	0.54	0.88
GLB	0.88	0.89	0.68	0.72	0.66	0.63	0.92
Asymptotic GLB	0.88	0.89	0.67	0.72	0.66	0.63	0.92

Tabel 7.7: Korrelasies tussen verskillende subtoetse vir die 2007-inname aan die Universiteit Stellenbosch, die Noordwes Universiteit en die Universiteit van Pretoria

	Totaal	Afd 1	Afd 2	Afd 3	Afd 4	Afd 5	Afd 6
Afdeling 1							
Skommelteks	0.43	1					
Afdeling 2							
Interpretasie van grafiese/visuele	0.47	0.19	1				
Afdeling 3							
Teksbegrip	0.89	0.32	0.33	1			
Afdeling 4							
Akademiese woordeskat	0.75	0.26	0.29	0.54	1		
Afdeling 5							
Tipes taalgebruik	0.47	0.15	0.17	0.31	0.35	1	
Afdeling 6							
Grammatika en teksverhoudings	0.66	0.20	0.26	0.37	0.41	0.38	1

Uit die bostaande kan daar gesien word dat twee van die ses korrelasies tussen die totaal van die toets en die afdelings die aanvaarde norm van 0.70 oorskry. Daar sou egter geargumenteer kon word dat 'n derde (Afdeling 6) naby genoeg aan die norm is. Verder kan gesien word dat agt van die vyftien korrelasies wanneer afdelings met mekaar vergelyk word tussen die aanvaarde norme van 0.3 en 0.5 val. Hierdie analise is in lyn met dit wat Van der Walt en Steyn (2007:147-148) gedoen het en

bevestig derhalwe laasgenoemde ondersoek. Op grond van die volgende stelling van dieselfde outeurs sou daar geoordeel kon word dat bostaande 'n aanduiding van konstruk geldigheid kan wees: "The internal correlations are probably as good as can be expected" (Van der Walt & Steyn (2007:148).

7.3.1.3.3 Sub-hipotese 3c: TAG beskik oor intertoetsgeldigheid

Intertoetsgeldigheid het te make met die beoordeling van korrelasies van 'n aantal toetse met mekaar. Dit is volgens Smit (1991:65) daarop gemik om gemeenskaplikhede tussen toetse te identifiseer en dan te bepaal of die onderskeie toetse dieselfde konstruk meet, al dan nie. Miller, Linn en Gronlund (2009:85) wys daarop dat

[t]he scores of any particular assessment can be expected to correlate substantially with the scores of other measures of the same or a similar construct. By the same token, lower correlations would be expected to be obtained with measures of a different ability or trait.

Vir doeleindes hiervan word drie metodes gevolg, naamlik die metode van kongruente geldigheid, die metode van faktorontleding van interkorrelasies verkry van 'n aantal toetse, en die metode van konvergente en diskriminante geldigheid, soos in Smit (1991:66) uiteengesit.

Die eerste metode, naamlik die van **kongruente geldigheid**, het te make met die ooreenstemmende kwaliteite, of anders gestel die gelykvormige aard, van twee toetse. Ten einde hierdie gelykvormigheid tussen twee toetse te ondersoek, moes daar dus met toetse gewerk word wat op min of meer dieselfde konstruk gebaseer is en wat deur dieselfde populasie afgelê is. Daar is derhalwe besluit om data van die Universiteit Stellenbosch vir hierdie ondersoek te gebruik, omdat die geleentheid hom by hierdie instelling voorgedoen het om verskillende toetse, wat almal in 'n mindere of 'n meerdere mate op dieselfde konstruk gebaseer is, én wat deur dieselfde populasie afgelê is met mekaar te vergelyk.

In 2007 het alle aansoekers as deel van dié Universiteit se toegangsbeleid 'n Afrikaanse en 'n Engelse toegangstoets (onderskeidelik TGTAfr en TGTEng genoem) geskryf, en wel in die jaar wat hul eerste studiejaar voorafgaan. Studente

het die keuse gehad in watter taal hul die toets wou aflê en slegs een van die twee toetse was verpligtend. TAG en TALL is egter verpligtend en addisioneel tot bogenoemde toetse tydens die verwelkomingsweek, die week voor klasse 'n aanvang neem, geskryf. Studente wat aansoek vir studie in die gesondheidswetenskappe gedoen het, het binne konteks van 'n nasionale ooreenkoms tussen alle Fakulteite Gesondheidswetenskappe egter nog 'n battery van toetse afgelê wat as die Gesondheidswetenskappe Plasingstoets (GWPT) bekendstaan. Hierdie battery het ook 'n taaltoets / 'n akademiese geletterdheidstoets as deel daarvan en daar is vroeër in hierdie studie daarna verwys, naamlik die PTEEP-toets van die Universiteit van Kaapstad.

Dieselfde poel studente wat vir hierdie analise gebruik is, is dus diegene wat almal TAG en TALL aan die begin van die akademiese jaar afgelê het – dus hoofsaaklik ingeskrewe studente ('n weglaatbare klein aantal studente wat hierdie toetse geskryf het, het uiteindelik nie geregistreer nie). Die statistiek vir die TGT Afr en TGT Eng, én die PTEEP verwys slegs na daardie studente wat wel TAG en TALL afgelê het, sowel as óf die TGT Afr óf die TGT Eng, en/óf die PTEEP. Slegs die ooreenstemmende dele van die populasies is dus telkens met mekaar vergelyk.

Tabel 7.8 hieronder is 'n uiteensetting van die beskrywende statistiek vir die bepaling van kongruente geldigheid vir die 2007-kohort aan die Universiteit Stellenbosch. Dit word opgevolg deur Tabel 7.9 waarin die korrelasies tussen die onderskeie datastelle voorgehou word. In hierdie tabel word telkens die Pearson-korrelasiekoëffisiënt gegee (aangedui deur r). Daarna word die variansieverklaring gegee (aangedui deur r^2). Dit word opgevolg deur die p-waarde ('n aanduiding van die waarskynlikheid dat die korrelasie wat uitgewys word deur die statistiese toets bloot toevallig is; soms ook 'n tipe 1-fout genoem) en ten slotte die aantal waarnemings wat vir die korrelasie gebruik is (aangedui deur N).

Tabel 7.8: Beskrywende statistiek vir die bepaling van kongruente geldigheid vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TAG	4160	51.78	18.80	0	100
TALL	4164	64.98	16.78	0	100
PTEEP	516	64.62	9.63	23.73	85.31
TGTAfr	2398	55.80	12.27	9.09	87.27
TGTEng	2108	53.26	13.79	12	88

Tabel 7.9: Korrelasies tussen verskillende datastelle vir die 2007-inname aan die Universiteit Stellenbosch vir die bepaling van kongruente geldigheid

		TAG	TALL	PTEEP	TGTAfr	TGTEng
TAG	r	1	0.39	0.37	0.67	0.27
	r ²		15.3%	13.9%	44.9%	7.4%
	p		<.0001	<.0001	<.0001	<.0001
	N		4159	516	2395	2103
TALL	r	0.39	1	0.69	0.38	0.60
	r ²	15.3%		46.9%	14.5%	36.0%
	p	<.0001		<.0001	<.0001	<.0001
	N	4159		516	2398	2108
PTEEP	r	0.37	0.69	1	0.54	0.50
	r ²	13.9%	46.9%		28.6%	25.0%
	p	<.0001	<.0001		0.0005	0.0022
	N	516	516		38	35
TGTAfr	r	0.67	0.38	0.54	1	0.32
	r ²	44.9%	14.5%	28.6%		10.4%
	p	<.0001	<.0001	0.0005		<.0001
	N	2395	2398	38		1037
TGTEng	r	0.27	0.60	0.50	0.32	1
	r ²	7.4%	36.0%	25.0%	10.4%	
	p	<.0001	<.0001	0.0022	<.0001	
	N	2103	2108	35	1037	

Die hoë en beduidende korrelasie tussen TAG en TGTAfr in Tabel 7.9 is te verwagte, omdat die TGTAfr, hoewel in verkorte vorm, op die konstruk van TAG gebaseer is; verder was daar ook oorvleueling tussen die toetsontwikkelaarspanne van die twee toetse. Dieselfde argument geld vir TALL en die TGTEng. Die hoë en beduidende korrelasie tussen TALL en PTEEP is ook te verwagte, aangesien die konstruk van TAG en TALL 'n verfyning is van die konstruk waarop PTEEP

gebaseer is (verwys Hoofstukke 3 en 4). Net so kan gesien word dat die variansieverklaring van TALL teenoor TAG laag is, wat beteken dat relatief min van die een toets deur die ander verklaar kan word. Dieselfde geld vir TGTAfr en TGTEng, en vir TAG met PTEEP en TGTEng onderskeidelik. Dit is ook te verwagte, omdat die tale waarin die twee toetse afgelê is, verskil, hoewel dit steeds in 'n mindere of 'n meerdere mate op dieselfde konstrunkte gebaseer is. Verder, en nog meer waarskynlik, is die feit dat sommige kandidate 'n bepaalde taal beter as 'n ander taal beheers en taalbeheersing speel dus ook 'n rol. Hierdie sou moontlik ook as argument kon dien vir die bestaanreg van toetse in minstens twee tale – 'n toets in slegs een taal sou nie noodwendig op die fynere nuanses van die twee tale kon inspeel nie en taal kan tot konstruk-irrelevante variansie bydra (sien die bespreking hierbo); dieselfde geld vir die kwessie van taalbeheersing. Ten slotte is dit interessant om daarop te let dat hoewel daar 'n laer korrelasie tussen byvoorbeeld TAG en TALL is as tussen TAG en TGTAfr, die korrelasie tussen eersgenoemde twee toetse steeds beduidend is. Dieselfde geld vir TAG met PTEEP en TAG met TGTEng. Daar blyk dus 'n verband te wees tussen die onderskeie toetse. Ten einde bogenoemde data te verifieer, is soortgelyke ondersoeke vir die 2008- en 2009-kohorte aan die Universiteit Stellenbosch gedoen.

Tabel 7.10 hieronder is 'n uiteensetting van die beskrywende statistiek vir die bepaling van kongruente geldigheid vir die 2008-kohort aan die Universiteit Stellenbosch, en Tabel 7.11 is 'n aanduiding van die korrelasies tussen die verskillende datastelle vir 2008.

Tabel 7.10: Beskrywende statistiek vir die bepaling van kongruente geldigheid vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TAG	4269	50.52	19.17	0	100
TALL	4259	66.23	16.65	7.00	100
PTEEP	479	66.70	8.94	27.43	85.71

Tabel 7.11: Korrelasies tussen verskillende datastelle vir die 2008-inname aan die Universiteit Stellenbosch vir die bepaling van kongruente geldigheid

		TAG	TALL	PTEEP
TAG	r	1	0.42	0.44
	r ²		17.7%	19.3%
	p		<.0001	<.0001
	N		4259	479
TALL	r	0.42	1	0.67
	r ²	17.7%		44.3%
	p	<.0001		<.0001
	N	4259		478
PTEEP	r	0.44	0.67	1
	r ²	19.3%	44.3%	
	p	<.0001	<.0001	
	N	479	478	

Vanaf 2008 word die TGTAfr en TGTEng aan die Universiteit Stellenbosch nie meer geskryf nie, omdat daar geoordeel is dat TAG en TALL meer omvattende toetse is en soos reeds in Tabel 7.11 hierbo gesien kan word dat die TGTAfr en TGTEng respektiewelik beduidend hoog korreleer met TAG en TALL. Voornemende studente aan die Fakulteit Gesondheidswetenskappe het egter steeds die PTEEP geskryf. Die korrelasies in Tabel 7.11 hierbo is weereens gedoen op slegs daardie studente wat eerstens TAG en TALL geskryf het, en tweedens slegs op daardie studente wat PTEEP, en TAG en TALL geskryf het. Dieselfde tendens as vir die 2007-kohort kom hier voor en word derhalwe nie weer bespreek nie, maar dit kan wel genoem word dat die relatief lae korrelasie tussen TAG en TALL inderdaad deur die verskil in taal verklaar kan word – sien Tabel 7.13 hieronder.

Vir die 2009-kohort is weereens 'n aantal toetse met mekaar vergelyk: onderskeidelik TAG, TALL, PTEEP en die akademiese geletterdheidsafdeling van die onlangs ontwikkelde 'National Benchmark Test' (NBT) wat ten doel het om aanvullende inligting tot die Nasionale Senior Sertifikaat te verskaf (verwys weer Hoofstuk 1 vir 'n uiteensetting van die problematiek met die gebruik van slegs die Nasionale Senior Sertifikaat as toelatingsmeganisme). Die NBT word bemark as 'n plasingstoets om vir universiteite 'n aanduiding te gee van of 'n student op byvoorbeeld hoofstroomonderwys geplaas behoort te word, óf op 'n verlengde graadprogram, óf dalk selfs nie geskik is vir hoër onderwysstudie nie. In praktyk

sal hierdie battery van toetse egter waarskynlik as toelatingsmeganismes gebruik word. Daar is tans nog nie uitsluitel oor die implementering of gebruik van hierdie toetse aan alle hoër onderwysinstellings nie. Boonop moet die betroubaarheid en geldigheid van hierdie toetse ook nog eers bepaal en openbaar gemaak word alvorens daar kritiek belangrike besluite oor persone en hul toekoms geneem kan word, veral as die toetse onder andere as toelatingsmeganismes gebruik gaan word. Loodsprojekte met die NBT is egter reeds landswyd onderweg en die data wat vir hierdie studie en onderstaande analise gebruik is, is gebaseer op 'n seleksie voornemende studente van die Universiteit Stellenbosch. Let ook daarop dat die Universiteit Stellenbosch se nuwe toelatingsbeleid in 2009 in werking getree het, wat beteken dat alle aansoekers verplig is om onder andere TAG en TALL af te lê, terwyl dit in die verlede slegs reeds toegelate studente was wat TAG en TALL afgelê het – verwys weereens Hoofstuk 1 van hierdie studie vir meer inligting hieroor. Die aantal studente wat sedert 2009 die toetse aflê is dus aansienlik meer as voorheen, maar soos reeds genoem het die grootte van die populasie nie 'n noemenswaardige invloed op die korrelasiekoëffisiënte gehad nie.

Tabel 7.12 hieronder is 'n uiteensetting van die beskrywende statistiek vir die bepaling van kongruente geldigheid vir die 2009-kohort aan die Universiteit Stellenbosch, en Tabel 7.13 is 'n aanduiding van die korrelasies tussen die verskillende datastelle vir 2009.

Tabel 7.12: Beskrywende statistiek vir die bepaling van kongruente geldigheid vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TAG	6106	47.32	17.28	0	100
TALL	6108	61.28	16.15	0	100
PTEEP	487	63.51	10.74	25	87.5
NBT	1065	68.13	9.18	30	89

Tabel 7.13: Korrelasies tussen verskillende datastelle vir die 2009-inname aan die Universiteit Stellenbosch vir die bepaling van kongruente geldigheid

		TAG	TALL	PTEEP	NBT
TAG	r	1	0.37	0.40	0.25
	r ²		13.3%	16.1%	6.2%
	p		<.0001	<.0001	<.0001
	N		6104	487	1064
TALL	r	0.37	1	0.68	0.70
	r ²	13.3%		46.3%	47.9%
	p	<.0001		<.0001	<.0001
	N	6104		487	1064
PTEEP	r	0.40	0.68	1	0.93
	r ²	16.1%	46.3%		87.2%
	p	<.0001	<.0001		0.0064
	N	487	487		6
NBT	r	0.25	0.70	0.93	1
	r ²	6.2%	47.9%	87.2%	
	p	<.0001	<.0001	0.0064	
	N	1064	1064	6	

Tabel 7.13 bevestig dit wat in Tabele 7.9 en 7.11 voorgekom het. Hier is dit weer die geval dat toetse wat in dieselfde taal afgelê word beduidend hoog met mekaar korreleer: TALL, PTEEP en die NBT onderskeidelik met mekaar (let daarop dat die NBT slegs in Engels beskikbaar was). Verder is die variansieverklaring van al drie hierdie toetse ook hoog, wat kan impliseer dat dit nie nodig is vir 'n student om al drie hierdie toetse af te lê nie. Die uitsonderlik hoë korrelasie en variansieverklaring tussen PTEEP en NBT is ook te verwagte, omdat laasgenoemde in wese maar 'n verwerking van eersgenoemde is. Dit is weereens interessant om daarop te let dat selfs TAG ('n Afrikaanse toets), hoewel laag, steeds beduidend met TALL, PTEEP en NBT korreleer.

Gebaseer op bogenoemde analyses en die gepaardgaande bespreking sou die afleiding dus gemaak kon word dat TAG (en derhalwe TALL ook) oor kongruente geldigheid beskik.

* * * * *

Die metode van **faktorontleding** van interkorrelasies verkry vir 'n aantal toetse behels die doen en interpretasie van faktoranalises vir verskeie toetse. Dit word normaalweg gedoen om drie doelwitte te bereik: eerstens, om te bepaal wat die getal gemeenskaplike faktore is wat tot interitemkorrelasies sal bydra; ten tweede, om te bepaal watter items tot prestasie in 'n toets bydra; en in die derde plek om 'n bepaling te kan maak van die persentasie variansie wat deur spesifieke faktore bepaal word. Na aanleiding hiervan kan 'n toets sodanig aangepas of ontwerp word dat dit 'n relatief suiwer meting van 'n spesifieke teoretiese konstruk gee. Faktoranalises (waar elke item as 'n faktor beskou word wat tot die konstruk van die toets bydra) lê meer of minder homogeniteit in die konstruk bloot, wat weer 'n invloed op konstrukgeldigheid het. Vir doeleindes van hierdie ondersoek sal ter wille van variasie op die bostaande analises die TAG 2007-, 2008- en 2009-data vir beide die Universiteit van Pretoria en die Noordwes Universiteit ondersoek word. Die data van die Universiteit Stellenbosch word dus hierdie keer by die ondersoek uitgesluit, onder andere ook omdat hierdie instelling nie op 'n jaar tot jaar grondslag dieselfde toetse as die ander instellings skryf nie.

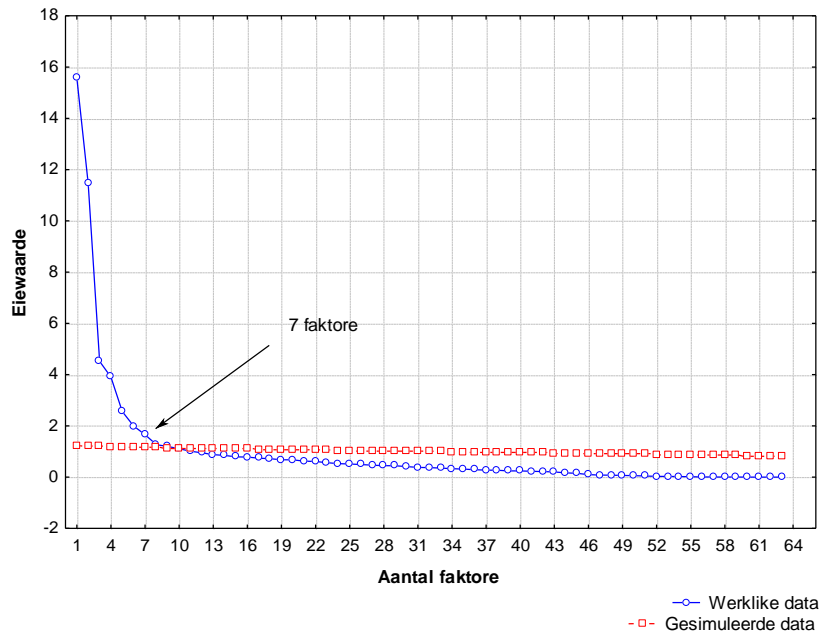
Vir hierdie ondersoek is, soos reeds genoem, die data van die Universiteit van Pretoria en die Noordwes Universiteit telkens vir die onderskeie jare van inname saam gegroepeer en daar is van die tegniek van 'Exploratory factor analysis', oftewel verkennende faktorontleding, gebruik gemaak. Hierdie tegniek het volgens De Coster (1998:2), en soos hierbo reeds gesuggereer, ten doel om onderliggende faktore (soms ook temas of subkonstrukte genoem) bloot te lê wat in 'n toets tot die variansie bydra en dus die response kon beïnvloed het. Die konstrukgeldigheid van elke afdeling van die toets word aan die hand van hierdie metode bepaal wat data sogenaamd reduceer en die kernkomponente of faktore daarvan uitwys. Items wat onder 'n faktor kluster moet met mekaar korreleer. Indien daar 'n kluster van items onder 'n spesifieke faktor is, is dit 'n aanduiding daarvan dat 'n enkele subkonstruk as't ware gemeet word. 'n Aanvaarbare resultaat sal dus uitwys dat die onderskeie items per afdeling van 'n toets kluster onder 'n enkele faktor, wat 'n aanduiding daarvan is dat dit spesifieke aspekte van die oorkoepelende konstruk meet.

Normaalweg word daardie afdelings waar daar nie klusters is nie beoordeel as 'n teken daarvan dat dit nie tot konstruegeldigheid van die toets bydra nie, tensy daar 'n argument daarvoor gevoer kan word. Van der Walt en Steyn (2007:149) verduidelik dit as “[t]o be construct valid, as few as possible factors that explain the maximum percentage of variance are required, with communalities as high as possible.”

Die hoeveelheid faktore sal uiteindelik afhang van waar die afsnypunt gestel word, en dit is gelykstaande aan die aantal eiewaardes ('eigenvalues') van die korrelasiematriks wat 'n waarde groter as 1 het. Hierdie word egter nie as 'n besonder goeie metode beskou nie en daar is gevolglik vir hierdie studie 'n sogenaamde Knakpunttoets ('Scree test') gedoen waar die genoemde eiewaardes van elke item in 'n dalende orde op 'n grafiek geplot word. Om die knakpunttoets beter te kwalifiseer, is gebruik gemaak van die metode van parallele analise om die optimale aantal faktore te bepaal (O'Connor, 2000).

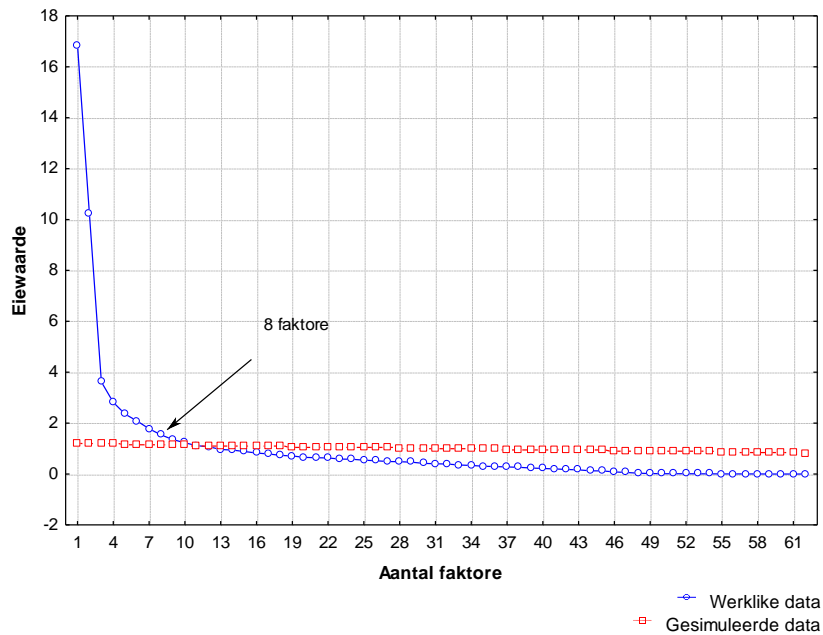
Hierdie faktore word dan geroteer deur die sogenaamde Varimaks-rotasie en/of die oblimaks-faktorrotasie (beide is in hierdie geval uitgevoer) wat uiteindelik 'n matriks sal oplewer wat die verwantskap tussen elke item op die toets en elke faktor sal aandui – hierdie matrikse is egter uitsonderlik groot en sal nie hier aangebied word nie, maar die essensie daarvan sal wel telkens direk na die knakpuntgrafiek aangebied word. Daar word normaalweg vir 'n faktorlading van bokant 0.5 gesoek en by alle verwysings hieronder van items wat kluster, is die faktorladings inderdaad 0.5 en hoër. Let daarop dat die data bestaan uit 0/1 response en daarom is die faktoranalise gebaseer op die sogenaamde tetrachoriese korrelasiematriks.

Figure 7.1, 7.2, en 7.3 toon die onderskeie knakpuntgrafieke aan vir die jare 2007, 2008, en 2009 vir die Universiteit van Pretoria en die Noordwes Universiteit gesamentlik.



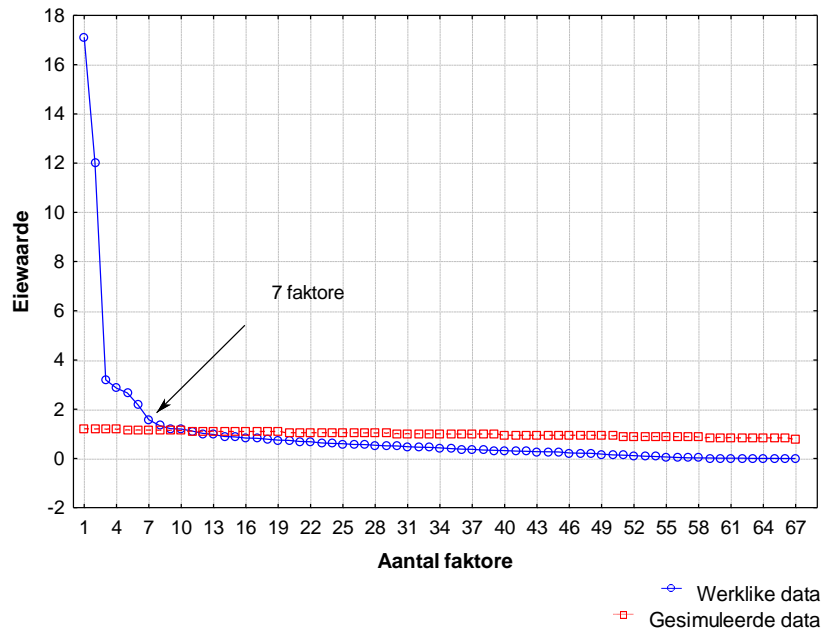
Figuur 7.1: Knappuntgrafiek vir 2007 vir die Universiteit van Pretoria en die Noordwes Universiteit gesamentlik

Uit hierdie eerste knoppuntgrafiek is dit duidelik dat daar sewe faktore / temas / subkonstrukte is wat bydra tot die variansie in response op TAG 2007. Items 1 – 5 (Skommelteks) het almal by Faktor 3 gelê, Items 6 – 12 (Interpretasie van visuele inligting) het almal by Faktor 4 gelê, Items 44 – 48 (Tekstipe) het almal by Faktor 5 gelê, en Items 51 tot 63 (Teksredigering) het almal by Faktor 1 gelê. Die res van die items het oor alle faktore heen versprei gelê. Die essensie hiervan is dat die sewe faktore eerstens 46.6% van die variansie verklaar. Ten tweede is dit 'n aanduiding daarvan dat sekere items by mekaar kluster – 'n saak is dus uit te maak vir die homogeniteit van die onderskeie afdelings van TAG (bostaande items meet telkens 'n spesifieke deel van die konstruk van TAG). Dit maak ook sin dat die meerderheid van daardie items wat nie kluster nie en wat tot 53.4% van die variansie bydra in die Teksbegrip-afdeling lê (wat die konstruk van TAG as geheel meet). Daar is wel sewe items in hierdie afdeling wat kluster, maar onder vier faktore. Wat die afdeling Akademiese Woordeskat betref, kan dit genoem word dat vier van die nege items onder een faktor gekluster het. Oorkoepelend gesien het hierdie twee afdelings dus 'n mate van heterogeniteit veroorsaak, maar dit is geregverdig en is reeds elders (cf. afdeling 5.2.2.3.2) bespreek.



Figuur 7.2: Knappuntgrafiek vir 2008 vir die Universiteit van Pretoria en die Noordwes Universiteit gesamentlik

Vir die 2008-inname is daar agt faktore wat bydra tot die variansie in response op TAG. 'n Soortgelyke patroon as vir die 2007-data ontwikkel hier, naamlik dat die skeiding volgens faktore grotendeels ooreenstem met die verskillende afdelings van die toets. In hierdie toets kluster Items 1 – 5 (Skommelteks) weereens saam, maar by Faktor 3 en Faktor 8. Vier van die ses items by die afdeling wat oor interpretasie van visuele inligting handel, kluster by Faktor 2. Items 12 – 16 (Tekstipe) kluster weereens almal saam, maar hierdie keer by Faktor 4, en Items 49 – 62 (Teksredigering) kluster saam by Faktor 1. Daar is by die interpretasie van hierdie datastel dus ook 'n argument te voer vir die homogeniteit van die onderskeie afdelings van TAG (bostaande items meet telkens 'n spesifieke deel van die konstruk van TAG). Daardie items wat nie noodwendig kluster nie en verspreid oor die verskillende faktore lê, kom weereens meestal by die Teksbegrip- en die Woordeskat-afdelings voor en dieselfde argument as hierbo geld dus hier ook, naamlik dat hierdie twee afdelings tot 'n mate van verdedigbare en verklaarbare heterogeniteit bygedra het.



Figuur 7.3: Knakpuntgrafiek vir 2009 vir die Universiteit van Pretoria en die Noordwes Universiteit gesamentlik

Vir die 2009-inname is daar, soos by die 2007-inname, sewe faktore wat bydra tot die variansie in response op TAG, en die patroon word herhaal, naamlik dat daardie items wat kluster, hoë faktorladings het (bokant 0.50) en meestal by een faktor lê. Daardie items met hoë faktorladings wat nie kluster nie, is weereens meestal by die Teksbegrip-afdeling.

Uit bostaande figure en die gepaardgaande bespreking kan afgelei word dat TAG se verskillende weergawes oor die jare heen tot intertoetsgeldigheid beskik. Die Woordeskat-afdeling is egter kommerwekkend, omdat 'n kluster van items daar te verwagte is en dit by nie een van die analyses voldoende gekorrespondeer het nie. Dit is wel so dat hierdie nie 'n maklike subtoets is om te ontwerp nie, en dat daar hoë uitvalsifers is van items by loodstoetse.

Bostaande resultate is oor die algemeen veral verblydend vanuit 'n konstrugeldigheidsperspektief, waarvan faktoranalises deel is. Benewens die feit dat daar uit die bostaande afgelei kan word dat TAG oor drie jaar, by twee instellings konsekwent dieselfde tipe resultaat opgelewer het, is dit ook 'n verifikasie van die werk van Van der Slik en Weideman (2005), waaroor daar reeds

in Hoofstuk 5 gerapporteer is, sowel as die werk van Van der Walt en Steyn (2007:148-149): eersgenoemde ondersoek het van ander statistiese tegniek as hierbo gebruik gemaak, terwyl laasgenoemde van dieselfde tegniek gebruik gemaak het.

* * * * *

Ten einde **konvergente en diskriminante geldigheid** te bepaal (die derde metode), word 'n spesifieke toets eerstens met ander toetse wat 'n meting van dieselfde konstruk is, gekorreleer. Dit word egter ook met ander toetse wat nie 'n meting van dieselfde konstruk is nie, gekorreleer om te bepaal of dit diskrimineer. 'n Hoë korrelasie dra dus by tot konvergente geldigheid en 'n lae korrelasie tot diskriminante geldigheid, en beide hierdie tipes geldigheid het weer 'n invloed op konstrukgeldigheid.

Hoewel konvergente geldigheid in wese dieselfde as kongruente geldigheid is en die analyses wat hierbo vir laasgenoemde gedoen is daarvoor aangebied sou kon word, is daar besluit om ter wille van verifikasie die saak tog verder te ondersoek, maar nog meer gefokus. Vir doeleindes van 'n ondersoek na die konvergente geldigheid van spesifiek TAG word die 2007-, 2008- en 2009-datastelle van die Universiteit Stellenbosch eerstens met TALL vir dieselfde tydperk vergelyk. Dit word verder weereens met die TGTAfr en die TGTEng vergelyk (waar van toepassing), en vir die 2009-kohort word dit ook met die akademiese geletterdheidsafdeling en waar van toepassing die kwantitatiewe geletterdheidsafdeling (soms ook die gesyferdheidsafdeling genoem) van die NBT vergelyk. Die redes hiervoor is presies dieselfde as hierbo bespreek vir kongruente geldigheid.

Net so word die data van TAG 2007, 2008 en 2009 wat by die Universiteit Stellenbosch afgelê is, ook ter wille van diskriminante geldigheid met dele uit die res van die toegangstoetsbattery wat by hierdie universiteit gebruik word, gekorreleer. Hierdie besluit is gebaseer daarop dat die studierigting waarvoor die student sal inskryf sal bepaal watter kombinasie van toetse in die battery geskryf moet word. So skryf 'n student wat byvoorbeeld vir ingenieurswese wil registreer

TAG en TALL, sowel as Fisiese Wetenskappe en Wiskunde. 'n Student wat weer van voorneme is om in die lettere en sosiale wetenskappe te registreer skryf byvoorbeeld TAG en TALL, en Denkvaardighede. Let daarop dat die battery uit ses toetse bestaan: TAG en TALL (wat beide in terme van die taalbeleid van die Universiteit van Stellenbosch, met die oog op plasing, afgelê móét word), Syfervaardighede, Wiskunde, Fisiese Wetenskappe, en Denkvaardighede.

Wat die analyses vir hierdie deel van die geldigheidsargument wel anders maak as hierbo by kongruente geldigheid, dog steeds omvattend en terselfdertyd hanteerbaar, is dat dit, soos reeds genoem, gefokus is op drie breë kategorieë, naamlik die Ekonomiese en Bestuurswetenskappe, die *Humaniora* en die *Scientiae* (alle studierigtings aan die Universiteit Stellenbosch kan onder hierdie drie kategorieë resorteer). TAG sal dus telkens met die toepaslike dele van die res van die Universiteit Stellenbosch se toegangstoetsbattery gekorreleer word, en waar van toepassing ook met die relevante afdelings van die NBT.

Tabel 7.14 hieronder is 'n uiteensetting van die beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2007-kohort vir die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch, en Tabel 7.15 is 'n aanduiding van die korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir 2007. Let daarop dat studente wat van voorneme is om in die Ekonomiese en Bestuurswetenskappe te studeer 'n kombinasie van beide TAG en TALL, Wiskunde en Syfervaardighede skryf.

Tabel 7.14: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2007-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TGTAfr	704	57.85	10.79	21.82	85.45
TGTEng	612	55.24	13.05	12	88
TGTSyfer	1036	49.04	18.27	6.67	100
TGTWisk	1205	45.22	17.74	10	96.67
TAG	1373	51.53	17.88	0	96
TALL	1373	65.23	15.86	19	99

Tabel 7.15: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2007-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch

		TGTAfr	TGTEng	TGTSyfer	TGTWisk	TAG	TALL
TAG	r	0.59	0.29	0.41	0.39	1	0.31
	r ²	34.9%	8.4%	17.2%	15.0%		9.9%
	p	<.0001	<.0001	<.0001	<.0001		<.0001
	N	704	612	1036	1205		1373

Soos verwag is die hoogste korrelasie van TAG met die TGTAfr. Hoewel die ander korrelasies ietwat laer is, is alles beduidend. Dit is egter interessant, en teen die intuïtiewe verwagting in, om daarop te let dat TAG 'n hoër korrelasie met die TGTSyfer en TGTWisk toon as met TGTEng of TALL. Een verklaring is dat syfervaardigheidstoetse miskien selfs meer taalafhanklik is as wat mens sou vermoed. Daar sou dus geredeneer kon word dat TAG meer as net akademiese geletterdheid meet en dat daar faktore is wat vir konstrukirrelevansie nagegaan moet word. By nadere ondersoek het dit geblyk dat gesyferdheid nie 'n swaar gewig in TAG dra nie, maar dat die wyse waarop vrae in die TGTSyfer geformuleer is, wel 'n sterk talige ondertoon het – die taal van gesyferdheid moet net so goed bemeester word as die vermoë om syferkundige berekeninge te kan doen. Dit val ook vreemd op dat die korrelasie met TGTEng en met TALL die laagste van alles is (sien die verklaring hierbo). Hoewel daar myns insiens hier nie blyke van diskriminante geldigheid (die verskillende toetse meet verskillende tipes vaardighede) na vore tree

nie, sou sommige kon redeneer dat die korrelasie met TGTWisk byvoorbeeld relatief laag is en as dit teenoor 'n relatief hoë korrelasie soos met TGTAfr vergelyk word, dan sou daar vir konvergente geldigheid aan die een kant en diskriminante geldigheid aan die ander kant geargumenteer kon word. In die literatuur is daar na die beste van my wete geen aanduiding van wat as laag genoeg en wat as hoog genoeg geag word vir 'n onderskeid tussen konvergente en diskriminante geldigheid nie. Bachman (1990:263) stel dit wel dat

[c]onvergence ... is the extent to which different measures of the same trait tend to agree, or converge. *Discrimination* is the extent to which measures of different traits, using either the same or different test methods, tend to produce different results. ... convergence would [therefore] be indicated by *high positive* correlations between the different measures of the same traits ... Discrimination, on the other hand would be indicated by *low* or zero correlations between measures of different traits ...

Binne konteks van hierdie aanhaling en van die voorbeelde wat wel in die literatuur oor konvergente en diskriminante geldigheid voorgehou word, kan die afleiding wel gemaak word dat daar blyke van konvergente geldigheid by TAG is, maar dit kan nog nie onomwonde gestel word dat hierdie toets oor diskriminante geldigheid beskik nie.

Vervolgens word dieselfde analises vir die 2008- en 2009-kohorte van voornemende studente van die Ekonomiese en Bestuurswetenskappe gedoen. Tabel 7.16 hieronder is 'n uiteensetting van die beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2008-kohort vir die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch, en Tabel 7.17 is 'n aanduiding van die korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir 2007. Let daarop dat daar besluit is om vanaf 2008 met die TGTAfr en TGTEng weg te doen (sien die verduideliking hierbo) en daarom verskyn dit nie weer hieronder in die tabelle nie.

Tabel 7.16: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2008-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TGTSyfer	1126	46.74	18.75	6.67	100
TGTWisk	1303	43.61	17.62	4.41	96.67
TAG	1405	50.03	17.94	0	95
TALL	1396	66.03	15.28	15	98

Tabel 7.17: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2008-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch

		TGTSyfer	TGTWisk	TAG	TALL
TAG	r	0.46	0.41	1	0.34
	r ²	21.2%	17.1%		11.3%
	p	<.0001	<.0001		<.0001
	N	1126	1303		1396

Dieselfde argument oor konvergente en diskriminante geldigheid as hierbo geld steeds.

Vervolgens word daar na die 2009-kohort vir die Ekonomiese en Bestuurswetenskappe gekyk. Tabel 7.18 is 'n uiteensetting van die beskrywende statistiek en Tabel 7.19 is 'n aanduiding van die korrelasies tussen TAG en die ander toetse wat vir die 2009-inname afgelê is, sowel as die toetse wat vir die loodsprojek van die NBT uitgevoer is. Let daarop dat die NBT uit verskeie afdelings bestaan en vir doeleindes van hierdie analise is daar besluit om van die akademiese geletterdheidsafdeling (AG) gebruik te maak ten einde uitsprake oor konvergente geldigheid te maak, sowel as van die kwantitatiewe geletterdheidsafdeling (KG; ook soms syfervaardighede genoem) ten einde uitsprake oor diskriminante geldigheid te kan maak. Vir hierdie analise word daar ook van beide TAG en TALL gebruik, omdat die NBT se loodsprojek slegs toetse in Engels beskikbaar gehad het.

Tabel 7.18: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2009-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TGTSyfer	1590	55.50	16.94	7.89	100
TGTWisk	1508	43.09	16.31	6.67	97.14
NBT (AG)	1033	68.17	9.21	30	89
NBT (KG)	1033	61.47	14.40	16	94
TAG	1753	48.57	15.80	0	97
TALL	1754	62.40	14.27	0	96

Tabel 7.19: Korrelasies tussen TAG en die ander toetse vir die 2009-inname van die Ekonomiese en Bestuurswetenskappe aan die Universiteit Stellenbosch

		TGTSyfer	TGTWisk	NBT (AG)	NBT (KG)	TAG	TALL
TAG	r	0.33	0.31	0.24	0.28	1	0.23
	r ²	10.9%	9.3%	5.8%	8.0%		5.3%
	P	<.0001	<.0001	<.0001	<.0001		<.0001
	N	1588	1507	1032	1032		1752
TALL	r	0.51	0.41	0.69	0.52	0.23	1
	r ²	25.6%	16.7%	47.5%	26.5%	5.3%	
	P	<.0001	<.0001	<.0001	<.0001	<.0001	
	N	1589	1507	1032	1032	1752	

Uit Tabel 7.19 kan gesien word dat TAG 'n hoër korrelasie met die TGTSyfer en TGTWisk toon as met die NBT(AG) en met TALL, wat interessant is omdat TAG en TALL op dieselfde konstruk gebaseer is en daar ook vlakke van ooreenkoms met die NBT(AG) behoort te wees. Dit is egter belangrik om hier in ag te neem dat TAG en TALL van jaar tot jaar weens sekuriteitsredes wissel en dat die korrelasies dus van jaar tot jaar behoort te verskil, én dat hierdie 'n geselekteerde deel van die spesifieke kohort is. TALL, aan die ander kant, toon hoë korrelasies met NBT(AG) en NBT(KG) wat waarskynlik weer onder andere daaraan toe te skryf is dat al drie hierdie toetse slegs in Engels afgelê is. Dit moet op hierdie punt egter ook genoem word dat TALL hoër met TGTSyfer en TGTWisk gekorreleer het as wat die geval met TAG was (die TGTSyfer en die TGTWisk was in óf Afrikaans óf Engels beskikbaar). Die laer korrelasies by TAG is moontlik daaraan toe te skryf dat 'n deel van die kohort wat TAG afgelê het se taalbeheersing in Afrikaans nie

noodwendig op dieselfde vlak is as daardie studente wat die toetse in Engels moes aflê nie. Dit moet ook nie uit die oog verloor word dat een van die afdelings in TAG en TALL, naamlik interpretasie van visuele inligting, in wese 'n meting van kwantitatiewe geletterdheid is wat aspekte van gesyferdheid kan insluit en dus deels tot die hoër korrelasies met TGTSyfer en TGTWisk aanleiding kon gegee het. 'n Ander verklaring wat reeds hierbo genoem is, is dat die TGTSyfer en die TGTWisk tot 'n groter mate afhanklik is van taal as wat op sigwaarde blyk. Laasgenoemde twee vra eintlik dringend na dieselfde soort van valideringsproses as wat hier vir TAG (en by implikasie TALL) gedoen word.

Op hierdie stadium kan dit wel genoem word dat daar vir die kategorie van die Ekonomiese en Bestuurswetenskappe vir die 2007-, 2008- en die 2009-kohorte aanduidings van konvergente geldigheid is, wat 'n verifikasie is van dit wat hierbo oor kongruente geldigheid aangebied is, maar dat daar nie sonder meer net aangevoer kan word dat TAG oor diskriminante geldigheid beskik nie. Benewens die bespreking hierbo, moet dit egter hier ook in ag geneem word dat die toetse wat hier met mekaar vergelyk word almal op die een of ander manier met mekaar verband hou, omdat die een byvoorbeeld uit die ander voortgevloei het / 'n verfyning van die ander is, die een aspekte van die ander sou kon insluit, en ook die feit dat al hierdie toetse min of meer dieselfde doel het, naamlik om gereedheid vir universiteitstudie te meet. In hierdie opsig sou daar gespekeer kon word dat die relatief lae korrelasies tog 'n aanduiding van diskriminante geldigheid kan wees.

Vervolgens word die 2007-, 2008-, en 2009-kohorte vir die *Humaniora* en die *Scientiae* ondersoek en kortliks bespreek. Tabelle 7.20, 7.22 en 7.24 is 'n uiteensetting van die beskrywende statistiek vir onderskeidelik die 2007-, 2008-, en 2009-kohorte wat vir toelating tot die *Humaniora* (Letter en Sosiale Wetenskappe, Regte, Teologie en Opvoedkunde) aansoek gedoen het. Tabelle 7.21, 7.23 en 7.25 is 'n aanduiding van die korrelasies tussen TAG en die ander toetse wat vir dieselfde kohorte aangeteken is, sowel as die toetse wat vir die loodsprojek van die NBT in 2009 uitgevoer is.

Studeute wat in 2007 vir toelating tot studie aan die Universiteit Stellenbosch vir 'n program in die *Humaniora* aansoek gedoen het, was verplig om beide TAG en TALL, beide TGTAfr en TGTEng en die Denkvvaardighedetoets af te lê. Daar was enkele uitsonderings wat nie Afrikaans as skoolvak gehad het nie en dan slegs TALL, TGTEng en Denkvvaardighede afgelê het. Hierdie getal is egter so klein dat dit nie 'n noemenswaardige effek op die analise gehad het nie en hierdie studeute is by die analise ingesluit.

Tabel 7.20: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2007-inname van die Humaniora aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TGTAfr	1084	53.19	13.59	9.09	85.45
TGTEng	1083	51.45	13.77	14	88
TGTDenk	1106	48.42	14.80	10	90
TAG	1276	50.89	18.24	0	94
TALL	1280	63.25	17.27	5	98

Tabel 7.21: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2007-inname van die Humaniora aan die Universiteit Stellenbosch

		TGTAfr	TGTEng	TGTDenk	TAG	TALL
TAG	r	0.71	0.31	0.35	1	0.39
	r ²	50.0%	9.8%	12.3%		15.0%
	p	<.0001	<.0001	<.0001		<.0001
	N	1081	1079	1102		1276

Die patroon word voortgesit: TAG het die hoogste korrelasie met TGTAfr en laer korrelasies met die ander toetse in die battery, hoewel nie laag genoeg om onomwonde te kan verklaar dat TAG oor diskriminante geldigheid beskik nie. Dit is weereens interessant om daarop te let dat die korrelasie met TALL (wat teoreties hoër behoort te wees) laer is vir studeute in kategorie *Humaniora* vir hierdie spesifieke jaar.

Tabel 7.22: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2008-inname van die Humaniora aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TGTDenk	1160	48.38	15.61	10	100
TAG	1321	48.35	18.91	0	100
TALL	1321	63.81	17.09	14	100

Tabel 7.23: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2008-inname van die Humaniora aan die Universiteit Stellenbosch

	TGTDenk	TAG	TALL
TAG r	0.40	1	0.39
r ²	15.7%		15.1%
p	<.0001		<.0001
N	1160		1321

Dieselfde tendens as hierbo geld ook vir die 2008-kohort. Tydens 2009 is die toelatingsbeleid vir die *Humaniora* verander en studente was verplig om beide TAG en TALL af te lê, sowel as die TGTDenk en die TGTSyfer. Voorts is die NBT(AG) en NBT(KG) ook onder hierdie poel studente afgeneem as deel van die loodsprojek. TALL word weereens ook hier in berekening gebring, omdat die NBT net in Engels beskikbaar was.

Tabel 7.24: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2009-inname van die Humaniora aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TGTSyfer	1879	46.17	16.11	3.33	96.67
TGTDenk	171	39.45	15.94	5.26	78.95
NBT(AG)	29	66.31	8.30	42	79
NBT(KG)	29	50.52	13.18	24	84
TAG	1933	44.74	15.91	0	96
TALL	1933	57.16	16.56	0	99

Tabel 7.25: Korrelasies tussen TAG en die ander toetse vir die 2009-inname van die Humaniora aan die Universiteit Stellenbosch

		TGTSyfer	TGTDenk	NBT (AG)	NBT (KG)	TAG	TALL
TAG	r	0.40	0.39	0.45	0.41	1	0.35
	r ²	16.2%	15.6%	20.2%	17.2%		12.5%
	p	<.0001	<.0001	0.0145	0.0253		<.0001
	N	1879	171	29	29		
TALL	r	0.59	0.63	0.80	0.85	0.35	1
	r ²	35.2%	39.6%	64.7%	72.5%	12.5%	
	p	<.0001	<.0001	<.0001	<.0001	<.0001	
	N	1879	171	29	29	1933	

'n Interpretasie van Tabel 7.25 dui daarop dat TAG minder konvergensie met die ander toetse toon as wat die geval met TALL blyk te wees. Dit is weereens interessant om daarop te let dat die korrelasie tussen TAG en TALL die laagste is, en die hoogste vir beide afdelings van die NBT. Daar is dus beslis steeds sprake van konvergente geldigheid, maar minder vir diskriminante geldigheid.

Die 2007-, 2008-, en 2009-kohorte vir die *Scientiae* word vervolgens hieronder geanaliseer en kortliks bespreek. Tabelle 7.26, 7.28 en 7.30 is 'n uiteensetting van die beskrywende statistiek vir onderskeidelik die 2007-, 2008-, en 2009-kohorte wat vir toelating tot hierdie kategorie (die Natuurwetenskappe, AgriWetenskappe en Ingenieurswese) aansoek gedoen het. Tabelle 7.27, 7.29 en 7.31 is 'n aanduiding van die korrelasies tussen TAG en die ander toetse wat vir dieselfde kohorte aangeteken is. Let daarop dat die loodsprojek vir die NBT nie op hierdie studente uitgevoer is nie, en dat studente wat om toelating tot studie aan die Fakulteit Gesondheidswetenskappe aansoek doen ook van hierdie analise uitgesluit word, omdat hulle benewens TAG en TALL slegs die gesondheidswetenskappeplasingstoets skryf en wat reeds hierbo ondersoek en bespreek is.

Studente wat in 2007 vir toelating tot studie aan die Universiteit Stellenbosch vir 'n program in die *Scientiae* aansoek gedoen het, was verplig om beide TAG en TALL, beide TGTAfr en TGTEng, sowel as die TGTWisk en TGTFisWet (Fisiese

Wetenskappe) af te lê. Let weereens daarop dat die TGTAfr en TGTEng vanaf 2008 wegval, omdat die toelatingsbeleid in hierdie jaar verander het en slegs TAG en TALL as taaltoetse verpligtend was.

Tabel 7.26: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2007-inname van die Scientiae aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TGTAfr	598	58.01	10.40	21.82	87.27
TGTEng	402	55.13	14.37	16	86
TGTWisk	849	49.75	18.33	6.67	96.67
TGTFisWet	783	53.92	18.22	9.09	100
TAG	1214	52.25	19.70	0	93
TALL	1214	65.47	17.17	9	100

Tabel 7.27: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2007-inname van die Scientiae aan die Universiteit Stellenbosch

		TGTAfr	TGTEng	TGTWisk	TGTFisWet	TAG	TALL
TAG	r	0.60	0.34	0.45	0.45	1	0.45
	r ²	36.1%	11.6%	20.4%	19.8%		19.8%
	p	<.0001	<.0001	<.0001	<.0001		<.0001
	N	598	401	848	782		1213

Tabel 7.28: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2008-inname van die Scientiae aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TGTWisk	975	49.10	18.44	6.67	94.12
TGTFisWet	892	51.24	18.11	6.06	96.97
TAG	1234	52.86	19.84	0	98
TALL	1234	67.76	17.51	7	100

Tabel 7.29: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2008-inname van die Scientiae aan die Universiteit Stellenbosch

		TGTWisk	TGTFisWet	TAG	TALL
TAG	r	0.49	0.47	1	0.52
	r ²	24.1%	22.1%		26.8%
	p	<.0001	<.0001		<.0001
	N	975	892		1234

Tabel 7.30: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2009-inname van die Scientiae aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TGTWisk	1475	45.67	16.63	8.57	97.14
TGTFisWet	1420	45.11	17.42	2.86	97.14
TAG	2075	48.28	18.92	0	95
TALL	2076	62.53	16.10	0	100

Tabel 7.31: Korrelasies tussen TAG en die ander toetse in die toegangstoetsbattery vir die 2009-inname van die Scientiae aan die Universiteit Stellenbosch

		TGTWisk	TGTFisWet	TAG	TALL
TAG	r	0.45	0.46	1	0.42
	r ²	20.2%	21.5%		17.9%
	p	<.0001	<.0001		<.0001
	N	1473	1419		2074

Uit bogenoemde tabelle behoort dit te blyk dat presies dieselfde tendens vir die *Scientiae* as vir die *Humaniora* en die Ekonomiese en Bestuurswetenskappe geld.

'n Vermoede het egter bestaan dat sou bostaande data ook per huistaal ondersoek word, omdat daar reeds aanduidings daarvan is dat taal 'n invloed kan hê op die korrelasiekoëffisiënt, dit moontlik verdere interessante data kan oplewer. Daar is gevolglik, en steeds as deel van die verifikasieproses, besluit om die 2009-data vir die Universiteit Stellenbosch as geheel ook per moedertaal (slegs vir diegene met Afrikaans of Engels as moedertaal) te ontleed. Vir doeleindes hiervan is die data van TAG, TALL, PTEEP (weereens slegs van toepassing op studente wat om

toelating vir studie in die gesondheidswetenskappe aansoek doen), en die akademiese geletterdheidsafdeling van die NBT (weereens slegs van toepassing op daardie studente wat dit vir die loodsprojek afgelê het) met mekaar vergelyk. Let daarop dat dit nie ook vir alle kohorte, soos hierbo, of vir ander onderskeidings gedoen word nie, omdat dit in die volgende afdeling wat oor groep-differensiële geldigheid handel, in meer besonderhede aan bod sal kom.

Tabel 7.32 is die beskrywende statistiek vir die moedertaalsprekers van Afrikaans vir die 2009-kohort. Tabel 7.33 is 'n aanduiding van die korrelasies tussen TAG, TALL, PTEEP, en die NBT vir studente met Afrikaans as huistaal. Tabelle 7.34 en 7.35 verteenwoordig dieselfde tipe data as bogenoemde twee tabelle, maar vir studente met Engels as moedertaal.

Tabel 7.32: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2009-inname vir studente met Afrikaans as moedertaal aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
NBT(AG)	255	62.94	10.72	25	84.5
PTEEP	635	66.21	9.81	30	89
TAG	3231	55.78	13.65	0	95
TALL	3231	57.66	15.43	0	100

Tabel 7.33: Korrelasies tussen TAG, TALL, PTEEP en die NBT(AG) vir die 2009-inname vir studente met Afrikaans as moedertaal aan die Universiteit Stellenbosch

		NBT (AG)	PTEEP	TAG	TALL
TAG	r	0.61	0.56	1	0.71
	r ²	37.3%	30.8%		49.7%
	p	<.0001	<.0001		<.0001
	N	634	255		3230
TALL	r	0.70	0.66	0.71	1
	r ²	49.2%	43.1%	49.7%	
	p	<.0001	<.0001	<.0001	
	N	634	255	3230	

Tabel 7.34: Beskrywende statistiek vir die bepaling van konvergente en diskriminante geldigheid vir die 2009-inname vir studente met Engels as moedertaal aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
NBT(AG)	168	66.02	9.84	39.3	87.5
PTEEP	342	71.40	6.79	47	88
TAG	2129	37.67	13.54	0	90
TALL	2131	67.40	14.31	17	99

Tabel 7.35: Korrelasies tussen TAG, TALL, PTEEP en die NBT(AG) vir die 2009-inname vir studente met Engels as moedertaal aan die Universiteit Stellenbosch

		NBT (AG)	PTEEP	TAG	TALL
TAG	r	0.23	0.50	1	0.45
	r ²	5.3%	24.9%		20.1%
	p	<.0001	<.0001		<.0001
	N	342	168		2128
TALL	r	0.51	0.70	0.45	1
	r ²	25.8%	48.8%	20.1%	
	p	<.0001	<.0001	<.0001	
	N	342	168	2128	

Uit die laaste vier tabelle kan afgelei word dat hoewel TAG, TALL, PTEEP en NBT(AG) onderling dieselfde konstruk mag meet, die moedertaal (of minstens vaardigheid in die moedertaal) inderdaad, en soos vermoed, 'n invloed het op die korrelasiekoëffisiënte. By die Afrikaanse studente is daar byvoorbeeld gevind dat hulle by wyse van rangorde tot 'n groot mate dieselfde posisie beklee, hetsy dit TAG of TALL is, terwyl die Engelse studente aansienlik swakker in Afrikaans as in Engels vaar en hul posisie op die rangorde dus daal. Hieruit is dit duidelik dat wanneer die groep gemeng is (alle taalgroepe saamgevoeg) die korrelasiekoëffisiënt laer sal neig, terwyl dit hoër sal neig by meer homogene groepe. Vergelyk byvoorbeeld die r-waarde van 0.705 vir studente met Afrikaans as moedertaal wanneer TAG en TALL met mekaar gekorreleer word, met die r-waarde van 0.365 (Tabel 7.13) wanneer alle studente by die analise ingesluit word.

Tabelle 7.16 tot 7.35 hierbo is 'n uiteensetting van die analyses wat vir verskillende kohorte aan die Universiteit Stellenbosch gedoen is. Die kohorte behels verskillende jare van inname, sowel as verskillende kategorieë studente: studente in die Ekonomiese en Bestuurswetenskappe, studente in die *Humaniora* en studente in die *Scientiae*, sowel as studente met onderskeidelik Afrikaans en Engels as moedertale. Hoewel hierdie dus 'n besonder gefokusde ondersoek was, met 'n besondere hoeveelheid detail ten einde verifikasie vir kongruente en diskriminante geldigheid te soek, het nie 'n enkele datastel enige van die ander weerspreek nie. Daar is uit die aard van die saak wel verskille, maar einde ten laaste kan daar met redelike sekerheid geargumenteer word dat TAG wel oor konvergente geldigheid beskik (wat 'n verifikasie is van kongruente geldigheid, hierbo bespreek). Daar is verder vae blyke van diskriminante geldigheid, omdat korrelasies tussen toetse wat veronderstel is om nie dieselfde bloudruk te hê nie wel laer neig, maar dit kan nie onomwonde as 'n konklusie gestel word nie. Voorts was daar ook hoër korrelasies tussen toetse wat in teorie veronderstel is om van mekaar te verskil, omdat die konstruksie daarvan verskil, wat die kwessie van diskriminante geldigheid by TAG as 'n non-argument sou kon verklaar.

Soos egter reeds telkens hierbo aangevoer, maar ter wille van die argument weer hier uiteengesit sal word, kan dit aan verskeie faktore toegeskryf word, soos die feit dat hierdie toetse almal vir presies dieselfde doelgroep ontwikkel is en min of meer dieselfde doel nastreef, naamlik om gereedheid vir universiteitstudie te meet; ook die feit dat homogene groepe geneig is tot hoër korrelasies en heterogene groepe tot laer korrelasies. Voorts is die een ook soms 'n funksie van die ander, omdat dit elemente kan meet wat ooreenstem. Verdere navorsing sou egter 'n toets met 'n heel ander doel, byvoorbeeld 'n persoonlikheidsvraelys, kon ontleed, en wel oor tyd heen, op dieselfde wyse as wat hierbo gedoen is; dit behoort hipoteties gesproke tot diskriminante geldigheid by te dra. Die beskikbare data is egter vir doeleindes van hierdie studie ontleed en geïnterpreteer en hoewel 'n argument vir diskriminante geldigheid dus nie moontlik is nie, sou daar terugskouend wel afgelei kon word dat

elemente van ooreenstemming in die verskillende toetse tot lae diskriminante geldigheid aanleiding gegee het.

7.3.1.3.4 Sub-hipotese 3d: TAG beskik oor groep-differensiële geldigheid

Groep-differensiële geldigheid het te make met hoe verskillende groepe op die toets presteer. Met verskillende groepe word geslagte, moedertaalgroepe, rasse, ensovoorts bedoel. Groepe kan ook in terme van die top presteerders (die boonste kwartiel) en die swakste presteerders (die onderste kwartiel) onderskei word. Sou daar differensiërende aspekte tussen verskillende groepe wees, kan dit gebruik word om die konstruk te regverdig. Hierdie tipe geldigheid hou ook verband met voorspellingsgeldigheid wat later bespreek sal word, omdat daar verskillende voorspellings vir verskillende groepe gemaak sal kan word.

Vir doeleindes van hierdie valideringsargument sal die volgende egter ondersoek word: daar sal eerstens bepaal word hoe daar aan die Universiteit Stellenbosch tussen geslagte gedifferensieer kan word; in die tweede plek sal die onderskeie rasgroeperinge wat aan hierdie universiteit studeer met mekaar vergelyk word (moedertaalgroeperinge sou ook hier teen mekaar afgespeel kon word, maar dit is nie hier gedoen nie, omdat dit ook by die afdeling oor konkurrente geldigheid gebruik is); en derdens sal die boonste en onderste kwartiel van verskillende kohorte met mekaar vergelyk word, om te bepaal of daar werklike verskille tussen sogenaamd sterker en sogenaamd swakker studente is - let daarop dat die gemiddelde Graad 12-prestasie hier as die kriterium gebruik is waarvolgens die boonste en die onderste kwartiel gedefinieer word, omdat dit steeds as die enkele beste voorspeller van akademiese sukses in Suid-Afrika geag word. Al drie hierdie ondersoeke is weereens vir die 2007, 2008 en 2009 kohorte gedoen.

Op hierdie punt blyk dit egter noodsaaklik te wees om eers weer die statistiese tegnieke te verduidelik wat vir onderstaande analyses gebruik is ten einde verwarring te vermy, omdat dieselfde toetse nie deurlopend soos by die analyses hierbo (Pearson se korrelasiekoëffisiënt) aangelê kon word nie; dit is ook in Hoofstuk 5 verduidelik. Die statistiese toetse wat vir onderstaande deel van die

studie aangewend is, is sogenaamde t-toetse (dieselfde toetse as wat in Hoofstuk 5 vir die bepaling van betroubaarheid aangewend is – let weereens op die verband tussen betroubaarheid en geldigheid), en ‘Analysis of Variance’ (ANOVA).

T-toetse word gebruik om te bepaal of die gemiddeld tussen twee veranderlikes (soms ook groepe genoem) werklik verskil en of die verskil bloot aan toeval toegeskryf kan word. Daar is as’t ware drie t-toetse wat hiervoor gebruik kan word. Die eerste is vir sogenaamde gepaarde waarnemings, byvoorbeeld waar dieselfde studente gebruik word, maar die toets verskil (tipies by ’n voor- en natoetsituasie). Die tweede is ’n toets wat aangelê word wanneer die groepe verskil, maar die toets wat gebruik word, word konstant gehou. Vir laasgenoemde is daar weer twee opsies: een vir wanneer die twee groepe gelyke variansies het, en die ander vir wanneer die variansies verskil. Daar moet dus eers ’n verdere toets (byvoorbeeld Bartlet of Levene se toets) aangelê word om te bepaal of die variansies verskil, ten einde ’n besluit te kan neem oor watter opsie om te gebruik.

ANOVA’s toets of daar beduidende verskille tussen die gemiddeld van meer as twee groepe is en dit kan slegs suksesvol uitgevoer word as die variansie dieselfde is vir al die groepe en die verskillende groepe se waardes normaal versprei is. Wanneer een faktor, maar verskillende groepe ter sprake kom, word die sogenaamde ‘oneway ANOVA’ gebruik om te bepaal of daar verskille rondom die gemiddeld van ’n bepaalde faktor is. Indien daar beduidende verskille tussen groepe gevind word, word agterna veelvoudige vergelykings gedoen om te bepaal watter groepe dan wel van mekaar verskil. Die toets vir homoskedastisiteit moet ook aangewend word, omdat die data vir t-toetse en vir ANOVA’s normaal versprei moet wees; dit word die nul-hipotesetoets genoem.

Die nul-hipotese vir die ANOVA is:

$$H_0: \text{gem}_1 = \text{gem}_2 = \dots = \text{gem}_c$$

waar c die aantal groepe verteenwoordig.

Indien die p -waarde > 0.05 , dan beteken dit dat daar nie ’n beduidende verskil tussen gemiddeld 1 en 2 (of c) is nie. Indien p -waarde < 0.05 , dan is daar ’n

beduidende verskil in gemiddeld. Levene se toets kan gebruik word om te bepaal of die variansie verskil, al dan nie en die nul-hipotese lyk dan as volg:

$$H_0: \text{var}_1 = \text{var}_2 = \dots = \text{var}_c$$

In die onderstaande ANOVA's het die aanname van normaliteit vir al die verskillende groeperinge gegeld, maar nie die aanname van homogene variansies nie. Streng gesproke behoort ander statistiese tegnieke gebruik te word indien hierdie aanname nie bevredig word nie. Aangesien die steekproewe wat vir die analyses gebruik is, egter redelik groot is, is die effek van die afwykings in die aannames beperk. Verder, veral gesien in die lig van die feit dat die analyses slegs 'n klein deel van hierdie proefskrif uitmaak, is dit nie geregverdig om binne konteks van die doel en bestek van hierdie studie in meer besonderhede hierop in te gaan nie. Uitsprake wat hieronder gemaak word, is dus geen sterk uitsprake nie, maar bloot 'n rapportering van tendense.

Tabelle 7.36 tot 7.38 is 'n aanduiding van die analyses wat vir 'n groepdifferensiële ondersoek na verskille in geslag vir die 2007-kohort aan die Universiteit Stellenbosch gedoen is.

Tabel 7.36: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per geslag vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	N	TAG Mean	TAG Std Dev	TAG Std Err	TAG 95.00%	TAG +95.00%
Totaal	4141	51.81	18.77	0.29	51.24	52.38
Manlik	1907	52.03	18.76	0.43	51.19	52.87
Vroulik	2234	51.62	18.79	0.40	50.84	52.40

Tabel 7.37: Levene se toets vir homogeniteit van variansies ten opsigte van geslag vir die 2007-inname aan die Universiteit Stellenbosch

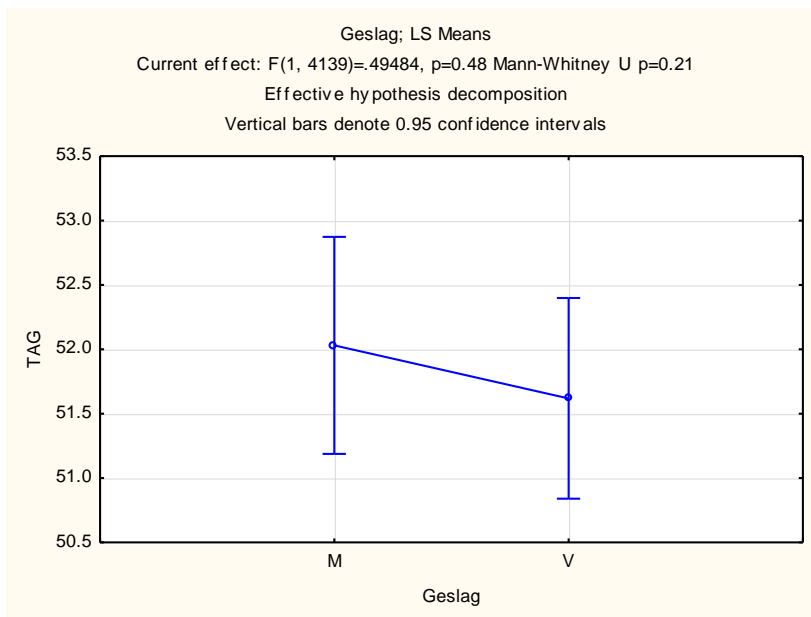
Veranderlike	MS - Effect	MS - Error	F	p
TAG	154.29	116.07	1.329	0.249

Hieruit kan gesien word dat die variansie nie beduidend verskil nie en 'n ANOVA kan gedoen word – let daarop dat daar wel getoets is vir normaliteit, maar die resultate word nie hier aangetoon nie.

Tabel 7.38: ANOVA vir verskil in gemiddelde TAG-prestasie per geslag vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	SS	Degr. Of Freedom	MS	F	p
Geslag	175	1	175	0.49	0.482

$p > 0.05$ en die gemiddelde prestasie op TAG tussen manlike en vroulike studente aan die Universiteit vir die 2007-inname is statisties gesproke dieselfde. Dieselfde data word grafies in Figuur 7.4 uitgebeeld:



Figuur 7.4: Grafiese voorstelling van verskil in vir die 2007-inname aan die Universiteit Stellenbosch

Tabelle 7.39 tot 7.41 is 'n aanduiding van die analises wat vir 'n groepdifferensiële ondersoek na verskille in geslag vir die 2008-kohort aan die Universiteit Stellenbosch gedoen is.

Tabel 7.39: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per geslag vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	N	TAG Mean	TAG Std Dev	TAG Std Err	TAG 95.00%	TAG +95.00%
Totaal	4269	50.52	19.17	0.29	49.95	51.10
Manlik	2092	50.12	19.32	0.42	49.29	50.95
Vroulik	2177	50.91	19.02	0.40	50.11	51.71

Tabel 7.40: Levene se toets vir homogeniteit van variansies ten opsigte van geslag vir die 2008-inname aan die Universiteit Stellenbosch

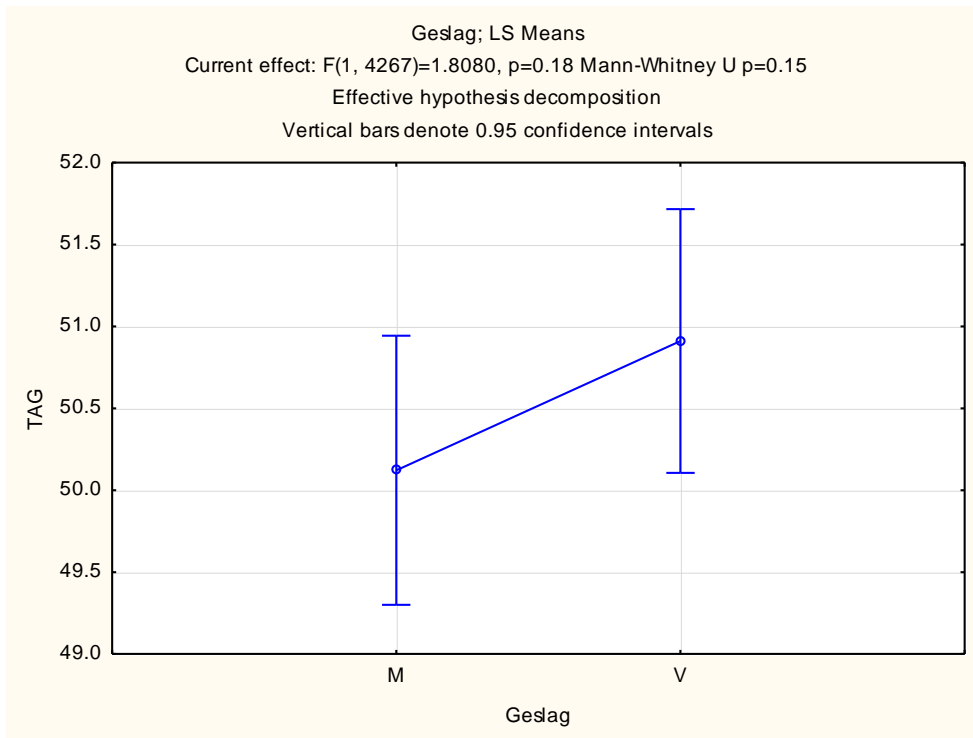
Veranderlike	MS – Effect	MS - Error	F	p
TAG	51.65550	118.3637	0.436413	0.508894

Hieruit kan gesien word dat die variansie nie beduidend verskil nie en 'n ANOVA kan gedoen word - let weerens daarop dat daar wel getoets is vir normaliteit, maar die resultate word nie hier aangetoon nie.

Tabel 7.41: ANOVA vir verskil in gemiddelde TAG-prestasie per geslag vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	SS	Degr. Of Freedom	MS	F	p
Geslag	664	1	664	1.81	0.178818

$p > 0.05$ en die gemiddelde prestasie op TAG tussen manlike en vroulike studente aan die Universiteit vir die 2008-inname is statisties gesproke dieselfde. Dieselfde data word grafies in Figuur 7.5 uitgebeeld:



Figuur 7.5: Grafiese voorstelling van verskil in vir die 2008-inname aan die Universiteit Stellenbosch

Tabelle 7.42 tot 7.44 is 'n aanduiding van die analyses wat vir 'n groepdifferensiële ondersoek na verskille in geslag vir die 2009-kohort aan die Universiteit Stellenbosch gedoen is.

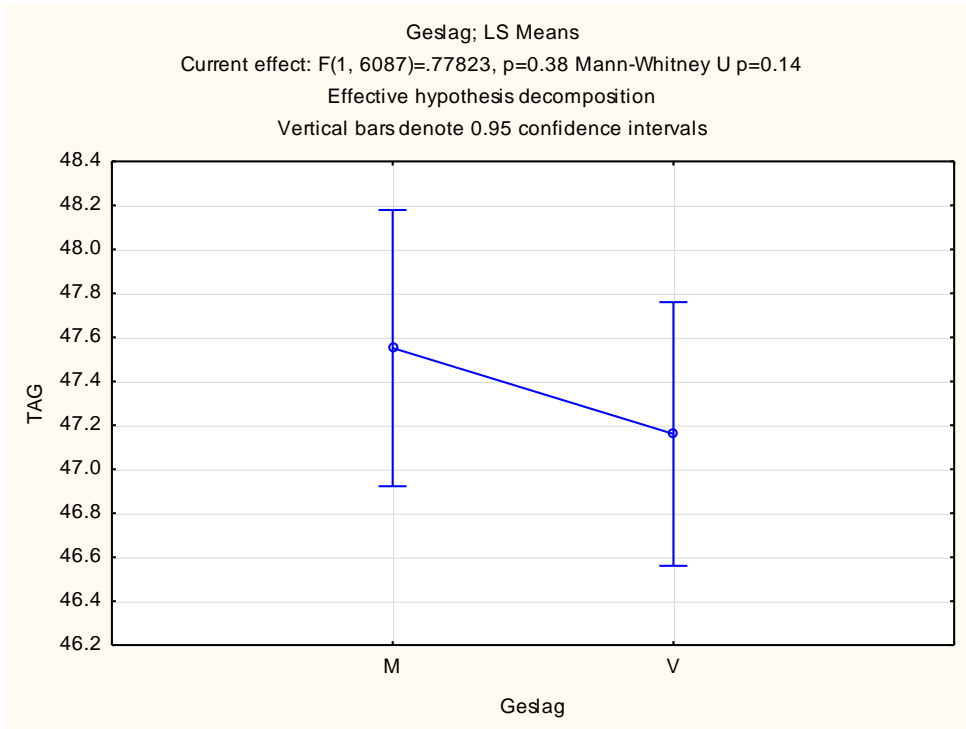
Tabel 7.42: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per geslag vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	N	TAG Mean	TAG Std Dev	TAG Std Err	TAG 95.00%	TAG +95.00%
Totaal	6089	47.35	17.26	0.22	46.91	47.78
Manlik	2903	47.55	17.86	0.33	46.90	48.20
Vroulik	3186	47.16	16.70	0.30	46.58	47.74

Tabel 7.43: Levene se toets vir homogeniteit van variansies ten opsigte van geslag vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	MS – Effect	MS - Error	F	p
TAG	1708.14	109.88	15.55	<0.0001

Hieruit kan gesien word dat die variansie wel beduidend verskil, en omdat hier slegs twee groepe ter sprake is kan die t-toets vir ongelyke variansie uitgevoer word om te bepaal of daar statisties 'n verskil tussen die geslagte vir hierdie kohort is. Figuur 7.6 is 'n grafiese voorstelling hiervan.



Figuur 7.6: Grafiese voorstelling van verskil in vir die 2009-inname aan die Universiteit Stellenbosch

Tabel 7.44: T-toets vir verskil in gemiddelde TAG-prestasie per geslag vir die 2009-inname aan die Universiteit Stellenbosch

	Variances	DF	t Value	Pr > t
TAG	Unequal	5935	0.88	0.379

$P > 0.05$: daar is dus geen beduidende verskil in prestasie tussen geslagte nie.

Uit bostaande tabelle is dit duidelik dat daar geen beduidende verskille tussen geslagte is nie en dit sou dus nie as argument vir die konstrugeldigheid van TAG aangewend kon word nie. Hoewel dit dus nie tot konstrugeldigheid bydra nie, is data soos hierdie wel nuttig wanneer daar oor die betroubaarheid van TAG

geargumenteer word, omdat die toets stabiel bly – dit is reeds in die vorige hoofstuk aangespreek. Hierdie inligting is ’n bevestiging van die werk wat Van der Slik (2009) hieroor gedoen het.

Tabelle 7.45 tot 7.47 is ’n aanduiding van die analyses wat vir ’n ondersoek na groepdifferensiële verskille in ras vir die 2007-kohorte aan die Universiteit Stellenbosch gedoen is. Dieselfde vertrekpunte as hierbo is vir hierdie analyses gebruik. Let daarop dat daar vir hierdie analyses slegs van die rasgroeperinge bruin, swart en wit gebruik gemaak is, omdat die Indiër-groep besonder klein is en daar geen sinvolle vergelyking met dit getref sou kon word nie.

Tabel 7.45: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	N	TAG Mean	TAG Std Dev	TAG Std Err	TAG 95.00%	TAG +95.00%
Totaal	4092	52.03	18.75	0.29	51.45	52.60
Bruin	667	46.00	14.09	0.55	44.93	47.07
Wit	3262	54.63	18.36	0.32	54.00	55.26
Swart	163	24.71	15.34	1.20	22.33	27.08

Tabel 7.46: Levene se toets vir homogeniteit van variansies ten opsigte van ras vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	MS – Effect	MS - Error	F	p
TAG	154.29	116.07	1.329	0.249

Hieruit kan gesien word dat die variansie nie beduidend verskil nie.

Vervolgens is die Bonferroni-toets gedoen om vas te stel tussen watter groepe die verskil voorkom en Tabel 7.47 is 'n aanduiding daarvan.

Tabel 7.47: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	Bruin – 46.00	Wit – 54.63	Swart – 24.71
Bruin		0.00	0.00
Wit	0.00		0.00
Swart	0.00	0.00	

Hieruit blyk dit dat daar wel statisties gesproke ook 'n verskil in prestasie per ras is: wit studente vaar beter as bruin studente, wat weer beter vaar as swart studente.

Tabelle 7.48 tot 7.50 is 'n aanduiding van die analyses wat vir 'n ondersoek na groepdifferensiële verskille in ras vir die 2008-kohorte aan die Universiteit Stellenbosch gedoen is. Dieselfde vertrekpunte as hierbo is vir hierdie analyses gebruik. Let daarop dat daar vir hierdie analyses weereens slegs van die rasgroeperinge bruin, swart en wit gebruik gemaak is, omdat die Indiër-groep besonder klein is en daar geen sinvolle vergelyking met dit getref sou kon word nie.

Tabel 7.48: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	N	TAG Mean	TAG Std Dev	TAG Std Err	TAG 95.00%	TAG +95.00%
Totaal	4230	40.24	19.17	0.29	50.07	51.22
Bruin	750	44.09	13.77	0.65	42.82	45.37
Wit	3285	53.79	18.75	0.31	53.18	54.41
Swart	195	22.83	14.91	1.27	20.33	25.33

Tabel 7.49: Levene se toets vir homogeniteit van variansies ten opsigte van ras vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	MS – Effect	MS - Error	F	p
TAG	7085.2	101.2	70.01	0.00

Hieruit kan gesien word dat die variansie wel beduidend verskil en die volgende tendense het geblyk:

Vervolgens is die Bonferroni-toets gedoen om vas te stel tussen watter groepe die verskil voorkom en Tabel 7.50 is 'n aanduiding daarvan.

Tabel 7.50: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	Bruin – 44.09	Wit – 53.79	Swart – 22.83
Bruin		0.00	0.00
Wit	0.00		0.00
Swart	0.00	0.00	

Hieruit blyk dit dat dieselfde tendens as vir 2007 ook hier geld, hoewel dit nie as 'n feit gestel kan word dat daar wel statisties beduidende verskille tussen die rasgroeperinge is nie.

Tabelle 7.51 tot 7.53 is 'n aanduiding van die analyses wat vir 'n ondersoek na groepdifferensiële verskille in ras vir die 2009-kohorte aan die Universiteit Stellenbosch gedoen is. Dieselfde vertrekpunte as hierbo is vir hierdie analyses gebruik. Let daarop dat daar vir hierdie analyses weereens slegs van die rasgroeperinge bruin, swart en wit gebruik gemaak is, omdat die Indiër-groep besonder klein is en daar geen sinvolle vergelyking met dit getref sou kon word nie.

Tabel 7.51: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	N	TAG Mean	TAG Std Dev	TAG Std Err	TAG 95.00%	TAG +95.00%
Totaal	5978	47.57	17.27	0.22	47.13	48.00
Bruin	1024	43.41	12.25	0.48	42.47	44.36
Wit	4556	50.88	16.45	0.23	50.43	51.33
Swart	398	20.28	15.94	0.77	18.76	21.79

Tabel 7.52: Levene se toets vir homogeniteit van variansies ten opsigte van ras vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	MS – Effect	MS - Error	F	p
TAG	4825.9	81.55	59.17	0.00

Hieruit kan gesien word dat die variansie wel beduidend verskil en die volgende tendense het geblyk:

Vervolgens is die Bonferroni-toets gedoen om vas te stel tussen watter groepe die verskil voorkom en Tabel 7.53 is 'n aanduiding daarvan.

Tabel 7.53: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	Bruin – 43.41	Wit – 50.88	Swart – 20.28
Bruin		0.00	0.00
Wit	0.00		0.00
Swart	0.00	0.00	

Hieruit blyk dit dat dieselfde tendens as vir 2007 en 2008 ook hier geld, hoewel dit steeds nie as 'n feit gestel kan word dat daar wel statisties beduidende verskille tussen die rasgroepe is nie.

Eerste blyke van groepdifferensiële geldigheid tree dus hier na vore. Die waarskuwing wat Messick hierbo in die bespreking oor konstruk-irrelevante

variansie oor die hantering van groepsverskille rig, moet egter in ag geneem word wanneer daar uitsprake oor verskille tussen individue en groepe gemaak word. So byvoorbeeld kan moedertaal of die vaardigheid in die taal wat gemeet word 'n invloed op prestasie op TAG hê – toetse om te bepaal wat die bepalende faktor is, is egter nie hier gedoen nie, omdat die kwessie van groepdifferensiële geldigheid nie die fokus van die studie is nie en maar slegs 'n deel van die groter validerings argument uitmaak, wat weer slegs 'n deel van die volle studie is.

'n Laaste poging ten einde groepdifferensiële geldigheid te bepaal, is om die populasie in twee groepe te verdeel, naamlik die boonste 25% en die onderste 25%. Die gemiddelde Graad 12 prestasie is as kriterium gebruik. Let daarop dat slegs die Nasionale Senior Sertifikaat se uitslae vir die 2009-kohort gebruik is, hoewel daar enkele studente van hierdie inname was wat nog die Senior Sertifikaat voltooi het, maar eers in 2009 tot universiteitstudie toetree het. Die afsny punte vir die bepaling van die boonste en onderste Graad 12-prestasie sien as volg daar uit:

2007	Onderste kwartiel = 68%
	Boonste kwartiel = 85.3%
2008	Onderste kwartiel = 67%
	Boonste kwartiel = 85%
2009	Onderste kwartiel = 66%
	Boonste kwartiel = 78.3%

Die onderste kwartiel is inderdaad so hoog soos hierbo aangedui, omdat die voederarea van die Universiteit Stellenbosch oor die algemeen uit bevoorregte gemeenskappe bestaan. Daar is wel 'n aantal studente wat uit benadeelde omgewings afkomstig is en by hierdie universiteit registreer, maar dit is duidelik dat hierdie universiteit steeds 'n voorkeurinstelling is vir wit studente. Dit blyk ook uit die verteenwoordigende aard van die inskrywingsgetalle waarvan die tabelle hierbo

'n aanduiding is. Tabelle 7.54 tot 7.62 hieronder is 'n aanduiding van die analises wat vir 'n groepdifferensiële ondersoek na verskille in prestasiegroepe vir die 2007-, 2008-, en 2009-kohorte aan die Universiteit Stellenbosch gedoen is.

Tabel 7.54: Beskrywende statistiek vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev
Onderste 25%	1036	65.49	16.21
Boonste 25%	1020	41.34	14.72

Tabel 7.55: Levene se toets vir homogeniteit van variansies ten opsigte van Graad 12-groep vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	MS – Effect	MS - Error	F	p
TAG	946.1	94.73	9.98	0.002

Die verskil in variansie tussen die boonste en die onderste 25% vir die 2007-kohort was beduidend ($p=0.0020$) en die t-toets vir ongelyke variansie is gevolglik vir onderstaande analise gebruik.

Tabel 7.56: T-toets vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2007-inname aan die Universiteit Stellenbosch

	Variiances	DF	t Value	Pr > t
TAG	Unequal	2041	33.92	<.0001

$P<0.001$: daar is 'n beduidende verskil in prestasie tussen die boonste en die onderste kwartiel vir hierdie kohort.

Tabel 7.57: Beskrywende statistiek vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev
Onderste 25%	1080	65.569	16.11
Boonste 25%	1078	39.893	15.94

Tabel 7.58: Levene se toets vir homogeniteit van variansies ten opsigte van Graad 12-groep vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	MS – Effect	MS - Error	F	p
TAG	151.54	94.28	1.607	0.205

Die verskil in variansie tussen die boonste en die onderste 25% vir die 2008-kohort was nie beduidend nie ($p=0.725$) en die t-toets vir gelyke variansie is gevolglik vir onderstaande analise gebruik.

Tabel 7.59: T-toets vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2008-inname aan die Universiteit Stellenbosch

	Variances	DF	t Value	Pr > t
TAG	Equal	2156	37.21	<.0001

$P < 0.001$: daar is 'n beduidende verskil in prestasie tussen die boonste en die onderste kwartiel vir hierdie kohort.

Tabel 7.60: Beskrywende statistiek vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev
Onderste 25%	1208	59.60	15.58
Boonste 25%	1204	39.67	13.51

Tabel 7.61: Levene se toets vir homogeniteit van variansies ten opsigte van Graad 12-groep vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	MS – Effect	MS - Error	F	p
TAG	2142.6	81.97	26.14	<0.0001

Die verskil in variansie tussen die boonste en die onderste 25% vir die 2009-kohort was beduidend ($p < 0.0001$) en die t-toets vir ongelyke variansie is gevolglik vir onderstaande analise gebruik.

Tabel 7.62: T-toets vir verskil in gemiddelde TAG-prestasie volgens Graad 12-groep vir die 2009-inname aan die Universiteit Stellenbosch

	Variances	DF	t Value	Pr > t
TAG	Unequal	2365	33.57	<.0001

$P < 0.001$: daar is 'n beduidende verskil in prestasie tussen die boonste en die onderste kwartiel vir hierdie kohort.

Bostaande analise het aangetoon dat daar vir drie kohorte blyke van groepdifferensiële geldigheid is, aangesien daar beduidende verskille tussen die boonste 25% en die onderste 25% van die populasie is.

Oorkoepelend sou daar dus geargumenteer kon word dat TAG oor groepdifferensiële geldigheid beskik, behalwe waar dit by geslag kom. Die vermoede bestaan ook dat bostaande veralgemeenbaar sal wees na TALL en ook na die Universiteit van Pretoria en die Noordwes Universiteit. Die rede hiervoor is dat dit artikuleer met die ondersoek van Van der Walt en Steyn (2007:144-145) wat aanvoer dat sogenaamd swakkeer kandidate inderdaad laer punte op TAG behaal, en sogenaamd sterker kandidate hoër punte – TAG diskrimineer dus geregverdiglik tussen die boonste en die onderste kwartiel en beskik dus na alle waarskynlikheid oor groepdifferensiële geldigheid.

7.3.1.3.5 Sub-hipotese 3e: TAG beskik oor domeinspesifieke geldigheid

Die domein wat TAG meet, akademiese geletterdheid, is in besonderhede in Hoofstuk 3, en ten dele ook in Hoofstuk 4 van hierdie studie aangebied. Dit word dus nie weer hier herhaal nie. Die uitspraak kan egter gemaak word dat die domein, en derhalwe die konstruk, deeglik gedefinieer is en die toets daarom wat hierdie aspek van konstrukgeldigheid betref voldoen aan die eise daarvan.

Ten slotte kan dit gestel word dat dit uit bostaande bespreking blyk dat TAG (en by implikasie TALL) oor interne geldigheid beskik. Hierdie uitspraak moet egter genuanseerd geles word, omdat daar velerlei veranderlikes is wat 'n invloed op die analises kon gehad het. 'n Ander navorser sou byvoorbeeld ander tegnieke kon aanwend of van ander datastelle gebruikmaak en moontlik by ander resultate en interpretasies kon uitkom. Daar is egter hierbo gepoog om op omvattende wyse die beskikbare data vir TAG (en waar van toepassing TALL) op sistematiese wyse te ontleed. Ten einde dit so oorwoë as moontlik te hanteer is verskillende kohorte se data ook oor tyd heen met mekaar vergelyk. Hierdie deurlopende verifikasie is wel 'n aanduiding (in sommige gevalle sterker en in sommige gevalle weer swakker) van interne geldigheid. Dit is egter slegs van toepassing op 'n spesifieke doel, naamlik die van plasing, en in 'n spesifieke konteks, naamlik die Suid-Afrikaanse hoëronderrigsgewing, waar die skoolindeksamenresultate blyk nie meer so akkuraat te voorspel as in die verlede nie en waar daar addisionele taalsteun aan studente verskaf moet word.

7.3.2 Eksterne geldigheid

Eksterne geldigheid verwys na kwaliteite / attribute van 'n toets wat ekstern daartoe lê. Hier gaan dit volgens Davies *et al.* (1999:220-222) veral oor kriterium-gerigte geldigheid wat voorspellingsgeldigheid, samevallende geldigheid, en soms ook konsekwensiële geldigheid insluit. Dit gaan in essensie hier oor hoe 'n spesifieke toets, in hierdie geval TAG, “relates to phenomena external to the test, for example other tests, to future performances in the ‘real’ world, and to educational and other systems in civil society” (Yeld, 2001:248), en dit word statisties uitgedruk.

7.3.2.1 Hipotese 4: TAG beskik oor voorspellingsgeldigheid

Soos reeds hierbo gesien kan word, verwys voorspellingsgeldigheid na die akkuraatheid waarmee 'n kriterium wat ekstern tot die toets lê, geantisipeer kan word (Davies *et al.*, 1999:149). Smit (1991:46) verduidelik dit ook as 'n aspek van geldigheid wat te make het met die voorspelling van toekomstige gedrag in 'n spesifieke situasie en word tipies in kontekste gebruik waar plasing of seleksie aan die orde is - in die konteks van hierdie studie gaan dit veral oor plasing op gepaste taalondersteuningskursusse. Messick (1989:17) stel dit egter nog meer spesifiek as die outeurs hierbo wanneer hy sê die fokus van voorspellingsgeldigheid is op "... selected relationships with measures that are critical for a particular applied purpose in a specific applied setting." Voorspellingsgeldigheid en deeglike ondersoek daarna is van kardinale belang veral wanneer dit by hoëbelangtoetse kom wat byvoorbeeld vir besluite oor toelating gebruik word en waarop Messick hierbo reeds sinspeel met woordkeuses soos 'critical' en 'applied'. Die rede hiervoor is dat voorspellings oor toekomstige gedrag so akkuraat as moontlik moet wees (Bachman, 1990:250), omdat dit mense se lewens kan beïnvloed en daar 'n regverdiging vir besluite wat oor byvoorbeeld toelating geneem is, behoort te wees. Binne konteks van die teoretiese raamwerk van hierdie studie gaan dit hier veral oor die weerklank wat die juridiese dimensie van ons bestaan vind in die leidende tegniese funksie van 'n meetinstrument van taalvermoë.

Hoofstuk 1 het aangetoon dat TAG egter 'n lae- tot mediumbelangtoets is, omdat dit hoofsaaklik vir plasing gebruik word en die meeste juridiese en politieke kwessies daardeur reeds uit die weg geruim is. Vir die Universiteit Stellenbosch word TAG as deel van die toegangstoetsbattery gebruik wat die belang daarvan verhoog. TAG of TALL dra by hierdie universiteit egter 'n gewig van slegs 13% wanneer die finale toelatingpunt bereken word (sien Bylaag B vir 'n verduideliking van hoe die finale toelatingpunt bereken word). Boonop tree fakulteitspesifieke toelatingsvereistes in werking nadat die toelatingpunt bereken is, wat die effek van TAG of TALL verder versag, omdat hierdie spesifieke vereistes slegs op skoolpunte gegrond is. Voorts is daar reeds eerste blyke dat die nuwe Nasionale Senior

Sertifikaat (NSS) redelik akkuraat voorspel en die nut van addisionele toetsbatterye vir die doeleindes van toelating word tans versigtig bevraagteken. Longitudinale data oor die NSS is egter nodig voordat finale besluite oor die gewig daarvan vir toelating geneem kan word. Die indruk is tans is dat die Universiteit Stellenbosch vir die volgende paar jaar óf nog van sy toegangstoetsbatterye gebruik sal maak, óf waarskynlik van die National Benchmark Tests (NBT's) waarin daar op nasionale vlak belangstelling is. Soos egter vroeër genoem, is hierdie toetse tans in 'n loodsfasie en daar moet nog bepaal word of dit die toets van die tyd gaan deurstaan. Let op hierdie punt weer daarop dat hoewel die NBT's as plasingsinstrumente voorgehou word, dit myns insiens in praktyk as deel van toelatingsmeganismes gebruik sal word. Hoe dit ook al sy, alle aanduidings is dat TAG en TALL by hierdie universiteit in die toekoms weer vir hul oorspronklike doel, naamlik plasing op gepaste taalondersteuningskursusse weens die diagnostiese aard daarvan, gebruik sal word. TAG en TALL gaan dus weer laebelangtoetse word en ondersoek na die voorspellingsgeldigheid daarvan word bloot ter wille van deeglikheid gedoen, eerder as om akkurate uitsprake oor individue se vermoëns te maak. Daar hoef dus nie vir doeleindes van hierdie studie met intensiewe en 'n verskeidenheid van ondersoek na die voorspellingsgeldigheid van TAG en TALL omgegaan te word nie. Daar is egter volgens Bachman (1990:251) 'n voorwaarde hieraan verbonde:

as long as one is only interested in predicting future performance. That is, it is possible to find or develop tests that will predict some future behavior ... In situations where predictive validity is the primary consideration, there is a tendency to simplify, to reduce the number of measures we use to the smallest set, or to the single measure that provides the greatest accuracy of the prediction."

Hoewel dit dus nie krities belangrik is om die akkuraatheid van die voorspellingswaarde van TAG en TALL te ondersoek nie (om Messick en Bachman hierbo se terme te gebruik), word dit tog gedoen, omdat akademiese taalvermoë geag word as bydraend tot akademiese sukses (Botha, Du Plessis & Menkveld, 2007a,b; en Frick, 2007). Dit artikuleer met die toetsontwikkelaars van TAG en TALL se argumente dat daar versigtig met TAG en TALL omgegaan moet word as deel van toelatingsmeganismes, omdat taal slegs een van 'n aantal veranderlikes is wat tot akademiese sukses bydra. Die vraag wat egter ontstaan, is wat aanvaarbare

vlakke van voorspelling ('n aanduiding van sukses of nie) behoort te wees, en daar is geen definitiewe aanduiding hiervan in die literatuur nie. Verder bemoelik kwessies soos populasiegrootte, homogene populasies, en die taal waarin die toets afgelê word ondersoek na die voorspellingsgeldigheid van 'n toets. Die kwessie van tydsverloop en ander invloede soos blootstelling aan die akademiese omgewing dra ook by tot laer voorspellingsgeldigheid. McNamara (1996:21) gee egter tog 'n vae aanduiding van waar die variansieverklaring behoort te lê wanneer hy verwys na ondersoek wat oor die voorspellingsgeldigheid van akademiese taaltoetse gedoen is. Hy wys daarop dat akademiese taalvaardigheid in hierdie ondersoek slegs 'n maksimum van 10% van akademiese prestasie in die algemeen verklaar het. In die analises wat hieronder volg, sal gesien word dat taal inderdaad, en ten regte ook, 'n lae maar beduidende persentasie van akademiese prestasie verklaar. Dit is nie nodig nie, en trouens selfs verkieslik, dat daar nie 'n hoë korrelasie tussen taal en akademiese prestasie in die algemeen moet wees nie (cf. McNamara, 1996; Botha, Du Plessis & Menkveld, 2007a,b; en Frick, 2007). Die teendeel is egter ook waar en Davies *et al.* (1999:149) se voorbeeld illustreer dit goed:

[A] test of English for academic purposes [EAP] is said to have high predictive validity if performance on the test correlates highly with performance ... on a subsequent academic course which is taught through the language under test.

Omdat hierdie studie ten dele die gebruik van TAG en TALL op die eerstejaarspopulasie in die algemeen by drie verskillende universiteite bestudeer, is daar nie op studente wat EAP-kursusse volg en waarna Davies *et al.* hierbo verwys, gefokus nie. Die problematiek wat in die vorige paragraaf uitgewys is, het ook tot hierdie besluit bygedra. Dit sou wel interessant wees om dit te ondersoek, maar dit val buite die bestek van hierdie studie.

Dit is dus origens nie krities belangrik dat daar akkurate voorspellings aan die hand van TAG en TALL oor die kriterium (in hierdie geval akademiese prestasie) gemaak hoef te word nie, omdat TAG en TALL as laebelangtoetse geag kan word. Daar is gevolglik besluit om slegs op sogenaamde korrelasiestudies te fokus, en ook omdat datastelle beskikbaar was waarop hierdie tipe analises gedoen kon word. Let daarop dat die fokus hier is op die verband tussen faktore en nie op die relatiewe

voorspellingsbydrae van die onderskeie faktore nie. Om hierdie rede is daar hier nie van regressie-analises gebruik gemaak nie, maar slegs van korrelasiestudies. Ander vorme van ondersoek na voorspellingsgeldigheid van meetinstrumente sluit sogenaamde ‘survival analyses’ in, en die vergelyking van toetsresultate met die oordele van die portuurgroep / vakspesialiste / dosente, ensovoorts. Van der Walt en Steyn, (2007:149-151) het wel al ten dele, maar meer implisiet as eksplisiet, hierop gefokus.

Yeld (2001:249-250) stel dit dat korrelasiestudies, soos die hieronder, slegs geregverdig is onder die volgende omstandighede:

- Daar moet sekerheid wees dat daar ’n verband tussen die toets en die kriterium bestaan en dit word slegs verkry deur vorige studies. Dit is wel die geval by hierdie ondersoek, veral as daar na die werk van Botha, Du Plessis & Menkveld (2007a,b), Frick (2007) en Van der Walt en Steyn, (2007) gekyk word.
- Daar moet sekerheid wees dat die verband wat daar tussen die toets en die kriterium bestaan op ’n redelik hoë korrelasie sal uitloop. Hierbo is egter geargumenteer dat dit nie altyd moontlik is nie weens verskeie redes en daar kan aanvaar word dat ’n lae, maar beduidende korrelasie met ’n variansieverklaring van om en by 10% voldoende sal wees vir hierdie ondersoek.
- Die gedragsaspek (in hierdie geval gemiddelde akademiese prestasie) wat geantisipeer word, moet nie te ver in die toekoms lê nie, omdat daar ’n groter wordende veelheid van veranderlikes is wat dit dan kan beïnvloed.
- Groepe (so groot as moontlik) en nie individue nie, moet ondersoek word.

Hieronder word daar na drie ondersoekes verwys. Die eerste is ’n ondersoek wat by die Universiteit van Pretoria deur Smit, Boraine en Owen in 2006 gedoen is en wat gebaseer was op verskillende datastelle afkomstig van die 2004-eerstejaarsinname aan hierdie universiteit. Die doel hiervan was om te bepaal tot watter mate inligting op die UP-databasis (soos matriekresultate en die M-telling wat ’n funksie van die

matriekresultate is) tesame met toetse soos die in die AARP-battery (waarna onder andere in Hoofstuk 3 verwys is en wat vir doeleindes van hierdie ondersoek uit drie toetse bestaan het: PTEEP, MACH, en MCOM), en TAG en TALL. Dit is vir die eerstejaarspopulasie van 2004 as geheel gedoen en was nie 'n poging om spesifieke ondersoeke wat deur fakulteite gedoen is te weerlê nie, maar eerder om aan die universiteitsowerhede perspektief te gee op die kwessie van die gebruik van toetse addisioneel tot die matriekresultate vir universiteitstoelating (Smit, Boraine & Owen, 2006:2). Daar was beperkinge aan hierdie ondersoek verbonde, veral in die opsig dat die sentrale stelsel van hierdie universiteit nog nie ten volle geïntegreer was met fakulteitstelsels en jaarboeke nie en bepaalde inligting ontbreek het. Verder was die ondersoek nêr op studente wat reeds toegelaat is tot universiteitstudie van toepassing en nie op alle aansoekers nie, en slegs 'n gedeelte van hierdie populasie het die AARP-toetse afgelê – dit was dus 'n geselekteerde groep waarop hierdie ondersoek gedoen is en die resultate daarvan verteenwoordig nie noodwendig die hele populasie nie. Die kriterium wat gebruik is en waarteen die genoemde toetse gemeet is, was sukseskoers. Sukseskoers is gemeet op een van drie moontlikhede : (i) die aantal krediete geslaag teenoor die aantal krediete voor geregistreer op eerstejaarsvlak (Pcredit genoem); (ii) die aantal krediete geslaag teenoor die aantal krediete deur die program op eerstejaarsvlak voorgeskryf (Pprogram genoem); en (iii) die aantal modules geslaag teenoor die aantal modules voor geregistreer in die eerstejaar (Pmod genoem). Tabele 7.63 en 7.65 is die beskrywende statistiek vir onderskeidelik TAG en TALL teenoor die ander veranderlikes, en Tabele 7.64 en 7.66 is die korrelasiematriks vir onderskeidelik TAG en TALL teenoor die ander veranderlikes.

Tabel 7.63: Beskrywende statistiek vir die bepaling van voorspellingsgeldigheid vir die 2004-inname aan die Universiteit van Pretoria

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TAG	3030	60.39	12.77	20	98
Pcredit	2589	76.47	27.13	0	100
Pprog	2347	76.78	27.73	0	138
Pmod	2589	79.82	24.44	0	100
PTEEP	801	51.17	18.28	3	82
MACH	670	48.61	21.66	2	97
MCOM	671	50.76	17.73	3	89
M-telling	2582	21.36	5.67	2	30

Tabel 7.64: Korrelasies tussen TAG, Pcred, Pprog, Pmod, PTEEP, MACH, MCOM, en M-telling vir die 2004-inname aan die Universiteit van Pretoria

		TAG	Pcredit	Pprog	Pmod	PTEEP	MACH	MCOM	M-telling
TAG	R	1	0.19	0.20	0.20	0.16	0.26	0.24	0.50
	r ²		3.7%	3.9%	4.0%	2.4%	6.7%	5.7%	25%
	P		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	N		2589	2347	2589	801	670	671	2582

Tabel 7.65: Beskrywende statistiek vir die bepaling van voorspellingsgeldigheid vir die 2004-inname aan die Universiteit van Pretoria

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
TALL	3280	62.88	19.32	0	100
Pcredit	2433	73.54	28.18	0	100
Pprog	2155	72.02	29.49	0	100
Pmod	2433	77.22	25.38	0	100
PTEEP	918	52.71	20.55	2	88
MACH	825	51.77	23.18	1	97
MCOM	821	45.81	19.63	1	91
M-telling	2330	20.30	5.78	3	30

Tabel 7.66: Korrelasies tussen TALL, Pcred, Pprog, Pmod, PTEEP, MACH, MCOM, en M-telling vir die 2004-inname aan die Universiteit van Pretoria

		TALL	Pcredit	Pprog	Pmod	PTEEP	MACH	MCOM	M-telling
TALL	r	1	0.30	0.29	0.28	0.49	0.39	0.53	0.51
	r ²		9%	8.2%	7.8%	23.8%	15%	28.2%	26%
	p		<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
	N		2433	2155	2433	918	825	821	2330

Uit die bostaande kan daar gesien word dat TAG statisties beduidend met Pcredit, Pprog en Pmod korreleer (al drie aanduiders van sukseskoers). Dit is in lyn met McNamara (1996) se argument hierbo dat taal beduidend met akademiese sukses behoort te korreleer, maar nie met 'n noemenswaardige variansieverklaring nie (10% of laer). Dieselfde geld vir TALL, hoewel hierdie toets hoër gekorreleer het met die genoemde drie aanduiders van akademiese sukses. Ofskoon die korrelasies van TAG en TALL onderskeidelik met PTEEP, MACH, MCOM en die M-telling nie 'n aanduiding van akademiese sukses is nie, is dit wel by bostaande tabelle ingesluit, omdat dit deel van die oorspronklike datastelle van die Universiteit van Pretoria was, en sou kon bydra tot verdere ondersoek oor kongruente en konvergente geldigheid. Die beduidende, en in sommige gevalle selfs relatief hoë korrelasies van TAG en TALL met hierdie veranderlikes is bloot 'n aanduiding daarvan dat die een ten dele deur die ander verklaar kan word en kan dus ook as 'n aanduiding van kongruente of konvergente geldigheid (soos hierbo bespreek) gesien word wat 'n verdere rede vir die insluiting daarvan hierbo is.

Die tweede en derde ondersoek waarna hierbo verwys is, is by die Universiteit Stellenbosch gedoen. Die eerste van hierdie twee is gebaseer op die werk van Van Dyk en Kistner (2008) waar daar ondersoek ingestel is na die 2005-, 2006-, en 2007-kohorte. Let daarop dat daar besluit is om ter wille van die leesbaarheid van die Van Dyk en Kistner-verslag nie die beskrywende statistiek in te sluit nie, aangesien dit ook 'n aantal ander analyses en interpretasies ingesluit het, wat die genoemde verslag geweldig omvattend gemaak het. Omdat dit daar ontbreek en dit nie meer toeganklik is nie, word dit dus ook nie in hierdie proefskrif aangedui nie. Tabel 7.67 hieronder word derhalwe net so uit die genoemde verslag aangehaal

(Van Dyk & Kistner, 2008:5) en gee 'n aanduiding van die korrelasies tussen die akademiese geletterdheidstoetse en die geweegde gemiddelde eerstejaarsprestasie aan die einde van die eerstejaar op universiteit. Volgens hierdie ondersoek word ongeveer 8% van die variansie van eerstejaarsprestasie deur taal / akademiese geletterdheid verklaar vir die drie kohorte. Die N-waarde neem ook toe oor die jare heen en die verklaring hiervoor is dat slegs 'n deel van die populasie in 2005 die toetse afgelê het, terwyl die hele populasie dit in 2006 en 2007 afgelê het.

Tabel 7.67: Korrelasies tussen TAG en TALL en geweegde gemiddelde eerstejaarsprestasie aan die Universiteit Stellenbosch vir die 2005-, 2006-, en 2007-innames onderskeidelik

		TAG	TALL
2005	r	0.283	0.267
	r ²	8.0%	7.2%
	p	<.0001	<.0001
	N	1777	1738
2006	r	0.287	0.269
	r ²	8.2%	7.3%
	p	<.0001	<.0001
	N	3728	3735
2007	r	0.291	0.243
	r ²	8.5%	5.9%
	p	<.0001	<.0001
	N	4216	4218

Wanneer hierdie data per fakulteit opgedeel word, is dit interessant om te sien dat daardie twee fakulteite waar akademiese geletterdheidsrisiko aldus TAG en TALL die hoogste is (Opvoedkunde en Lettere en Sosiale Wetenskappe), dieselfde fakulteite is waar die variansieverklaring die laagste is. Die omgekeerde blyk ook 'n tendens te wees, naamlik dat daardie fakulteite waar akademiese geletterdheidsrisiko aldus TAG en TALL laer is, ook daardie fakulteite is waar die variansieverklaring hoër neig. Tabel 7.68 is ook oorgeneem uit die Van Dyk en Kistner-verslag (2008:10) en is 'n aanduiding van die korrelasiekoëffisiënte en variansieverklarings per fakulteit vir die drie kohorte hierbo saamgevoeg. Wat vreemd opval is dat, soos reeds genoem, die variansieverklaring vir die Fakulteite

Lettere en Sosiale Wetenskappe en Opvoedkunde laer is as vir byvoorbeeld die Fakulteite Gesondheidswetenskappe, Ingenieurswese en Natuurwetenskappe. 'n Moontlike rede hiervoor is dat hierdie fakulteite (veral die eerste twee van laasgenoemde drie) oor die algemeen sterker studente trek en die populasie dus meer homogeen is, terwyl daardie fakulteite waar die variansieverklaring laer is weer oor die algemeen 'n meer heterogene populasie trek wat akademiese vermoëns betref. Dit is ook interessant om te let op die aansienlik hoër variansieverklarings vir die Fakulteite Regsgeleerdheid, Teologie en Ekonomiese en Bestuurswetenskappe; dieselfde argument (populasiesamestelling) as hierbo kan waarskynlik ook hier geld. Wat wel verblydend is, is dat die korrelasies oor die algemeen beduidend is en daar met redelike sekerheid aanvaar kan word dat die TAG- en TALL-resultate minstens vir 'n deel die variansie in akademiese prestasie op eerstejaarsvlak verklaar. Verder toon die analise ook dat die variansieverklaring, of dit nou vir die groep as geheel is, of vir subgroepe, in lyn is met internasionale tendense. Dit is ook min of meer in lyn met bevindinge wat uit die analise vir die Universiteit van Pretoria gemaak is en die twee ondersoekte sou dus gebruik kon word om mekaar te verifieer. Op grond van die verifikasie sou daar aangevoer kon word dat TAG en TALL oor voorspellingsgeldigheid beskik in terme van eerstejaarsprestasie, hoewel lae voorspellingsgeldigheid, wat in lyn is met internasionale tendense (sien argument van McNamara, 1996 hierbo).

Tabel 7.68: Korrelasies tussen TAG en TALL en geweegde gemiddelde eerstejaarsprestasie per fakulteit aan die Universiteit Stellenbosch vir die 2005-, 2006-, en 2007-innames saamgevoeg

		TAG	TALL
Agri-wetenskappe	r	0.17	0.10
	r ²	2.9%	0.9%
	p	<.0001	0.0179
	N	596	595
EBW	r	0.37	0.26
	r ²	13.5%	6.9%
	p	<.0001	<.0001
	N	2593	2600

Gesondheidswetenskappe	r	0.26	0.29
	r ²	6.6%	8.7%
	p	<.0001	<.0001
	N	634	634
Ingenieurswese	r	0.27	0.23
	r ²	7.1%	5.3%
	p	<.0001	<.0001
	N	748	746
Lettere en Sosiale Wetenskappe	r	0.24	0.29
	r ²	5.5%	8.5%
	p	<.0001	<.0001
	N	1860	1866
Natuurwetenskappe	r	0.28	0.25
	r ²	8.1%	6.3%
	p	<.0001	<.0001
	N	1041	1041
Opvoedkunde	r	0.15	0.22
	r ²	2.3%	4.8%
	p	0.0139	0.0003
	N	266	265
Regsgeleerdheid	r	0.34	0.25
	r ²	11.8%	6.3%
	p	<.0001	0.0015
	N	157	157
Teologie	r	0.38	0.33
	r ²	14.1%	11.2%
	p	0.0079	0.0189
	N	49	49

In 'n verdere ondersoek wat by die Universiteit Stellenbosch gedoen is, is die 2007-, 2008, en 2009-kohorte vir die analise gebruik. In die onderstaande berekeninge verwys Jaar 1, Jaar 2, en Jaar 3 onderskeidelik na die geweegde gemiddeld wat aan die einde van elke jaar van studie behaal is: Jaar 1 dus na die resultate aan die einde van die eerste studiejaar, Jaar 2 die einde van die tweede studiejaar, ensovoorts. Let daarop dat die aantal krediete wat 'n vak tel, die aantal krediete is wat die gewig bepaal in die geweegde gemiddeld; die gemiddeld word dus as't ware geweeg volgens die aantal krediete. In hierdie data word daar nie in die tweede of derde studiejaar 'n onderskeid getref tussen werklike tweedejaarsvakke en/of werklike derdejaarsvakke nie en of die student byvoorbeeld nog eerstejaars- of

tweedejaarsvakke volg wat hy / sy gedruip het nie. Hoewel dit beperkend van aard is (soos ook die geval met die ondersoek aan die Universiteit van Pretoria hierbo), kon dit nie in die analise geakkommodeer word nie as gevolg van die instelling se stelsels wat nie noodwendig in alle opsigte die onderskeid maak nie.

Tabel 7.69 is die beskrywende statistiek vir die 2007-inname aan die Universiteit Stellenbosch. Tabel 7.70 gee 'n aanduiding van die korrelasiekoëffisiënte en variansieverklarings van TAG met die 2007-inname se geweegde eerstejaars-, tweedejaars-, en derdejaarsgemiddeld. Die variansieverklaring neem, soos verwag, af met die tyd saam, wat in lyn is met een van die gevare wat Yeld hierbo uitgewys het wanneer korrelasiestudies oor tyd heen gedoen word. Tabel 7.71 toon voorts aan wat die gemiddelde prestasie op TAG was van daardie studente wat die graad in drie jaar voltooi het, en die gemiddelde prestasie op TAG van daardie studente wat nog besig is met hul studies of dit reeds gestaak het.

Tabel 7.69: Beskrywende statistiek vir die bepaling van voorspellingsgeldigheid vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
Jaar 1	4060	52.74	18.53	0	96.51
Jaar 2	3540	55.39	15.10	0	94.35
Jaar 3	3408	56.31	14.84	0	92.67
TAG	4057	51.97	18.80	0	96

Tabel 7.70: Korrelasies tussen TAG met geweegde eerstejaars-, tweedejaars, en derdejaarsgemiddeld vir die 2007-inname aan die Universiteit Stellenbosch

		Gemiddeld: Jaar 1	Gemiddeld: Jaar 2	Gemiddeld: Jaar 3
TAG	R	0.29	0.24	0.19
	r ²	8.6%	5.7%	3.8%
	P	<.0001	<.0001	<.0001
	N	4057	3540	3408

Tabel 7.71: Gemiddelde prestasie in TAG volgens driejarige graad in minimum tydperk behaal vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	N	Mean: TAG	Std Dev
Slaag graad: JA	1112	55.70	18.2
Slaag graad: NEE	1541	47.96	17.4

Tabelle 7.72 en 7.73 is onderskeidelik die beskrywende statistiek vir die 2008-inname en die korrelasies en variansieverklarings tussen TAG en die geweegde eerstejaars-, en tweedejaarsgemiddeld vir die 2008-inname aan die Universiteit Stellenbosch. Die twee tabelle daarna is vir die 2009-inname aan hierdie universiteit.

Tabel 7.72: Beskrywende statistiek vir die bepaling van voorspellingsgeldigheid vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
Jaar 1	4191	53.10	18.20	0	94.94
Jaar 2	3661	54.99	15.64	0	93.77
TAG	4191	50.61	19.13	0	100

Tabel 7.73: Korrelasies tussen TAG met geweegde eerstejaars-, en tweedejaarsgemiddeld vir die 2008-inname aan die Universiteit Stellenbosch

		Gemiddeld: Jaar 1	Gemiddeld: Jaar 2
TAG	r	0.31	0.25
	r ²	9.8%	6.2%
	p	<.0001	<.0001
	N	4191	3661

Tabel 7.74: Beskrywende statistiek vir die bepaling van voorspellingsgeldigheid vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	N	Mean	Std Dev	Minimum	Maximum
Jaar 1	4237	53.10	18.89	0	95.78
TAG	4235	50.05	17.05	0	97.0

Tabel 7.75: Korrelasies tussen TAG met geweege eerstejaarsgemiddeld vir die 2009-inname aan die Universiteit Stellenbosch

		Gemiddeld: Jaar 1
TAG	r	0.249
	r ²	6.2%
	p	<.0001
	N	4235

Bostaande vier tabelle is 'n bevestiging van ander analyses wat in hierdie afdeling oor voorspellingsgeldigheid aangebied is. Dit word weereens beklemtoon dat die ondersoek wat vir hierdie afdeling gedoen is bloot ter wille van konseptuele skerpheid gedoen is en dit kan dus nie geïnterpreteer word asof TAG (en by implikasie TALL) as toelatingsinstrument aangewend kan word nie. Soos reeds genoem was die fokus hier bloot om te bepaal of daar 'n statisties beduidende verband tussen die faktore is en nie om te bepaal wat die relatiewe voorspellingsbydrae van elke faktor is nie – dit word tipies en eerder gebruik by toelatingstoetse. Hoewel dit in lyn is met internasionale tendense en daar geargumenteer sou kon word dat TAG oor voorspellingsgeldigheid beskik, kan dit nie uit die oog verloor word nie dat TAG en TALL ontwerp is met die doel om addisionele inligting te verskaf oor studente se risikovolke in terme van akademiese taalvermoë. Soos reeds genoem, word 'n gedeelte van akademiese prestasie aan taal toegeskryf en vanuit hierdie perspektief sou daar wel geoordeel kon word dat toetse soos TAG en TALL belangrik is, omdat die vroegetydige identifisering van taalrisiko en die instel van gepaardgaande steunmeganismes, soos kursusse, daartoe kan bydra dat studente gemakliker in die akademiese omgewing oor die weg sal kom.

7.3.2.2 Hipotese 5: TAG beskik oor samevallende geldigheid

Samevallende geldigheid verwys na die verwantskap tussen 'n toetsresultaat en die resultaat van 'n ander meting wat min of meer op dieselfde tyd plaasgevind het (Yeld, 2001:273). Samevallende geldigheid word ook soms konkurrente geldigheid genoem en hieruit kan afgelei word dat dit gaan oor die mate waartoe variansie in die een meting, en op 'n gegewe moment, met variansie in die ander meting

ooreenstem. Smit (1991:62) stel dit dat samevallende geldigheid te make het met die “huidige status [van ’n toetsling] met betrekking tot spesifieke vaardighede en ander gedragsaspekte.” Hierdie tipe geldigheid word gewoonlik saam met voorspellingsgeldigheid bespreek, omdat eersgenoemde gaan oor die teenswoordige status en laasgenoemde oor die toekomstige status van ’n kandidaat ten opsigte van die kriterium.

Samevallende geldigheid word soms, en moontlik met goeie reg, met groep-differensiële, konvergente, en diskriminante geldigheid, wat hierbo bespreek is, verwar (verwys byvoorbeeld Robinson Kurpius & Stafford, 2006:150). Die rede hiervoor is dat dieselfde tegnieke gebruik kan word om dit te bepaal, naamlik om kontrasgroepe (soos ras, geslag, moedertaal, ensovoorts) teen mekaar af te speel, om verskillende toetse met mekaar te korreleer, en om met beoordelingskale te werk. Soos egter reeds aangevoer, is dit vir doeleindes van hierdie studie juis belangrik om konseptueel skerp te wees in die valideringsproses, sodat tegniese regverdiging vir die ontwerp (in hierdie geval TAG en TALL) moontlik is. Dit tesame met die tegnies-analitiese verdedigbaarheid van die konstruk onderliggend aan die toets sal daartoe bydra dat ’n instrument soos TAG wel die toets van die tyd deurstaan en geldig verklaar kan word om te gebruik onder bepaalde omstandighede.

Wanneer samevallende geldigheid ondersoek word, vind daar myns insiens egter ’n klemverskuiwing plaas in die opsig dat die fokus nie meer soseer is op hoe die variansie in toetsresultate van kontrasgroepe verskil of ooreenstem nie, of op wat die korrelasiekoëffisiënte behoort te wees nie, maar eerder op of dit dan hoegenaamd nodig is om ’n ander toets te ontwerp as daar reeds ’n legitieme en relevante produk op die mark is (Yeld, 2001:273). Daar sou byvoorbeeld geargumenteer kon word dat dit nie nodig was om ’n toets soos TAG of TALL te ontwikkel nie, omdat daar reeds ’n toets soos PTEEP op die mark was, en soos reeds gesien is daar beduidende hoë korrelasies tussen hierdie twee toetse (veral die Engelse weergawes daarvan). Verder ontstaan die vraag of dit dan werklik nodig was om ’n toets soos die NBT te ontwikkel, met toetse soos PTEEP en TAG en

TALL op die mark. Davies *et al.* (1999:30) stel die problematiek met ondersoek na samevallende geldigheid, en binne konteks van die voorbeeld wat in die vorige twee sinne aangebied is, dan tereg ook as

[t]he problem of concurrent validity is how much emphasis to put on the existing criterion measure. If the criterion is itself valid, then the need for the test under trial is questionable. On the other hand, if the test under trial represents a new departure, then there will be little point in validating it against the existing test, which is presumably regarded as an imperfect measure of the relevant construct of ability.

Soos telkemale reeds in die vorige twee hoofstukke, sowel as hierdie een, aangevoer, word toetse met 'n bepaalde doel ontwerp en is slegs toepaslik vir gebruik in 'n spesifieke konteks (verwys Hoofstuk 3). Daar is ook aangevoer dat daar gewoonlik 'n dringendheid aan die werk van veral toegepaste taalkundiges is, naamlik om vinnig met oplossings vir taalprobleme vorendag te kom. Boonop moet hierdie oplossings ook ekonomies haalbaar en volhoubaar wees. Daar ontstaan dus 'n geneigdheid om toetse vir doeleindes aan te wend waarvoor dit nie ontwerp is nie: so kan PTEEP byvoorbeeld benewens sy doel om potensiaal te meet en dus as toelatingsmeganisme gebruik te word ook aangewend word om homself voor te doen as 'n plasingstoets van akademiese geletterdheid. Net so kan TAG en TALL vir ander doeleindes aangewend word as waarvoor dit ontwerp is, naamlik 'n toelatingsinstrument eerder as 'n plasinginstrument na aanleiding van 'n meting van akademiese geletterdheid. En, soos reeds genoem, bestaan daar 'n risiko dat die NBT wat in wese ook 'n toets is wat potensiaal en gereedheid vir universiteitstudie meet, gebruik kan word as toelatingsmeganisme eerder as die plasingmeganisme waarvoor dit gepropageer word. Hierdie paragraaf dien dus eintlik as vermaning vir universiteitsowerhede om versigtig met toetse om te gaan, sodat dit binne konteks van die teoretiese raamwerk van hierdie studie aan die eise van veral die etiese en die juridiese dimensies van ons bestaan sal voldoen en nie vir doeleindes aangewend word waarvoor dit nie bedoel is nie.

Aangesien konseptuele skerpheid 'n fokus van hierdie hoofstuk is, word daar vervolgens 'n uiteensetting aangebied van die ondersoek wat na die samevallende geldigheid van TAG onderneem is, met die inagneming van die problematiek wat in

die aanhaling van Davies *et al.* hierbo en die bespreking daarna uitgewys is. Dit word gedoen binne die konteks van TAG en TALL wat as unieke plasingstoetse (in Afrikaans en in Engels – nie vertaal nie) vir akademiese geletterdheid ontwerp is vir die Suid-Afrikaanse hoërsonderwysomgewing.

Ondersoeke na samevallende geldigheid kan soos reeds genoem op drie maniere aangepak word, naamlik die van beoordeling van variansie in kontrasgroepe, korrelasieondersoeke, en die gebruik van beoordelingskale. Die eerste twee is die tegnieke wat die meeste in die literatuur omskryf en gebruik word, waarskynlik omdat dit die mees haalbare is. Laasgenoemde is prakties gesproke moeiliker haalbaar en die kwessie van inter-beoordelaar-betroubaarheid bring weer 'n problematiek van sy eie mee. Indien ondersoeke wat van beoordelingskale gebruik maak verder nie sistematies en weldeurdrag aangepak word nie, is daar 'n menigte veranderlikes wat tot konstruk-irrelevante variansie kan bydra. Hoewel daar anekdotiese getuienis is van dosente dat die studente in hul klasse wat synde deur TAG en TALL as kandidate met hoë of lae risiko aangemerkt is, ooreenstem met hul ervarings van die betrokke studente se akademiese geletterdheidsvermoëns, word dit as gevolg van bogenoemde redes nie hier aangespreek nie en daar is besluit om vir doeleindes van hierdie deel van die studie slegs op eersgenoemde twee metodes te fokus, wat hieronder bespreek sal word.

Die eerste het te make met die gebruik van kontrasgroepe. Kontrasgroepe verwys na verskille soos geslag, ras, moedertaal, vermoë op die kriterium (in hierdie geval akademiese geletterdheid), ensovoorts. Aangesien geslag, ras en vermoë ('n vergelyking van die boonste en die onderste kwartiel van die populasie) reeds hierbo ondersoek en bespreek is, word dit nie hier weer herhaal nie, behalwe om aan te toon dat gegewe hierdie inligting daar wel rede is om te glo dat TAG oor samevallende geldigheid beskik, veral wat ras en vermoë betref. Die invloed van moedertaal op die kriterium is egter doelbewus nie hierbo ondersoek nie, omdat dit hier verken sal word. Dieselfde statistiese tegnieke as vir groep-differensiële, en

konvergente en diskriminante geldigheid word weer hier gebruik – verwys veral afdeling 7.3.1.3.4 vir ’n uiteensetting daarvan.

Vir die ondersoek na verskillende moedertaalgroepe is die 2007-, 2008-, en 2009-innames van die Universiteit Stellenbosch ontleed. Die indeling in moedertaalgroepe word by hierdie universiteit gedoen op grond van dit wat die studente self aangedui het as hul moedertaal, te wete Afrikaans, Engels, Afrikaans/Engels, SA Taal (enige van die oorblywende amptelike Suid-Afrikaanse tale), en Ander (alle ander tale wat nie onder die voorgenoemde kan resorteer nie, byvoorbeeld ander Afrikatale en Europese tale).

Tabelle 7.76 tot 7.78 is ’n aanduiding van die analyses wat vir ’n samevallende ondersoek na verskille in moedertaal vir die 2007-inname aan die Universiteit Stellenbosch gedoen is. Let daarop dat daar hier ook slegs tendense te bespeur is, omdat die ANOVA-aannames vir nie een van die ondersoeksjare gegeld het nie. Daar word dus net soos by die afdeling hierbo wat oor groepdifferensiële geldigheid gehandel het geen sterk uitsprake gemaak nie. Die data was in alle gevalle normaal versprei.

Tabel 7.76: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per moedertaal vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	N	TAG Mean	TAG Std Dev	TAG Std Err	TAG 95.00%	TAG +95.00%
Totaal	4032	51.92	18.69	0.29	51.34	52.49
Afrikaans	2582	59.98	15.33	0.30	59.39	60.56
Engels	1342	38.56	14.78	0.41	37.75	39.37
SA Taal	108	25.11	14.60	1.46	22.26	27.97

Tabel 7.77: Levene se toets vir homogeniteit van variansies ten opsigte van moedertaal vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	MS – Effect	MS - Error	F	p
TAG	1041.6	80.25	12.978	<0.0001

Levene se toets hierbo toon aan dat daar beduidende verskille is in die variansie van die drie moedertaalgroepe en die aannames vir 'n ANOVA geld dus nie. Die tendens word egter waargeneem dat daar verskille in gemiddeld tussen die verskillende moedertaalgroepe is. Soos reeds genoem, kon 'n ANOVA dus nie gedoen word nie en Bonferroni se toets is vervolgens uitgevoer. Tabel 7.78 is 'n aanduiding hiervan:

Tabel 7.78: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2007-inname aan die Universiteit Stellenbosch

Veranderlike	Afrikaans – 59.98	Engels – 38.56	SA Taal – 25.11
Afrikaans		0.00	0.00
Engels	0.00		0.00
SA Taal	0.00	0.00	

Uit die bostaande kan daar reeds 'n tendens gesien word, naamlik dat studente met Afrikaans as moedertaal statisties gesproke beter presteer as Engelssprekende studente wat weer beter as sprekers van ander Suid-Afrikaanse presteer. Let daarop dat hierdie weereens geen sterk uitspraak is nie; dit is bloot 'n tendens wat waargeneem word. Omdat die steekproewe egter relatief groot is, is die effek van die afwykings van die aannames klein en daar is wel afleidings hieruit te maak, naamlik dat die verskille tussen moedertaalgroepe as blyke van samevallende geldigheid sou kon dien, en net so vir groep-differensiële geldigheid, omdat dit tog lyk of die toets diskrimineer tussen groepe waar moedertaal 'n invloed sou kon hê op prestasie op TAG. Dit kan egter nie as feit gestel word nie.

Vervolgens word soortgelyke analyses vir die 2008- en 2009-kohorte van die Universiteit Stellenbosch voorgehou en Tabele 7.79 tot 7.84 is 'n voorstelling hiervan.

Tabel 7.79: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per moedertaal vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	N	TAG Mean	TAG Std Dev	TAG Std Err	TAG 95.00%	TAG +95.00%
Totaal	4118	50.63	19.19	0.30	50.05	51.23
Afrikaans	2583	59.25	16.03	0.30	58.65	59.84
Engels	1414	37.38	14.35	0.41	36.58	38.18
SA Taal	121	21.80	13.44	1.40	19.05	24.54

Tabel 7.80: Levene se toets vir homogeniteit van variansies ten opsigte van moedertaal vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	MS – Effect	MS - Error	F	p
TAG	2539.8	85.13	29.84	<0.0001

Vervolgens is die Bonferroni-toets gedoen om vas te stel tussen watter groepe die verskil voorkom en Tabel 7.81 is 'n aanduiding hiervan.

Tabel 7.81: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2008-inname aan die Universiteit Stellenbosch

Veranderlike	Afrikaans – 59.25	Engels – 37.38	SA Taal – 21.80
Afrikaans		0.00	0.00
Engels	0.00		0.00
SA Taal	0.00	0.00	

Die 2009-inname se resultate het as verifikasie gedien vir bogenoemde twee innames in die opsig dat daar bevind is dat dieselfde tendens as hierbo heers. Tabelle 7.82 tot 7.84 is 'n aanduiding van die analyses wat hiervoor gedoen is.

Tabel 7.82: Beskrywende statistiek vir die bepaling van verskil in gemiddelde TAG-prestasie per moedertaal vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	N	TAG Mean	TAG Std Dev	TAG Std Err	TAG 95.00%	TAG +95.00%
Totaal	5576	47.51	17.04	0.23	47.06	47.96
Afrikaans	3231	55.78	13.65	0.24	55.30	56.25
Engels	2129	37.67	13.54	0.29	37.09	38.24
SA Taal	216	20.91	14.91	1.01	18.91	22.91

Tabel 7.83: Levene se toets vir homogeniteit van variansies ten opsigte van moedertaal vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	MS – Effect	MS - Error	F	p
TAG	535.77	70.17	7.635	0.0005

Vervolgens is die Bonferroni-toets gedoen om vas te stel tussen watter groepe die verskil voorkom en Tabel 7.84 is ’n aanduiding hiervan.

Tabel 7.84: Bonferroni se toets vir die bepaling van verskil in gemiddelde TAG-prestasie per ras vir die 2009-inname aan die Universiteit Stellenbosch

Veranderlike	Afrikaans – 55.78	Engels – 37.67	SA Taal – 20.91
Afrikaans		0.00	0.00
Engels	0.00		0.00
SA Taal	0.00	0.00	

Uit die bostaande kan daar afgelei word dat daar blyke van samevallende geldigheid by TAG is as die moedertaalsprekers van Afrikaans, Engels en SA Taal met mekaar vergelyk word, omdat daar oënskynlike diskriminasie tussen hierdie groeperinge is. Die 2007-, 2008- en 2009-kohorte se analyses het mekaar oor en weer geverifieer en soos reeds hierbo blyk dit asof die toetse diskrimineer tussen groepe waar moedertaal ’n invloed sou kon hê op prestasie. Hierdie inligting moet egter ook saam met die analyses wat hierbo vir groep-differensiële, en konvergente en diskriminante geldigheid gedoen is, geles word om ’n finale oordeel te kan vel oor

samevallende geldigheid. Bostaande moet ook geïnterpreteer word teen die agtergrond daarvan dat daar aanvaar word dat moedertaalsprekers van Afrikaans se prestasie op TAG beter sal wees as die van sprekers van ander moedertale en daar sou dus kritiek uitgespreek kon word oor die veralgemening wat hier plaasvind. Ten einde dit teen te werk, word daar eerstens terugverwys na die ondersoek wat gedoen is vir groep-differensiële geldigheid op die boonste en die onderste kwartiel van die innames. Daar word tweedens egter veral terugverwys na die ondersoek wat vir konvergente en diskriminante geldigheid gedoen is en waar daar 'n onderskeid tussen die populasie as geheel en daardie studente met respektiewelik Afrikaans en Engels as moedertale getref is. Juis dit is bevestiging daarvan dat sistematiese interpretasie van 'n veelheid van argumente aan die orde moet wees wanneer toetse ondersoek word: as 'n stel data aangebied word as bewyse vir 'n bepaalde tipe geldigheid, móét dit in samehang met ander bewyse gelees word, omdat die een die ander minstens ten dele kan verklaar. Hierdie stelling is weereens tekenend daarvan dat konseptuele vaagheid daartoe kan bydra dat empiriese data misïnterpreteer kan word. Misïnterpretasie van empiriese inligting, veral inligting wat betrekking het op die objektiewe resultate (die effekte van die tegniese meetproses, wat die fisies-energetiese dimensie van ons bestaan analogies reflekteer), kan aanleiding gee tot probleme op juridiese en etiese vlak. Dit moet dus weer hier beklemtoon word dat die konstitutiewe en regulatiewe kante van die teoretiese raamwerk van hierdie studie in wisselwerking met mekaar gesien moet word. Daar sal inderdaad soms kompromieë na die een of die ander kant aangegaan moet word, maar in tegniese sin moet daar altyd 'n balans wees tussen byvoorbeeld die verdedigbaar-diskriminerende aard (gelding) van toetse, en die ewe oorwoë omsien na die belange van diegene op wie die resultate 'n impak gaan hê. Dit is juis na hierdie tegniese-etiese kwessies waarna daar in die volgende afdeling verwys gaan word.

7.3.2.2 Hipotese 6: TAG beskik oor konsekwensiële geldigheid

Konsekwensiële geldigheid handel oor die potensiële sosiale en juridiese gevolge wat 'n toetsresultaat op 'n toetsling kan hê. In 'n land soos Suid-Afrika is dit

daarom van kardinale belang om op die sosiale en politieke implikasies van toetsresultate te let. Die bestek van hierdie studie laat nie toe dat hierdie kwessie naastenby tot sy volle reg ondersoek kan word nie, omdat dit in hoofsaak fokus op die konstitutiewe kant van die teoretiese raamwerk waarteen dit gekontekstualiseer is. 'n Studie soos die van Rambiritch (s.a.) wat op die regulatiewe kant van die teoretiese raamwerk van hierdie studie gebaseer is, sal die kwessie van die sosiale meer tot sy reg laat kom. Hoewel hierdie studie dus geen poging is om selfs net voor te gee dat dit die antwoorde op sosiale kwessies rondom toetsing het nie, wat die Rambiritch-studie behoort te bereik, behoort dit teen hierdie tyd duidelik te wees dat die konstitutiewe en die regulatiewe nie kan funksioneer sonder wisselwerking nie, wat goed deur die volgende woorde van Davies *et al.* (1999:31) geïllustreer word:

Consequential validity places the onus on the test developer to provide evidence not only that test interpretation and use are both in line with intended testing purposes, but that they are also in line with other social values, in other words, to ensure that the score based interpretations and actions reflect accepted social values.

Vir doeleindes van hierdie studie word daar dus geen ondersoek na die konsekwensiële geldigheid van TAG gedoen nie en daar sal derhalwe ook nie uitsprake daarvoor gemaak word nie. Wat hieronder volg, is egter 'n argument vir juis waarom ondersoek na die sosiale implikasies van toetse van belang is.

'n Hele aantal interpretasies van konsekwensiële geldigheid word geassosieer met die sosiale implikasies van die gebruik van toetse. So byvoorbeeld kan vrae oor die gebruik van toetse op verskillende groepe gestel word, die formaat waarin die toetsitems voorkom, die wyse waarop die gebruik van toetse toetslinge affekteer tydens en veral na die toets op die kort-, medium- en langtermyn, ensovoorts. 'n Toets wat aan die eise van konsekwensiële geldigheid voldoen, moet myns insiens daarom eerstens regverdig en geskik wees vir die doelgroep. Dit moet dus vry wees van aspekte wat tot bevooroordeelings sal bydra, oftewel konstruk-irrelevante verteenwoordiging, maar binne konteks van dit wat hierbo bespreek is (omdat dit juis nodig is dat 'n toets moet kan diskrimineer). Wat egter van belang is, is dat dieselfde patrone wat vir een groep ontvou ook vir 'n ander groep behoort te

ontvou, gegewe die spesifieke konteks. 'n Toets behoort minstens te meet wat dit voorgee om te meet (sien ook die bespreking in die vorige hoofstuk en ten dele in hierdie hoofstuk hieroor) en juis nie dit meet wat dit nie veronderstel is om te meet nie, soos geslag, ouderdom, ras, ensovoorts. Let egter weereens daarop dat dit soms nodig is dat toetse juis móét kan diskrimineer tussen onderskeie groepe, afhangend van die konteks. Trouens, indien die kwessie van bevooroordeling in toetse ondersoek sal word, bestaan die vermoede dat daar 'n oneindige hoeveelheid maniere sal wees waarop 'n toets sosiaal nie aanvaarbaar sal wees nie, en daarom is dit belangrik om weer die belangrike kwessie van konteks uit te lig: 'n toets moet geskik wees vir 'n bepaalde doel, op 'n bepaalde tydstip, in 'n bepaalde omgewing. Daar is geen manier waarop daar gewaak kan word teen alle moontlike veranderlikes wat tot konstruk-irrelevante variansie kan bydra nie en pogings om dit te probeer doen sal waarskynlik terwyl dit die een probleem aanspreek, weer 'n ander probleem op 'n ander plek veroorsaak. Wat belangrik is, is dat daar wel toegewings gemaak sal moet word ten einde 'n gebalanseerde oordeel oor die geldigheid van 'n toets te kan vel.

Ten tweede behoort 'n toets wat aan die eise van konsekwensiële geldigheid voldoen minstens bruikbaar, regverdigbaar en verdedigbaar te wees vir die doel waarvoor dit aangewend word. In aansluiting by die vorige paragraaf moet dit egter in ag geneem word dat toetsvalidering juis gaan oor 'n proses en hoe 'n saak ten tyde van hierdie proses vanuit verskeie kante beskou kan word. Dit bring weereens die tema na vore wat deur die analises in hierdie hoofstuk gedemonstreer word, te wete dat taaltoetsontwerpers in staat moet wees om op tegnies-sistematiese wyse 'n veelheid van bewyse en analises byeen te kan bring in 'n argument ter ondersteuning van die geldigheid van die meetinstrument wat hul ontwerp het. Die rede hiervoor is dat 'n toets of 'n toetsresultaat maar net een veranderlike in die proses van validering is. Validering gaan juis oor 'n beoordeling van die proses as geheel en daarom is die teoretiese raamwerk waarteen hierdie studie gekontekstualiseer is na my oordeel besonder bruikbaar vir 'n argument oor die sosiale aspekte rondom toetse: dit belig nie net die grondliggende en teoreties-

verdedigbare aspekte (wat die basis van die ontwerp vorm) van 'n instrument soos 'n toets nie. Dit ontsluit ook hierdie grondliggende aspekte deur ondersoek in te stel na die artikulasie van die ontwerp, die implementering daarvan, die bruikbaarheid daarvan, die regverdiging daarvoor, die teoretiese verdedigbaarheid daarvan, ensovoorts.

Om hierdie redes kan die toetsontwikkelaars van 'n toets nie altyd verantwoordelik gehou word vir die onverantwoordelike gebruik daarvan nie, veral nie in kontekste wat buite hul beheer is nie (Davies *et al.* 1999:31). Studies soos hierdie een en die van Rambiritch (s.a.) is egter van kardinale belang om juis die gronde te verskaf waarop toetse soos TAG en TALL met verantwoordelikheid gebruik kan word.

Ten slotte kan dit gestel word dat dit uit bostaande bespreking blyk dat TAG (en by implikasie TALL) wel tot 'n mate oor eksterne geldigheid beskik. Dit moet egter weereens in ag geneem word dat daar 'n velerlei van veranderlikes is wat 'n invloed op die analyses en interpretasies daarvan kon gehad het en dat dit afhanklik is vanuit watter perspektief dit geïnterpreteer word.

7.4 Samevatting

In die vorige twee hoofstukke is die belangrike kwessies van betroubaarheid en geldigheid aangeraak, en die verband tussen die twee is aangetoon. Hierdie hoofstuk het ten doel gehad om 'n geldigheidsargument vir TAG op te bou as deel van die proses van validering daarvan. Vroeër in hierdie studie is dit aangevoer dat 'n geldigheidsargument volgens Fulcher en Davidson (2007:20) aan vier eise moet voldoen, naamlik die eis van eenvoud (konseptuele skerpheid / fokus), die eis van koherensie (byeenbring van verskillende aspekte tot 'n samehangende geheel), die eis van bewysbaarheid (voorspellings kan slegs gemaak word op veranderlikes wat toetsbaar is), en die eis van omvattendheid (volledigheid van die argument). Daar is daarom in hierdie hoofstuk, wat veral in samehang met Hoofstuk 6 gelees moet word, so omvattend moontlik te werk gegaan deur 'n hele aantal hipoteses te ondersoek. Hierdie hipoteses is elkeen in 'n mindere of 'n meerdere mate bewys.

Die krag van 'n valideringsondersoek lê egter nie net in die bewyse nie, maar veral ook in die argument as geheel. Daarom moet die hipoteses in samehang met mekaar, maar ook binne konteks van dit wat in Hoofstukke 5 en 6 gestel is, gelees word. Vanselfsprekend is die inligting vervat in Hoofstukke 3 en 4 ook nou verweef met die bespreking in Hoofstuk 6. Teen die teoretiese raamwerk wat in Hoofstuk 2 uiteengesit is en die konteks is waarteen die res van die studie gelees behoort te word, sou daar dus geargumenteer kon word dat 'n toets soos TAG, teen die huidige ortodoksie in (verwys Hoofstuk 6), inderdaad oor geldigheid beskik. Die punt is uiteindelik dat die effek of krag van 'n toets aangetoon kan word deur dit as 'n eenheid in 'n veelheid van tegnies-relevante argumente te hanteer.

'n Finale kwessie: die verantwoordelike interpretasie van toetsresultate

8.1 Inleidend

In die vorige drie hoofstukke is die belangrike kwessies van toetsbetroubaarheid, toetsgeldigheid en toetsvalidering uitvoerig bespreek. Teorie is deurlopend op praktyk van toepassing gemaak en daar is eerstens bevind dat TAG 'n betroubare meetinstrument is. Teen die huidige ortodoksie in, wat alles begripsmatig onder geldigheid wil laat ressorteer, is daar in die tweede plek geargumenteer dat geldigheid wel 'n attribuut van 'n toets kan wees en dat 'n toets soos TAG gelding ('n objektiewe effek dus) kan hê. TAG is derdens aan 'n sistematiese proses van geldigmaking (validering) onderwerp en daar is aangetoon dat hierdie toets inderdaad as 'n geldige meetinstrument geag kan word vir die doeleindes waarvoor dit ontwerp is. Hoofstuk 3 het ten vierde 'n verdere belangrike kwessie aangeraak, naamlik die van die teoretiese regverdigbaarheid (soms ook konstruktgeldigheid genoem) van die toets. Daar is in hierdie hoofstuk aangetoon hoe daar konseptueel te werk gegaan is om die onderbou van TAG uit te pluus en uiteindelik te artikuleer.

Die vraag wat egter nog nie in hierdie studie beantwoord is nie, is die van die interpretasie van toetsresultate en aanverwante kwessies, met ander woorde die kwessie wat deur Messick c.s. geag word as deurslaggewend te wees in hulle

ortodokse opvatting dat dit die geldigheid van die resultate van 'n toets is, met verwysing na vele stelle gegewens waaroor dit in die proses van toetsvalidering eintlik gaan. Op hierdie punt is dit weereens belangrik om daarop te wys dat die interpretasie van 'n toetsresultaat nie 'n toets geldig of ongeldig kan maak nie, soos wat die huidige ortodoksie dit graag voorhou. Wat vanselfsprekend wel uiters belangrik is, is dat 'n toetsresultaat verantwoordelik geïnterpreteer moet word. Die doel met hierdie hoofstuk is daarom om kortliks, en veral in aansluiting by die voorlaaste afdeling van Hoofstuk 7 wat oor konsekwensiële geldigheid gehandel het, aan te toon hoe daar in werklikheid met die resultate van TAG, en aanverwante kwessies, omgegaan word.

8.2 Die akkommodering van 'n verskeidenheid van perspektiewe

Die opmerkings van Van Schalkwyk (2008:27, 28) waarna in Hoofstuk 3 verwys is, blyk in aansluiting by die bostaande 'n goeie illustrasie te wees van die wisselwerking wat daar tussen die konstitutiewe en die regulatiewe kante van die teoretiese raamwerk van hierdie studie behoort te bestaan. Dit is naamlik dat hoewel die konstruk van TAG 'n oorwoë en deeglik geartikuleerde definisie van akademiese geletterdheid blyk te wees, wat deur empiriese gegewens verder ondersteun word, dit egter juis as gevolg hiervan die risiko kan inhou dat dit ondermynend kan wees tot spesifiek onderrig en leer wat in 'n dinamiese sosiale konteks plaasvind en waar daar 'n veelheid van perspektiewe is wat geakkommodeer sal moet word wanneer toetsresultate geïnterpreteer word. Hierdie kwessie word uitvoerig in Shohamy (2001) en Chapelle *et al.* (2009) bespreek en het verskeie kere reeds in hierdie studie aan bod gekom. Die punt is uiteindelik dat toetsresultate gevolge vir toetslinge kan inhou, veral waar dit by hoëbelang toetse kom, maar selfs ook by lae- tot mediumbelangtoetse soos TAG en TALL. Die potensiële effekte daarvan moet daarom geantisipeer en so omsigtig moontlik hanteer word ten einde 'n sosiaal-verantwoordbare en regverdigte uitkoms te bewerkstellig. Dit kan, soos daar deur Chapelle *et al.* (2009:23) opgemerk word, slegs gebeur wanneer “a mass of heterogeneous evidence becomes comprehensible

... through the flow of narrative, a story as it were.” Dit is wat in die vorige hoofstuk ’n sistematiese argument wat uit ’n verskeidenheid gegewens en datastelle opgebou is, genoem is.

Daar word gevolglik gepoog om in die volgende afdeling van hierdie hoofstuk die storie van TAG by wyse van spreke klaar te vertel deur ’n versoerende perspektief te bied op die interpretasie van die resultate van TAG (en by implikasie TALL) en die bruikbaarheid en toepaslikheid daarvan binne die institusionele konteks(te) waarin dit aangewend word.

8.3 ’n Versoerende perspektief op die resultate van TAG

Uit die empiriese gegewens van hierdie studie blyk dit, soos reeds hierbo gesuggereer, dat TAG ’n meetinstrument is wat die toets van die tyd deurstaan het. Benewens dit kan daar ook ’n teoretiese regverdiging vir die konstruk daarvan aangebied word. Die vraag wat egter nou opduik, is wat met die resultate van TAG gedoen word, omdat die beste gegewens en teoretiese regverdiging nie noodwendig tot die etiese gebruik daarvan sal bydra nie. Die antwoord hierop is dat daar op ’n uiters omsigtige en genuanseerde wyse daarmee omgegaan word. Die bespreking hieronder is ’n poging om dit te illustreer.

8.3.1 ’n Onderbou van tegnies-teoretiese verdedigbaarheid

Die teoretiese verdedigbaarheid van ’n toets is die hoofkwessie in die res van hierdie hoofstuk wat telkens sal verwys na pogings wat aangewend is om TAG en TALL, en die interpretasie van die resultate daarvan, se teoretiese regverdiging so duidelik as moontlik te artikuleer.

Daar is byvoorbeeld reeds elders in hierdie studie melding daarvan gemaak dat daar vanuit die staanspoor gepoog is om ’n basis te skep wat daartoe sal lei dat daar verantwoordelik met TAG se resultate omgegaan kan word, ten einde tot tegnies-wetenskaplike gevolgtrekkings te kan kom, wat verdedigbaar is in die lig van teoretiese gegewens. Selfs die besluit dat TAG en TALL hoofsaaklik uit

meervoudige keuse items moes bestaan, was uit 'n tegniese hoek gesien, dus nie lukraak nie, maar oorwoë. Benewens die bruikbaarheidsvoordele hieraan verbonde, word die problematiese kwessie van inter-betroubaarheid tussen beoordelaars uit die weg geruim en alle studente word in hierdie opsig derhalwe regverdig behandel. Daar is inderdaad deur hierdie besluit met 'n mate van die gesigsgeldigheid van die toetse weggedoen, maar na 'n ondersoek is daar byvoorbeeld gevind dat die skryfafdeling van die toets (wat aanvanklik ter wille van gesigsgeldigheid ingesluit is) weggelaat kan word, omdat die meervoudige keuse items alleen voldoende is om uitsprake oor die toets en die vermoëns van studente te kon maak. Hierdie toegewing het met verloop van tyd soveel positiewe gevolge gehad, dat die invloed daarvan op die gesigsgeldigheid van die toets vandag nie meer 'n besondere kwessie is nie.

Verder dra die feit dat sekere eise in die vorm van statistiese parameters gestel is, en waaraan alle items moet voldoen, voordat dit by 'n finale produk ingesluit word, ook daartoe by dat die toets eerstens verdedigbaar is, en tweedens dat die resultate daarvan regverdig gebruik sal word. Daar sal in die res van hierdie hoofstuk ook gesien kan word dat tegniese oorwoënheid 'n deurlopende tema is – die keuse van tegnieke dra by tot die verkryging van objektiewe resultate wat die subjektiewe interpretasie van 'n toetsresultaat meer geregverdig maak.

8.3.2 Bepaling en hantering van afsnypunte

Hoewel die kwessie van afsnypunte in Hoofstuk 5 aan die orde gekom het, is dit net ten dele bespreek en 'n meer geartikuleerde omskrywing blyk dus hier in orde te wees. Die afsnypunte van TAG en TALL vir die Universiteit van Pretoria is aanvanklik bepaal in lyn met die argumente wat vir die ELSA-Plus se afsnypunte aangevoer is, naamlik dat ongeveer die onderste derde van die populasie as risikokandidate uitgewys moet word, of anders gestel die toets moet druipe. Hoewel daar 'n ongemak met hierdie redenasie was, het die anekdotiese getuienis van deskundiges soos taaldosente tog bevestig dat die afsnypunt min of meer korrek was. Die getuienis het behels dat daardie studente wat as risikokandidate deur die

toets aangemerkt is inderdaad in taalondersteuningskursusse tuis hoort en dat daardie studente wat as kandidate met lae tot geen risiko aangemerkt is se akademiese taalvermoë beslis op 'n hoër vlak was as die van die ander studente.

Daar is egter besluit om ter wille van tegniese verfyning, en in aansluiting by die vorige punt, met wiskundige modellering 'n meer akkurate afsnypunt te bepaal wat gebaseer was op historiese data, naamlik dat ongeveer die onderste derde steeds as risikokandidate uitgewys moet word. Dit het daarop neergekom dat die afsnypunt vir beide TAG en TALL altyd op ongeveer 10% onder die gemiddelde prestasie van die betrokke toets was. Dit is as volg geartikuleer: 'n student toon risiko wanneer sy / haar prestasie op die toets ongeveer 10% onder die gemiddeld van sy / haar portuurgroep is. Verder is daar voorsiening vir potensiële misklassifikasies en grensgevalle gemaak, wat hieronder aan bod sal kom.

Hierdie afsnypunt is egter met verloop van tyd sodanig verfyn dat dit tans wiskundig as volg uitgedruk word: 'Cut point = (mean - 0.6 σ)'. Hierdie berekening van die afsnypunt word egter nie rigied toegepas of afgedwing nie en verskil van instelling tot instelling, omdat die kontekste verskil. So verkies een instelling byvoorbeeld om hierdie formule te gebruik en 'n ander neem wel daarvan kennis, maar kan byvoorbeeld net 'n beperkte aantal studente op 'n bepaalde kursus akkommodeer en pas die afsnypunt daarvolgens aan. Logistiese vermoë en institusionele kapasiteit noodsaak dus verstellings aan die meer algemene aanduidings van waar die afsnypunt moet lê.

Omdat die toetse vir plasing gebruik word, was dit egter ook belangrik om die resultaat fyner uit te druk as bloot 'n aanduiding van risiko of nie, soos die geval met die ELSA-Plus was (wat natuurlik tot stigmatisering bygedra het). Dit word in meer besonderhede hieronder bespreek.

Wat hier van belang is, is dat alle instellings waar TAG afgelê word min of meer dieselfde vetrekpunt vir die bepaling van afsnypunte gebruik. Verder word die parameters vir grensgevalle so wyd moontlik getrek om studente so ver moontlik

die voordeel van die twyfel te gee - dit word in meer besonderhede by die volgende punt bespreek.

8.3.3 Potensiële misklassifikasie van grensgevalle

Potensiële misklassifikasies en die hantering van grensgevalle is, in teenstelling met bostaande punt oor afsnypunte, uitvoerig in Hoofstuk 5 bespreek en dit word daarom nie weer hier herhaal nie. Wat hier egter wel van belang is, is dat daarop gelet moet word dat daar op 'n uiters verantwoordbare en tegnies-gefundeerde wyse met daardie studente wat nét bo of nét onder die afsnypunt lê, omgegaan word. Aspekte soos die waargenome telling teenoor die werklike telling en die metingsfout word byvoorbeeld hier in ag geneem – Hoofstuk 5 kan vir meer besonderhede hieroor geraadpleeg word.

Misklassifikasies by TAG en TALL lê in die reël tussen 0.15 en 0.29 om die gemiddeld, wat beteken dat daar in praktyk aan almal wat tussen 0.075 en 0.145 standaardafwykings van die afsnypunt lê, 'n tweede geleentheid gebied word om 'n toets te herskryf.

Die verdiskontering van bogenoemde bring mee dat TAG en TALL se resultate op 'n verantwoordbare wyse gebruik word alvorens studente in 'n bepaalde kategorie of risikoband geplaas word.

8.3.4 Kategorisering van toetslinge

Die kategorisering van studente in vyf risikobande het bygedra daartoe dat die toetsresultate minder gestigmatiseer was, omdat dit nie meer 'n slaag- / druip-indruk geskep het nie. Daar is verder ook nie meer na 'n enkele resultaat as afsnypunt verwys nie, wat aanvanklik problematies was, omdat die afsnypunte van jaar tot jaar verskil het. Die bekendmaking van resultate het ook bygedra tot nog beter plasing van studente op gepaste ondersteuningskursusse. Die berekening van die afsnypunte vir die onderstaande risikobande is reeds in Hoofstuk 5 bespreek en word nie weer hier herhaal nie. Daar moet wel opgelet word dat die poging om toetsresultate sosiaal meer aanvaarbaar te maak deur dit in bande bekend te maak

besonder positiewe gevolge onder studente en universiteitsowerhede tot gevolg gehad het. Die bande sien as volg daar uit:

Band 1	Uiters hoë risiko
Band 2	Hoë risiko
Band 3	Risiko (word ook gebruik om grensgevalle te identifiseer)
Band 4	Lae risiko
Band 5	Lae tot geen risiko nie

8.3.5 Gevolge vir toetslinge

Die gevolge van toetsresultate op studente word reeds deels deur die vorige punt gesuggereer. Dit kom naamlik daarop neer dat studente op gepaste taalondersteuningskursusse geplaas of minstens daarvoor aanbeveel word, soos die geval by die Universiteit van Pretoria blyk te wees. Tot en met 2009 is studente aan die Noordwes Universiteit óf verplig om 'n taalondersteuningskursus te volg, óf daarvan vrygestel. Vanaf 2010 word alle studente by hierdie instelling egter verplig om die een of ander vorm van taalsteun te volg. Die model wat by die Universiteit Stellenbosch tans stelselmatig, en in 'n mindere of 'n meerdere mate, geïmplementeer word en myns insiens prysenswaardig is, is een van verpligte taalontwikkeling (die verwerwing van een van die amptelike tale van die Wes Kaap), én een van verpligte akademiese taalvaardigheidsontwikkeling in een van die twee onderrigtale van dié universiteit.

Voorts is daar by minstens twee van hierdie universiteite ook die geleentheid vir toppresterders gebied om hulself, na 'n semester se bywoning van die toepaslike kursus, te kan vrystel van verdere klasbywoning deur middel van hul prestasies op andersoortige assesserings as toetse wat deel uitmaak van die kursus. By al drie bogenoemde instellings is die kursusse tans boonop deel van die kredietdraende programaanbod en dus nie 'n addisionele las op enige student nie.

Uit die bostaande behoort dit duidelik te wees dat daar gepoog word om die moontlike nadelige gevolge van die toetsresultate so min as moontlik te maak.

Verder word die aanvanklik stigmatiserende aard van die resultate deesdae ook tot die minimum beperk.

8.3.6 Deursigtigheid

Daar word deurlopend gepoog om ten volle deursigtig te wees oor die onderbou van, die ondersoek na, en die gebruik en interpretasie van TAG. Dit word gedoen deur kontinu navorsing te onderneem oor sake wat met die toetse verband hou. Die uitkomst van die navorsing word in die publieke domein geplaas deur dit in eweknie-geëvalueerde vaktydskrifte te publiseer (daar het byvoorbeeld al meer as 14 artikels oor TAG en TALL die lig gesien), en hierdie studie is 'n voorbeeld van een van twee omvattende nagraadse ondersoek wat tans op doktrale vlak daarvoor onderneem word. Verder word daar ook op 'n deurlopende basis verslag aan universiteitsowerhede oor die toetse en die potensiële effekte daarvan gelewer ten einde ingeligte besluitneming tot voordeel van alle belanghebbendes te bevorder. Inligting oor die toetse word ook vooraf aan toetslinge beskikbaar gestel deur van die elektroniese media (webwerwe, e-posse en sms'e) en die gedrukte media (brosjures, inligtingstukke en briewe) gebruik te maak – hier hang die presiese aard en inkleding daarvan van instelling tot instelling af. Soveel as moontlik pogings tot groter deursigtigheid dra by tot die sosiale regverdiging van die toets, maar soos die ondersoek van Van der Walt en Steyn (2007:149-151) aangetoon het, kan daar steeds meer gedoen word om dit te bevorder.

8.3.7 Koste

In Hoofstuk 1 is gesuggerer dat die gebruik van die ELSA-plus op 'n gegewe moment onbekostigbaar sou word en dat ander moontlikhede ondersoek sou moes word, sodat daar groter beheer oor die koste uitgeoefen kon word. Op daardie tyd was die ontwikkeling van 'n nuwe toets die beste opsie, veral aangesien dit nie net die koste-element was wat ongemak veroorsaak het nie (verwys Hoofstuk 1 vir 'n volledige bespreking). Die gevolge van die besluit wat geneem is, is dat die koste vir die ontwikkeling, implementering en onderhoud van TAG en TALL besonder laag is - dit kos tans maar ongeveer R10/toets. Hierdie faktor bring ook

mee dat die koste wat van studente vir hierdie toetse verhaal moet word óf minimaal óf nie van toepassing is nie. Verder dra die feit dat TAG en TALL tans onder die ICELDA-samewerkingsooreenkoms bedryf word daartoe by dat koste-implikasies versprei is tussen vier instellings.

8.3.8 Bruikbaarheid

Soos ook in Hoofstuk 1 genoem was die bruikbaarheid van die ELSA-Plus problematies en 'n toets moes derhalwe ontwerp word wat as't ware meer gebruikersvriendelik was. Daar moet egter ook melding van die doeltreffendheid waarmee TAG en TALL gebruik word, gemaak word: so word gemaklik 10 000 studente op 'n dag getoets met resultate wat binne twee dae, en soms binne 18 uur, daarna geverifieer en beskikbaar is vir interpretasie. Verder is die stelsels van die universiteite reeds sodanig aangepas dat die toetsuitslag onmiddellik die nodige effekte vir byvoorbeeld registrasie het. Dit dra alles by tot die ekonomiese volhoubaarheid van die toets.

8.4 Samevatting

In hierdie hoofstuk is daar gepoog om die storie van TAG (en TALL) af te sluit deur in aansluiting by die res van die studie te toon hoe daar in praktyk met TAG en TALL omgegaan word om die gevolge van die interpretasie van die resultate so laag as moontlik te hou. Daar is op direkte en indirekte gevolge gefokus deur agt kwessies op die tafel te plaas wat elkeen kortliks bespreek is. Direkte gevolge verwys na daardie gevolge wat 'n onmiddellike en 'n direkte effek op die toetsling kan hê, terwyl indirekte gevolge verwys na die geïmpliseerde aard van die gebruik van die toetse en die interpretasie van die resultate daarvan.

Daar is derhalwe gepoog om in hierdie hoofstuk op versoerende wyse om te gaan met die inligting in spesifieke hoofstukke 3, 5, en 7 wat daarop dui dat TAG 'n bruikbare en geldige toets is wat doen wat dit veronderstel is om te doen, naamlik om 'n aanduiding van die akademiese geletterdheidsvlakke van eerstejaarstudente te gee, sodat hulle onder andere op gepaste taalondersteuningskursusse geplaas kan

word. Om die prentjie dus af te rond is daar, in aansluiting by bogenoemde hoofstukke, gepoog om aan te toon hoe 'n oorwoë en 'n tegnies-verantwoordbare ingesteldheid daartoe bydra dat TAG en TALL, en die interpretasie van die resultate daarvan, op regverdige en verantwoordelike wyse gebruik word. Hierdie hoofstuk word vervolgens afgesluit met die volgende woorde van Chapelle *et al.* (2009:320) wat dit goed opsom: “[t]he research seeks evidence that helps to support intended interpretations and uses with the aim of developing a test whose interpretations and uses can be justified.”

Slot en samevatting

9.1 Inleidend

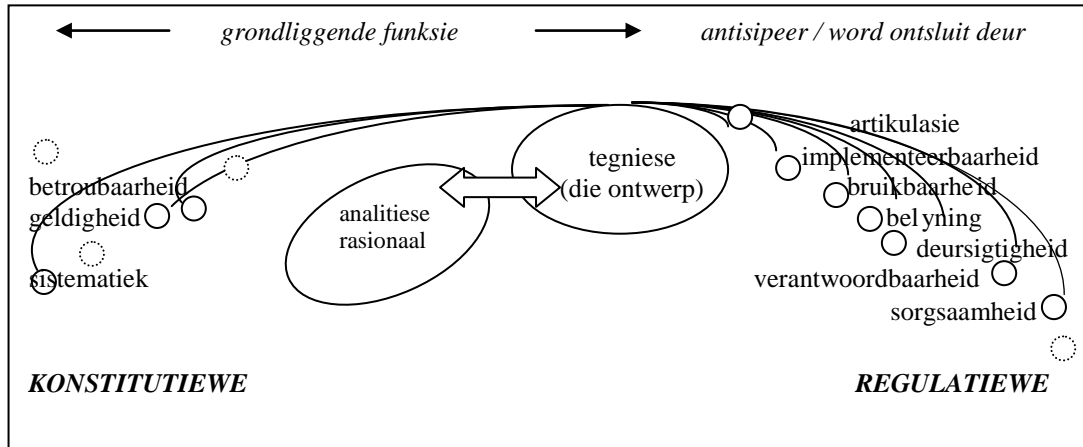
Daar is vroeër in hierdie studie verwys na Shohamy (2001:16) wat aanvoer dat dit uiteindelik tog maar die prestasie op 'n spesifieke toets wat op 'n gegewe moment by 'n spesifieke geleentheid afgeneem is, wat daartoe aanleiding kan gee dat die toetsling verreikende, dikwels onomkeerbare, gevolge in die gesig moet staar. Net so is daar kennis geneem van Weideman (2006:71) wat beweer dat daar altyd 'n oënskynlike dringendheid aan die werk van toegepaste linguïste is sodat hulle selde die tyd neem om deeglik te reflekteer oor die teoretiese onderbou van hul tegniese ontwerpe, sowel as oor die implikasies of gevolge wat die gebruik daarvan kan hê. Die nadraai hiervan is dat daar in die eerste plek soms geen deeglik geformuleerde teoretiese rasionaal vir toegepaste linguïstiese ontwerpe is nie, en ten tweede dat daar nie ten volle sosiale verantwoordings gedoen kan word oor die implikasies wat die gebruik van die artefak inhou nie.

Hierdie studie was in sy geheel 'n poging om terug te staan en 'n spesifieke toegepaste linguïstiese instrument, 'n taaltoets (TAG), in heroënskou te neem en in die proses die storie daarvan (Shohamy, 2001) te vertel. Die doel met hierdie proefskrif was daarom om, teen die agtergrond van die teoretiese raamwerk waarop dié studie berus, 'n perspektief te bied op verantwoordbare en verantwoordelike toetsontwerp en -gebruik binne die institusionele konteks(te) waarin dit aangewend

word, en die regverdigte en billike interpretasie van toetsresultate daarvoor. Die raamwerk, wat die studie as geheel ondersteun, is deurlopend van toepassing gemaak op TAG (en ter illustrasie soms ook op TALL). Daar is gepoog om aan die hand van die besprekings en bevindinge van veral Hoofstukke 3, 5, 6, 7, en 8 artikulasie te gee aan die navorsing wat oor die afgelope aantal jaar oor TAG en TALL gedoen is, maar sonder om enigsins te suggereer dat hierdie studie die finale woord kan of gaan spreek oor die verantwoordbaarheid of die regverdiging daarvan. Dit is bloot 'n aanvanklike en voorlopige oordeel wat met verloop van tyd deur verdere navorsing aangevul en verder gevoer sal moet word, 'n punt waartoe hieronder by afdeling 9.3 teruggekeer word.

TAG is op sistematiese wyse ondersoek en daar is in lyn met die kenmerke van die sesde generasie van die toegepaste taalkunde gepoog om 'n verskeidenheid perspektiewe in die proses van ondersoek te akkommodeer, hoewel die bestek van die studie 'n beperkende invloed daarop gehad het. Die uitdaging, om Weideman (2003d:10) se woorde te mag omvorm, was egter nie soseer in die erkenning van die veelheid van perspektiewe opgesluit nie, maar eerder in die ondersoek na die ooreenstemmende en oorvleuelende kenmerke daarvan en die interaksie wat dit met mekaar het. Dit bring ons presies by die punt van die teoretiese raamwerk waarop hierdie proefskrif gebou is, naamlik dat daar voortdurend terugskouend én voorspellend te werk gegaan moet word wanneer 'n toegepaste taalkundige oplossing (soos TAG en TALL) vir 'n taalprobleem voorgehou word: antisipatories in die opsig dat die toegepaste taalkundige oplossings vir taalprobleme byvoorbeeld die moontlike gevolge van die gebruik daarvan in ag moet neem (as deel van die leidende of regulatiewe aard daarvan) en terugskouend in die opsig dat daar retrospektief 'n teoretiese begroning moet wees om op terug te val (as deel van die basis, of konstitutiewe aard daarvan). Met so 'n benadering word tegniese-konstitutiewe begrippe soos betroubaarheid en geldigheid aan regulatiewe idees, soos verantwoordbaarheid en sorgsaamheid verbind. Artikulasie tussen die grondliggende funksies (van konstitutiewe aard) van 'n toets, en die ontslote regulatiewe aard daarvan, is dus van belang en word goed geïllustreer deur Figuur

9.1 wat reeds elders in hierdie studie aangebied is, maar wat duidelikshalwe weer hier herhaal word. Hierdie figuur is oorgeneem en verwerk uit Weideman (2009a:20).



Figuur 9.1: Illustrasie van artikulasie tussen konstitutiewe en regulatiewe elemente

Wat volg is 'n verbesonderde opsomming van dié studie en die gevolgtrekkings waartoe gekom is. Dit word opgevolg deur enkele aanbevelings vir verdere navorsing.

9.2 Opsomming en gevolgtrekkings van die studie

Hoofstuk 1 het die agtergrond tot die studie verskaf en 'n uiteensetting van die rasionaal daarvoor en die uiteindelijke probleem wat ondersoek is, gebied. Daar is aangevoer dat universiteite 'n behoefte het aan onder andere akademiese geletterdheidstoetse ten einde ingeligte besluite oor studente se akademiese taalbehoefte en die nodige ondersteuning te kan neem. TAG is as so 'n toegepaste linguistiese artefak voorgehou en daar is daarop gewys dat daar behoefte is aan 'n gesistematiseerde ondersoek na veral die teoretiese verdedigbaarheid daarvan.

In Hoofstuk 2 is die teoretiese raamwerk waarteen die res van die studie gekontekstualiseer is, aangebied. Daar is begin deur die toegepaste linguistiek as dissipline te definieer deur na drie tradisies en / of sewe generasies daarvan te verwys. Dit is opgevolg deur 'n bespreking van die fases waardeur toegepaste

linguistiese ontwerpe (as oplossings vir taalprobleme) gaan. Die hoofstuk is afgesluit met 'n argument vir die toegepaste linguistiek as dissipline van ontwerp en daar is 'n raamwerk daarvoor voorgehou wat in die inleidende deel van hierdie samevattende hoofstuk reeds ter sprake gekom het en daarom nie hier verder bespreek sal word nie.

Hoofstuk 3 het die onderbou van TAG (die teoretiese verdedigbaarheid daarvan) in besonderhede bespreek. 'n Historiese oorsig oor taaltoetsontwikkeling is aangebied en verskillende benaderings tot die talige aspek van akademiese geletterdheid is bespreek. Daar is verder gefokus op die redes waarom daar 'n nuwe konstruksie vir TAG en TALL ontwikkel moes word en hoe die keuse van konstruksie uiteindelik geregtig kan word in die lig van verskillende perspektiewe tot die toegepaste linguistiek in die algemeen en taaltoetsing in die besonder. 'n Konstruksie vir TAG en TALL is uiteindelik voorgehou.

Hoofstuk 4 het voortgebou op Hoofstuk 3 in die opsig dat dit een van die groot uitdagings van toetsontwerp aanspreek, naamlik die belyning van die konstruksie met die spesifikasies wat uiteindelik as taaktipes en items geoperasionaliseer word. 'n Voorbeeld van hoe 'n spesifieke taaktipe omvorm is tot een wat hoë produktiwiteit toon en waarvan die belyning met die spesifikasies en konstruksie goed is, is aangebied. Aan die hand van die teoretiese raamwerk waarop die studie gebaseer is, het verskeie ontwerpbeginsels vir taaltoetse in hierdie hoofstuk aan die orde gekom.

Een van die fundamentele oorwegings in toetsontwerp, te wete betroubaarheid ('n konstitutiewe begrip), is in Hoofstuk 5 uitvoerig bespreek. Betroubaarheid verwys na die konsistensie en stabiliteit van toetsresultate en in hierdie hoofstuk is verskillende vorme van betroubaarheid van TAG ondersoek aan die hand van enkele toepaslike statistiese bewerkings. Aspekte wat 'n invloed op die betroubaarheid van 'n toets kan hê is ook uitgewys en beskryf en daar is opmerkings oor die veralgemeenbaarheid van toetsresultate gemaak. In hierdie

hoofstuk is daar tot die gevolgtrekking gekom dat TAG 'n uitsonderlik betroubare toets is.

In aansluiting by die vorige hoofstuk, het Hoofstuk 6 gekonsentreer op die kwessie van geldigheid wat naas betroubaarheid as nog 'n fundamentele oorweging in toetsontwerp geag word. Hierdie begrip is vanuit drie perspektiewe beskou, maar die kern daarvan lê in die bespreking van die verruimde standpunt daaroor, naamlik dat dit terselfdertyd 'n attribuut van 'n toets kan wees en as konstitutiewe begrip ontsluit kan word deur verdere (regulatiewe) idees met betrekking tot taaltoetsing. Daar is verder tot die gevolgtrekking gekom dat daar beslis 'n argument te voer is vir onderskeibare konsepte soos nut, regverdiging, betroubaarheid en geldigheid, hoewel hierdie konsepte ook by mekaar aansluit in die opsig dat dit die belang van konstruktiewe geldigheid (die teoretiese verdedigbaarheid van die toets) weer na vore laat tree.

Die vorige hoofstuk was teoreties-inleidend tot Hoofstuk 7 wat gefokus het op toetsbekragtiging of validering in dié sin dat dit verskillende soorte geldigheid en gelyklopende bekragtigingsprosesse ondersoek het ten einde vas te stel of TAG werklik meet wat dit voorgee om te meet, naamlik akademiese geletterdheid. Daar is hoofsaaklik van korrelasiestudies vir die analise van data in hierdie hoofstuk gebruik gemaak. Die interne en eksterne geldigheid van TAG is bepaal en op die wyse is daar aan die hand van 'n sistematiese hantering van die proses van validering 'n geldighedsargument vir dié toets opgebou. Daar is gepoog om die geldighedsargument tegelykertyd eenvoudig, koherent, omvattend en bewysbaar te hou. In hierdie hoofstuk is daar finaal tot die gevolgtrekking gekom dat TAG 'n geldige toets is.

Hoofstuk 8 was 'n poging om die storie van TAG af te sluit deur 'n versoberende perspektief te bied op die interpretasie van die resultate, en die bruikbaarheid en toepaslikheid daarvan in gegewe kontekste. Dit is gedoen in noue aansluiting by die vorige drie hoofstukke deur aan te toon dat 'n oorwoë en 'n tegniese-

verantwoordbare ingesteldheid nodig is om TAG op regverdige en verantwoordelike wyse te kan gebruik.

9.3 Verdere navorsingsmoontlikhede

Hierdie studie het nog lank nie die laaste woord oor TAG en TALL gespreek nie en daar is talle aspekte wat verder ondersoek kan word. Daar is derhalwe, en op deurlopende wyse, opmerkings oor verdere navorsingsmoontlikhede gemaak, waarvan die belangrikste hieronder gelys word:

Dit wat in Hoofstuk 1 as die probleemstelling en hipotese van hierdie studie uiteengesit is, sou byvoorbeeld ook vanuit ander perspektiewe benader en met behulp van 'n ander metodiek ondersoek kon word.

Die teoretiese raamwerk wat in Hoofstuk 2 voorgehou is, kan ook verder ondersoek en uitgebou word - daar is reeds verwys na minstens een ander doktrale studie wat hiermee gemoeid is. Die bruikbaarheid van die raamwerk op ander toegepaste linguistiese artefakte sou ook ondersoek kon word. Verder sou die keuse van juis hierdie raamwerk as onderbou van hierdie studie bevraagteken kon word en dieselfde studie sou dus vanuit 'n ander teoretiese perspektief benader kon word.

Hoofstuk 3 leen homself tot 'n veelheid van verdere navorsingsmoontlikhede. So kan daar byvoorbeeld gevra word of die konstruk van akademiese geletterdheid waarop uiteindelik besluit is, voldoende en omvattend genoeg is. Die onderliggende taalteorieë waarop daar gefokus is, sou ook bevraagteken kon word en daar sou uiteindelik 'n heel ander konstruk vir die toetsing van akademiese geletterdheid ontwikkel kon word. Daar is verder reeds verskeie studies oor die invloed en rol van identiteit op akademiese geletterdheid onderneem en daar sou ondersoek hierna binne 'n taaltoetskonteks gedoen kon word.

In Hoofstuk 4 word daar melding van gemaak dat ondersoek na die gewigte per vraag of afdeling in TAG moontlik nodig is ten einde 'n toets met 'n geregverdigde en gebalanseerde hoeveelheid items te produseer. Die doeltreffendheid van die

verskillende taak- en itemtipes van TAG sou ook ondersoek kon word, en daar sou met spesifieke taaktipes (soos in hierdie hoofstuk aangetoon is) geëksperimenteer kon word ten einde 'n toets ter tafel te lê wat produktiewe items met hoë belyning met die konstruk insluit. Die invloed van die demografiese samestelling van die onderskeie instellings op die diskriminasiewaardes en die moeilikheidsgraad van die toets sou ook nagevors kon word. In Hoofstuk 4 is daar ook aangedui dat die keuse van tekste en die wyse waarop vrae geformuleer word, verder ondersoek kon word.

Hoofstuk 5 bied 'n legio moontlikhede vir verdere navorsing, veral omdat dit ten dele empiries van aard is. Replikasiestudies op ander populasies en in ander omstandighede sou byvoorbeeld gedoen kon word om dit wat in hierdie hoofstuk aan bod gekom het te bevestig of te weerlê. 'n Aspek wat egter dringend verdere ondersoek vereis, is die veralgemeenbaarheid van TAG en TALL se toetsresultate, omdat dit 'n invloed op die betroubaarheid (en uiteindelik geldigheid ook) van TAG en TALL kan hê.

Hoofstuk 6 sou konseptueel vanuit 'n ander perspektief benader kon word en dit staan lesers van hierdie proefskrif vry om vir of teen die standpunt wat in hierdie hoofstuk ingeneem is te argumenteer. 'n Aspek wat wel aandag behoort te geniet, is die kwessie van interpretasie, omdat daar eerste blyke in die jongste literatuur is dat alles nie meer onder konstrukgeldigheid verenig hoef te word nie, maar wel onder interpretasie. Daar is in hierdie studie suggesties gemaak dat ook hierdie benadering nie korrek sou wees nie en dat daar liefst steeds 'n duidelike onderskeid tussen tegniese begrippe wat toetsing onderlê, gemaak moet word. Daar het egter nog te min inligting hieroor verskyn om gefundeerde uitsprake daarvoor te kan maak.

Hoofstuk 7 sou net soos Hoofstuk 5 anders benader kon word en bied ook 'n verskeidenheid van verdere navorsingsmoontlikhede. Replikasiestudies sou hier ook aan die orde kon gewees het. 'n Aspek wat moontlik meer aandag kan geniet, is die kwessie van hoe studente se persepsies oor die toets, en die gepaardgaande intervensie, oor 'n jaar verander. Net so kan verdere ondersoeke na die verhoging

van die deursigtigheid van TAG ook gedoen word. Aspekte wat tot konstruk-irrelevante variansie bydra, sou ook verder ondersoek kon word.

Soos reeds genoem behoort daar verdere ondersoek na spesifiek die kwessie van die interpretasie van toetsresultate te wees. Aangesien die fokus van hierdie studie op die konstitutiewe kant van die teoretiese raamwerk wat daarvoor gebruik is fokus, is daar nie in besonderhede aandag gegee aan die regulatiewe aspekte van die raamwerk nie. Soos genoem, is daar reeds 'n doktorsale studie hieroor onderweg. Hoofstuk 8 het wel enkele kwessies aangeroer wat hiermee verband hou en verder ondersoek sou kon word.

9.4 Ten slotte

Alhoewel die moontlikhede vir verdere navorsing op 'n hele aantal aspekte wys wat nog nagevors kan word en daar selfs die vermoede sou kon bestaan dat hierdie studie ten dele oop is vir kritiek, is die slotsom waartoe hierdie studie uiteindelik kom, naamlik dat TAG 'n toets van hoë kwaliteit is, eintlik nie te vermy nie. In daardie sin dien dit as voorbeeld van hoe akademiese geletterdheidstoetse met sorgsaamheid gemaak kan, en behoort te, word.

Die planne wat toegepaste taalkundiges maak om taalprobleme uit die weg te probeer ruim, gee blyke daarvan dat sulke tegniese ontwerpe nooit selfgenoegsaam is en by wyse van spreke op eie voete kan staan nie, maar dat dit eerder werk is wat diensbaar bly aan – soos in hierdie geval – groterwordende getalle mense. Toegepaste taalkundige ontwerpe is dus altyd en noodwendig afhanklik van die mate waartoe dit verantwoordelik gebruik kan word. In die woorde van Weideman (2007a:602):

Applied linguistic designs therefore find their meaning in the *service* (or *disservice*) that they will perform for other human beings ... [T]he care with which designs are made points to the love that we show for humanity. This love is evident even in the technical artefacts that we create.

BIBLIOGRAFIE

- Alderson, C.; Clapham, C.; Wall, D. 1995. *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Alternative Admissions Research Project (AARP). S.a. PTEEP (Placement Test in English for Educational Purposes). Kaapstad: University of Cape Town.
- American Educational Research Association, American Psychological Association, National Council for Measurement in Education. 1999. *Standards for educational and psychological testing*. Washington, DC: Educational Testing Services.
- Angoff, W.H. 1988. Validity: an evolving concept. Pp. 19–32 in Wainer, H. & Braun, I.H. (eds.) 1988.
- Arce-Ferrer, A.J. & Castillo, I.B. 2006. Investigating postgraduate college admission interviews: Generalisability theory, reliability and incremental predictive validity. *Journal of Hispanic Higher Education*, 6(2):118-134.
- Astin & Oseguera. 2002. *Degree attainment rates at American colleges and universities*. Los Angeles: Higher Education Research Institute, UCLA.
- Babbie, E. & Mouton, J. 2006. *The practice of social research*. Oxford: Oxford University Press.
- Bachman, L.F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F. 2005. *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Bachman, L.F. & Palmer, A.S. 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

- Bargate, K. 1999. Mathematics as an indicator of success in first year accounting programmes at Technikon Natal. *South African Journal of Higher Education*, 13(1):139-143.
- Baynham, M. & Prinsloo, M. 2001. New directions in literacy research. *Language and Education*, 15(2&3):83-91.
- Bencze, J.L. 2005. Constructivism. [Intyds]. Beskikbaar: <http://leo.oise.utoronto.ca/~lbencze/Constructivism.html>. Besoekop 9 Desember 2005.
- Blacqui re, A. 1989. Reading for survival: Text and the second language student. *South African Journal for Higher Education*, 3(1):73-82.
- Blanckenberg, H.C. 1999. *Die onderrig van Afrikaans as alternatiewe taal in 'n uitkomsgebaseerde onderwysbenadering*. Ongepubliseerde VDO-Studiegids. Johannesburg: Nasionale Private Kolleges.
- Blanton, L.L. 1994. Discourse, artefacts and the Ozarks: Understanding academic literacy. *Journal of Second Language Writing*, 3(1):1-16. Herdruk (as Hoofstuk 17: 219-235) in Zamel, V. & Spack, R. (reds.), 1998.
- Bourdieu, P., Passeron, J. & De Saint Martin, M. 1994. *Academic discourse: Linguistic misunderstanding and professorial power*. Cambridge: Polity Press.
- Borsboom, D., Mellenbergh, G.J. & Van Heerden, J. 2004. The concept of validity. *Psychological Review*, 111(4):1061-1071.
- Botha, H.L. & Cilliers, C.D. 1999. Preparedness for university study: Designing a thinking skills test. *South African Journal of Higher Education*, 13(1):144-152.
- Botha, H.L., Du Plessis, S.A. & Menkveld, H. 2007a. *'n Ontleding van deurvloeikoerse aan die US – 'n verslag deur die Tredhousisteam*. Stellenbosch: Universiteit Stellenbosch.

- Botha, H.L., Du Plessis, S.A. & Menkveld, H. 2007b. *'n Profiel van die 1997-2006-kohorte eerstejaarstudente – 'n verslag deur die Tredhousisteem*. Stellenbosch: Universiteit Stellenbosch.
- Boughey, C. 2000. Multiple metaphors in an understanding of academic literacy. *Teachers and Teaching: Theory and practice*, 6(3):279-290.
- Boughey, C. 2002. Naming students' problems: an analysis of language-related discourses at a South African university. *Teaching in Higher Education*, 7(3): 295-307.
- Boyuwoye, O. 2002. Stressful experiences of first year students of selected universities in South Africa. *Counselling Psychology Quarterly*, 15(3):277-290.
- Brindley, G. 2002. Issues in language assessment. Pp.459-470 in Kaplan, R.B. (red.), 2002.
- Brown, H.D. 2004. *Language assessment: Principles and classroom practice*. New York: Pearson Education.
- Brown, J.D. 1996. *Testing in language programmes*. Upper Saddle River, New Jersey: Prentice Hall Regents.
- Brown, J.D. 2004. Research methods for applied linguistics: Scope, characteristics, and standards. Pp.476-500 in Davies, A. & Elder, C. (reds.), 2004.
- Brumfit, C. 1997. How applied linguistics is the same as any other science. *International Journal of Applied Linguistics*, 7(1):86-94.
- Bubb, E.S. 1991. *The adaptation of coloured first year students at an Afrikaans university*. Ongepubliseerde magisterverhandeling. Johannesburg: Randse Afrikaanse Universiteit.
- Butler, H.G. & Van Dyk, T.J. 2004. An academic English language intervention for first year engineering students. *South African Journal of Linguistics*, 22(1&2):1-8.

- Canale, M. 1983. From communicative competence to communicative language pedagogy. Pp.2-27 in Richards, J.C. & Schmidt, R.W. (reds.), 1983.
- Canale, M. & Swain, M. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics*, 1(1):1-47.
- Carrol, B.J. 1980. *Testing communicative performance: An interim study*. Oxford: Pergamon Press.
- Carrol, J.D. 1961. Fundamental considerations in testing for English language proficiency of foreign students. *Testing*. Washington: Center for Applied Linguistics.
- Cattell, R.B. 1946. *Description and measurement in personality*. New York: World Book Company.
- Cattell, R.B. 1966. The scree test for the number of factors. *Multivariate behavioural research*, 1: 245-276.
- Chapelle, C.A. 1994. Are C-tests valid measures for L2 vocabulary research? *Second Language Research*, 2(10): 157-187.
- Chapelle, C.A. 1999. Validity in language assessment. *Annual review of applied linguistics*, 19:254-272.
- Chapelle, C.A., Enright, M.K. & Jamieson, J.M. 2009. *Building a validity argument for the Test of English as a Foreign Language*. Oxon: Routledge.
- Cito. 2005. *TiaPlus, User's Manual* ©. Arnhem: Cito, M & R Department.
- Clark, J.L.D. 1978. Psychometric considerations in language testing. Pp.15-30 in Spolsky, B. (red.), 1978.
- Clark, J.L.D. 1983. Language testing: Past and current status – directions for the future. *Modern Language Journal*, 67:431-443.

- Cliff, A.F. & Hanslo, M. 2005. The use of 'alternate' assessments as contributors to processes for selecting applicants to health sciences' faculties. Referaat gelewer by die Europe Conference for Medical Education, Amsterdam, Desember 2005.
- Cliff, A.F., Hanslo, M., Ramaboa, K. & Visser, A. 2005. *Third annual report to the Health Sciences Consortium on the use of Health Sciences Placement Tests. AARP Research report.* Kaapstad: University of Cape Town.
- Cliff, A.F., Ramaboa, K. & Pearce, C. 2007. The assessment of entry-level students' academic literacy: does it matter? *Ensovoort*, 11(2): 33-48.
- Cliff, A.F. & Yeld, N. 2006. Test domains and constructs: Academic literacy. Pp. 19-27 in Griesel, H. (red.) *Access and entry level benchmarks: The national benchmark tests project.* Pretoria: Hoër Onderwys, Suid-Afrika.
- Cliff, A.F., Yeld, N. & Hanslo, M. 2003. Assessing the academic literacy skills of entry-level students, using the Placement Test in English for Educational Purposes (PTEEP). Referaat gelewer by die Bi-annual conference of the European Association for Research in Learning and Instruction (EARLI), Padova, Italië.
- Cooper, P. & Van Dyk, T.J. 2003. Measuring vocabulary: A look at different methods of vocabulary testing. *Perspectives in Education*, 21(1): 67-80.
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly*, 34(2): 213-238.
- Cronbach, L.J. 1984. *Essentials of psychological testing.* 4^{de} Uitgawe. New York: Harper & Row.
- Cronbach, L.J. 1988. Five perspectives on validity argument. Pp. 3-17 in Wainer, H. & Braun, I.H. (reds.), 1988.
- Cronbach, L.J. & Meehl, P.E. 1955. Construct validity in psychological tests. *Psychological Bulletin*, 52: 281-302.

- Davidson, F. & Lynch, B.K. 2002. *Testcraft*. New Haven: Yale University Press.
- Davies, A. & Elder, C. (reds.) 2004. *The handbook of applied linguistics*. Victoria, Australië: Blackwell.
- Davies, A. 1968. *Language testing symposium*. Londen: Oxford University Press.
- Davies, A. 1990. *Principles of language testing*. Cambridge: Basil Blackwell.
- Davies, A. 2008. TESOL, applied linguistics, and the butterfly effect. Bydrae tot simposium: Theory in TESOL. *TESOL Quarterly*, 42(2): 296-298.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. & McNamara, T. 1999. *Dictionary of language testing*. (Series editor: Milanovic, M.). Cambridge: Cambridge University Press.
- Davies, A. & Elder, C. (reds.) 2004. *The handbook of applied linguistics*. Victoria, Australië: Blackwell.
- Davies, A. & Elder, C. 2004. Applied linguistics: Subject to discipline. Pp.1-15 in Davies, A. & Elder, C. (reds.) 2004.
- Davies, A. & Elder, C. 2005. Validity and validation in language testing. Pp.795-813 in Hinkel, E. (red.) 2005.
- De Coster, J. 1998. *Overview of factor analysis*. [Intyds]. Beskikbaar: <http://www.stat-help.com/notes.html>. Besoek 16 September 2010.
- Department van Onderwys. 2001. *Education in South Africa: Achievements since 1994*. Pretoria: Department of Education.
- De Zeeuw, J. 1971. *Algemene Psigodiagnostiek I*. Amsterdam: Swets & Zeitlinger.
- De Zeeuw, J. 1978. *Algemene Psigodiagnostiek II*. Amsterdam: Swets & Zeitlinger.
- Douglas, D. 2000. *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.

Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2004. TAG (Toets van Akademiese Geletterdheidsvlakke). Pretoria: Universiteit van Pretoria.

Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2004. TALL (Test of Academic Literacy Levels). Pretoria: Universiteit van Pretoria.

Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2005. TAG (Toets van Akademiese Geletterdheidsvlakke). Pretoria: Universiteit van Pretoria.

Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2005. TALL (Test of Academic Literacy Levels). Pretoria: Universiteit van Pretoria.

Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2006. TAG (Toets van Akademiese Geletterdheidsvlakke). Pretoria: Universiteit van Pretoria.

Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2006. TALL (Test of Academic Literacy Levels). Pretoria: Universiteit van Pretoria.

Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2007. TAG (Toets van Akademiese Geletterdheidsvlakke). Pretoria: Universiteit van Pretoria.

Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2007. TALL (Test of Academic Literacy Levels). Pretoria: Universiteit van Pretoria.

Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2008. TAG (Toets van Akademiese Geletterdheidsvlakke). Pretoria: Universiteit van Pretoria.

Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2008. TALL (Test of Academic Literacy Levels). Pretoria: Universiteit van Pretoria.

- Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2009. TAG (Toets van Akademiese Geletterdheidsvlakke). Pretoria: Universiteit van Pretoria.
- Eenheid vir Akademiese Geletterdheid, Universiteit van Pretoria. 2009. TALL (Test of Academic Literacy Levels). Pretoria: Universiteit van Pretoria.
- Farhady, H. 1979. The disjunctive fallacy between discrete-point and integrative tests. *TESOL Quarterly*, 13(3):347-357.
- Ferreira, J.G. 1992. *Onderrigstrategieë vir die oorbrugging van skool na universiteit*. Ongepubliseerde doktorsale tesis. Pretoria: Universiteit van Suid-Afrika.
- Foxcroft, C.D. & Stumpf, R. 2005. *What is matric for?* Referate en aanbiedinge by die Umalusi en CHET seminaar oor *Matric: What is to be done?* Pretoria, 23 Junie 2005.
- French, J.W. & Michael, W.B. 1966. The nature and meaning of validity and reliability. In Grondlund, N.E. (red.), 1968.
- Frick, B.L. 2007. *Confident or confused? A report on first-year student experience in selected modules on campus*. Plakkaataanbieding by die Universiteit Stellenbosch se eerste Kongres oor die Akademieskap van Onderrig en Leer. Stellenbosch, 22-23 Mei.
- Fulcher, G. 1997. An English language placement test: Issues in reliability and validity. *Language Testing*, 14(2):113-138.
- Fulcher, G. 2010. *Practical language testing*. Londen: Hodder Education.
- Fulcher, G. & Davidson, F. 2007. *Language testing and assessment: an advanced resource book*. Abingdon, Oxon: Routledge.
- Gardner, H. 2000. *The disciplined mind: Beyond facts and standardized tests – the K-12 education that every child deserves*. New York: Penguin Putnam.

- Gee, J.P. 1990. *Social linguistics and literacies: Ideology in discourses*. Londen: Falmer Press.
- Gee, J.P. 1998. What is literacy? Pp. 51-59 in Zamel, V. & Spack, R. (reds.), 1998: Herdruk van die 1987-artikel in *Teaching and Learning: the Journal of Natural Enquiry*, 2:3-11.
- Gee, J.P. 2003. Opportunity to learn: A language-based perspective on assessment. *Assessment in Education*, 10(1):27-46.
- Geisler, C. 1994. *Academic literacy and the nature of expertise: reading, writing and knowing in academic philosophy*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Grabe, W. 2002. Applied linguistics: An emerging discipline for the twenty-first century. Pp.3-12 in Kaplan, R.B. (red.). 2002.
- Graddol, D. (red.), 2001. *Applied linguistics for the 21st century, AILA review*, 14.
- Griesel, H. (red.), 2006. *Access and entry level benchmarks: The national benchmark tests project*. Pretoria: Hoër Onderwys, Suid-Afrika.
- Gronlund, N.E. (red.), 1968. *Readings in measurement and evaluation*. New York: MacMillan.
- Halliday, M.A.K. 1978. *Language as social semiotic*. Londen: Edward Arnold.
- Halliday, M.A.K. 1985. *An introduction to functional grammar*. Londen: Edward Arnold.
- Hambleton, R.K., Swaminathan, H. & Rogers, H.J. 1991. *Fundamentals of item response theory, volume 2*. Newbury Park: Sage Publications.
- Hatch, E. & Faraday, H. 1982. *Research design and statistics for applied linguistics*. Rowley, Mass.: Newbury House.
- Hay, H.R. & Marais, F. 2004. Bridging programmes: Gain, pain or all in vain. *South African Journal of Higher Education*, 18(2):59-75.

- Hinkel, E. (red.) 2005. *Handbook of research on second language teaching and learning*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Horne, T.J. 2002. Seminaar aangebied by die Eenheid vir die Ontwikkeling van Taalvaardigheid, Universiteit van Pretoria. Ongepubliseerde opmerkings en bespreking.
- Hough, B. & Horne, T.J. 2001. ELSA-Plus (English Literacy Skills Assessment vir die tersiêre omgewing). Johannesburg: Hough & Horne.
- Hughes, A. 1989. *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hughes, A. 2003. *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hymes, D.H. 1972. On communicative competence. Pp. 269-293 in Pride, J.B. & Holmes, J. (reds.), 1972.
- Illiterate matrices. 2004. *Natal Witness*, 27 September 2004:8.
- Ingram, E. 1978. The psycholinguistic basis. Pp.1-14 in Spolsky, B. (red.), 1978.
- Inter-institutional Centre for Language Development and Assessment (ICELDA). 2010. TAG (Toets van Akademiese Geletterdheidsvlakke). Stellenbosch: ICELDA.
- Inter-institutional Centre for Language Development and Assessment (ICELDA). 2010. TALL (Test of Academic Literacy Levels). Stellenbosch: ICELDA.
- Ivanič, R. 2004. Discourses of writing and learning to write. *Language and Education*, 18(3):220-245.
- Ivanič, R., Edwards, R., Barton, D., Martin-Jones, M., Fowler, Z., Hughes, B., Mannion, G., Miller, K., Satchwell, C., & Smith, J. 2009. *Improving learning in college: Rethinking literacies across the curriculum*. Oxon: Routledge.

- Jakobovits, L.A. 1970. *Foreign language learning: A psycholinguistic analysis of the issue*. Rowley Massachusetts: Newbury House.
- Jawitz, J. 1995. Performance in first- and second-year engineering at UCT. *South African Journal of Higher Education*, 9(1):101-108.
- Jordan, R.R. 1997. *English for academic purposes. A guide and resource book for teachers*. Cambridge: Cambridge University Press.
- Kane, M.T. 1992. An argument-based approach to validity. *Psychological Bulletin*, 112(3):527-535.
- Kane, M.T., Crooks, T. & Cohen, A. 1999. Validating measures of performance. *Educational measurement: issues and practice*, 18(2): 5-17.
- Kaplan, R.B. (red.). 2002. *The Oxford handbook of applied linguistics*. Oxford: Oxford University Press.
- Kelley, T.L. 1927. *Interpretation of educational measurements*. New York: Macmillan.
- Kern, R. 2000. *Literacy and language teaching*. Oxford: Oxford University Press.
- Klein-Braley, C. 1994. *Language testing with the C-test: A linguistic and statistical investigation into the strategies used by C-test takers, and the prediction of C-test difficulty*. Ongepubliseerde doktrale tesis. Duisburg: Universiteit Duisburg.
- Kohn, A. 2000. *The case against standardized testing*. Westport, CT: Heinemann.
- Kress, G. 1989. *Linguistic processes in sociocultural practice*. Oxford: Oxford University Press.
- Kumaravadivelu, B. 2003. *Beyond methods: Macrostrategies for language teaching*. Londen: Yale University Press.
- Lado, R. 1961a. *Language testing: The construction and use of foreign language tests*. New York: McGraw-Hill.

- Lado, R. 1961b. *Language testing*. Londen: Longman.
- Lane, S. 1999. *Validity evidence for assessment*. [Intyds]. Beskikbaar: http://nciea.org/publications/ValidityEvidence_Lane99.pdf. Besoek 10 November 2009.
- Larsen-Freeman, D. & Cameron, L. 2008. *Complex systems and applied linguistics*. Oxford: Oxford University Press.
- Larsen-Freeman, D. & Weideman, A.J. 2008. *Uncharted territory: Emerging paradigms and the foundations of applied linguistics*. Projekvoorstel vir die “ACLS collaborative research awards”.
- Lea, M.R. & Street, B.V. 1998. Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2):157-173.
- Leibowitz, B. 2001. *Students' prior learning and their acquisition of academic literacy at a multilingual South African university*. Ongepubliseerde doktorsale proefskrif. Sheffield: University of Sheffield.
- Leibowitz, B. & Mohamed, Y. (reds.). 2000. *Routes to writing in Southern Africa*. Cape Town: Silk Road International.
- Le Roux, J. (red.), 1993. *The black child in crisis: A socio-educational perspective*. Pretoria: J.L. van Schaik.
- Lillis, T. 2003. Student writing as ‘academic literacies’: Drawing on Bakhtin to move from critique to design. *Language and Education*, 17(3):192-207.
- Linn, R.L. (red.) 1989. *Educational measurement*. New York: MacMillan.
- Lyman, H.B. 1991. *Test scores & what they mean*. 5^{de} Uitgawe. Englewood Cliffs, New Jersey: Prentice Hall.

- Macdonald, C.A. 1990. *Crossing the threshold into standard three in black education: The consolidated main report of the Threshold Project*. Pretoria: Human Sciences Research Council.
- Madsen, H.S. 1983. *Techniques in testing*. New York: Oxford University Press.
- Magnussen, D. 1966. *Test theory*. Ontario: Don Mills.
- Mamaila, K. 2001. R1,3bn spent on dropout students. *Star*, 15 May 2001: 1.
- Marais, F.C. 2008. *An investigation into the significance of listening proficiency in the assessment of academic literacy levels at Stellenbosch University*. Ongepubliseerde magister verhandeling. Stellenbosch: Universiteit Stellenbosch.
- McKenna, S. 2004a. *A critical investigation into the discourses used to construct Academic Literacy at the Durban Institute of Technology*. Unpublished PhD thesis. Rhodes University.
- McKenna, S. 2004b. The intersection between academic literacies and student identities. *South African Journal of Higher Education*, 18(3):269–280.
- McKenzie, K. & Schweitzer, R. 2001. Who succeeds at university? Factors predicting academic performance in first year Australian university students. *Higher Education Research and Development*, 20(1):21-33.
- McNamara, T. 1996. *Measuring second language performance*. Londen: Longman.
- McNamara, T. 2000. *Language testing*. Oxford: Oxford University Press.
- McNamara, T. 2003. Looking back, looking forward: Rethinking Bachman. *Language testing*, 20(4):466-473.
- McNamara, T. 2004. Language testing. Pp.763-783 in Davies, A. & Elder, C. (reds.) 2004.

- McNamara, T. 2008. Mapping the scope of theory in TESOL. Bydrae tot simposium: Theory in TESOL. *TESOL Quarterly*, 42(2):302-305.
- McNamara, T. & Roever, C. 2006. *Language testing: The social dimension*. Oxford: Blackwell.
- Medina, N. & Neill, D.M. 1990. *Fallout from the testing explosion*. Cambridge, MA: National Center for Fair and Open Testing.
- Messick, S. 1980. Test validity and the ethics of assesment. *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. Pp. 33–46 in Wainer, H. & Braun, I.H. (reds.), 1988.
- Messick, S. 1989. Validity. Pp.3-103 in Linn, R.L. (red.), 1989.
- Messick, S. 1994. The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2):13-23.
- Messick, S. 1995. Validity of psychological assessment: Validation of inferences from person's responsesand performances as scientific inquiry into score meaning. *American Psychologist*, 50(9):741-749.
- Messick, S. 1996. Validity and washback in language testing. *Language Testing*, 13:242-256.
- Messick, S. (red.) 1999. *Assessment in higher education. Issues of access, quality, student development, and public policy*. Mawah, New Jersey.: Lawrence Erlbaum.

- Miller, M.D., Linn, R.L. & Gronlund, N.E. 2009. *Measurement and assessment in teaching*. Upper Saddle River, New Jersey: Pearson Education.
- Ministerie van Onderwys. 2001. *National plan for higher education*. Pretoria: Ministry of Education.
- Mkhabela, M. & Malan, M. 2004. By die 40% eerstejaars druij. Skoksyfer kos belastingbetalers R1 miljard p.j. *Rapport*, 26 September 2004:2.
- Moss, P.A. 1994. Can there be validity without reliability? *Educational Researcher*, 23(2):5-12.
- Moss, P.A. 1996. Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher*, 25(1):20-28.
- Nation, P. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nel, C. 2007. *Die oorgang van skool na universiteit: 'n Teoretiese raamwerk vir 'n pre-universitêre intervensie*. Ongepubliseerde doktorsale tesis. Stellenbosch: Universiteit Stellenbosch.
- Newton-Smith, W.H. (red.) 1999. *A companion to the philosophy of science*. Oxford: Blackwell.
- Nitko, A.J. 2004. *Educational assessment of students*. 4^{de} Uitgawe. Upper Saddle River, New Jersey: Prentice-Hall.
- Norton, K.& Toohey, K. (reds.), 2004. *Critical pedagogies and language learning*. Cambridge: Cambridge University Press.
- Ntshoe, I.M. 2002. National plan for higher education in South Africa: A programme for equity and redress or globalised competition and managerialism? *South African Journal of Higher Education*, 16(2):7-10.

- O'Connor, B.P. 2000. Behavioural research methods. *Instrument sand computers*, 32(3):396-402.
- Oller, J.W. 1973. Discrete-point tests versus tests of integrative skills. Pp.184-199 in Oller, J.W. & Richards, J.C. (reds.), 1973.
- Oller, J.W. 1975. Assessing competence in ESL. Pp.25-36 in Palmer, S. & Spolsky, B. (reds.), 1975.
- Oller, J.W. 1978. Pragmatics and language testing. Pp.39-57 in Spolsky, B. (red.), 1978.
- Oller, J.W. 1979. *Language tests at school: A pragmatic approach*. Londen: Longman.
- Oller, J.W. & Richards, J.C. (reds.) 1973. *Focus on the learner: Pragmatic perspectives for the language teacher*. Rowley, Massachusetts: Newbury House.
- Palmer, S. & Spolsky, B. (reds.) 1975. *Papers on language testing*. TESOL, Washington D.C.
- Paras, J. 2001. Crisis in mathematics education. Student failure: Challenges and possibilities. *South African Journal of Higher Education*, 15(3):66-73.
- Paxton, M. 1994. Evaluation of writing programmes – merely an issue of measurement? Pp.239-250 in Weideman, A.J. (red.), 1994.
- Pennycook, A. 2004. Critical applied linguistics. Pp.784-807 in Davies, A. & Elder, C. (reds.) 2004.
- Perkins, D.M. 1991. Improvement of reading and vocabulary skills at the University of Transkei. *South African Journal of Education*, 11(4):231-235.
- Pierce, W. 2003. *Student preparedness for college work*. [Intyds]. Beskikbaar: <http://academic.pg.cc.md.us/~wpeirceMCCCTR/ctrep~1.html>. Besoek 24 Maart 2004.

- Popham, W.J. 1997. Consequential validity: right concern – wrong concept. *Educational Measurement: Issues and practice*, 16(2):9-13.
- Popham, W.J. 2008. *Classroom assessment: what teachers need to know*. Boston: Pearson Education.
- Pretorius, E.J. 1995. Reading as an interactive process: Implications for studying through the medium of a second language. *Communicatio*, 21(2):33-43.
- Pride, J.B. & Holmes, J. (reds.). 1972. *Sociolinguistics: selected readings*. Harmondsworth: Penguin.
- Rademeyer, A. 2005a. Te veel studente ‘hoort nie werklik op universiteit’: Tot soveel as 30% van eerstejaars het nie vrystelling nie. *Die Burger*, 14 Mei 2005:4.
- Rademeyer, A. 2005b. Klas van 2008. *Insig*, 28 Februarie 2005:40.
- Rajagopalan, K. 2004. The philosophy of applied linguistics. Pp.397-420 in Davies, A. & Elder, C. (reds.) 2004.
- Rambiritch, A. S.a. *Accessibility, transparency and accountability as regulative conditions for a post-graduate test of academic literacy*. Ongepubliseerde en onvoltooide doktorsale tesis. Bloemfontein: Universiteit van die Vrystaat.
- Reddy, V. 2006. Introduction. Pp. xii-xix in Reddy, V. (red.), 2006. *Marking matrix: Colloquium proceedings*. Kaapstad: HSRC Press.
- Richards, J.C. & Schmidt, R.W. (reds.). 1983. *Language and communication*. Londen: Longman.
- Rivers, W. 1968. *Teaching foreign language skills*. Chicago: The University of Chicago Press.
- Robinson-Kurpius, S.E. & Stafford, M.E. 2006. *Testing and measurement. A user-friendly guide*. California: Sage Publications Inc.

- Roets, N. 1990. *Afrikaans vir akademiese doeleindes*. Ongepubliseerde doktorsale tesis. Johannesburg: Randse Afrikaanse Universiteit.
- Schmitt, N. (red.), 2002. *An introduction to applied linguistics*. Londen: Longman.
- Schmitt, N. & Celce-Murcia, M. 2002. An overview of applied linguistics. Pp.1-16 in Schmitt, N. (red.), 2002. *An introduction to applied linguistics*. Londen: Longman.
- Schuurman, E. 1972. *Techniek en toekomst: Confrontatie met wijsgerige beschouwingen*. Assen: Van Gorcum.
- Schuurman, E. 2005. *The technological world picture and an ethics of responsibility: Struggles in the ethics of technology*. Sioux Center, Iowa: Dordt College Press.
- Scott, I.; Yeld, N. & Hendry, J. 2007. A case for improving teaching and learning in South African higher education. *Higher Education Monitor* No. 6. Pretoria: Council on Higher Education. [Intyds]. Beskikbaar: <http://www.che.ac.za/documents/d000155/index.php>. Besoek 17 September 2009.
- Searle, J.R. 1969. *Speech acts: An essay in the philosophy of language*. Londen: Cambridge University Press.
- Shohamy, E. 1997. *Critical language testing and beyond*. Paper presented at the annual conference of the American Association of Applied Linguistics in March 1997 in Orlando, Florida.
- Shohamy, E. 2001. *The power of tests: a critical perspective on the uses of language tests*. Harlow, Essex: Pearson Education.
- Shohamy, E. 2004. Assessment in multicultural societies: Applying democratic principles and practices to language testing. Pp. 72-92 in Norton, K.& Toohey, K. (reds.), 2004.

- Skehan, P. 1988. Language testing. Part I. *Language Teaching*. Januarie 1988: 11-221.
- Sluijter, C. 1998. *Toetsen en beslissen: Toetsing bij doorstroombeslissingen in het voortgezet onderwijs*. Ongepubliceerde proefschrift. Arnhem: Cito.
- Smit, G.J. 1991. *Psigometrika: Aspekte van toetsgebruik*. Silverton: HAUM Opvoedkundige Uitgewers.
- Smit, C., Borraine, H.& Owen, R. 2006. *Statistical analysis of AARP and TALL results: 2004, UP candidates*. Ongepubliceerde verslag. Pretoria: Universiteit van Pretoria.
- Spolsky, B. 1973. What does it mean to know a language, or how do you get someone to perform his competence? Pp.164-176 in Oller, J.W. & Richards, J.C. (reds.), 1973.
- Spolsky, B. 1978. Introduction: Linguists and language testers. Pp.v-x in Spolsky, B. (red.). 1978. *Advances in language testing*. Arlington: Centre for Applied Linguistics.
- Spolsky, B. 1995. *Measured words*. Oxford: Oxford University Press.
- Squelch, J. 1993. Towards a multicultural approach to education in South Africa, Pp.171-201 in Le Roux, J. (red.), 1993.
- Sridhar, S.N. 1993. What are applied linguistics? *International Journal of Applied Linguistics*, 3(1):3-16.
- Starfield, S. 2000. Assessing students' writing. Pp.102-117 in Leibowitz, B. & Mohamed, Y. (reds.), 2000.
- Street, B. 1984. *Literacy in theory and practice*. Cambridge: Cambridge University Press.

- Street, B. 1993. *Cross cultural approaches to literacy*. Cambridge: Cambridge University Press.
- Street, B. 1995. *Social literacies: critical approaches to literacy in development, ethnography and education*. Londen: Longman.
- Sundberg, N.D. 1977. *Assessments of persons*. New Jersey: Prentice Hall.
- Thesen, L. 1997. Voices, discourse and transition: In search of new categories in EAP. *TESOL Quarterly*, 31(3):487-511.
- Thesen, L. & Van Pletzen, E. 2006. *Academic literacy and the languages of change*. Londen: Continuum.
- Thorndike, R.L. & Hagen, E. 1969. *Measurement and evaluation in psychology and education*. New York: John Wiley & Sons.
- Trout, J.D. 1999. Measurement. Pp. 265-276 in Newton-Smith, W.H. (red.), 1999.
- Tyler, L. 1963. *Tests and measurement*. Englewood Cliffs, New Jersey: Prentice Hall.
- Umalusi. 2004. *Investigation into senior certificate examination*. Summary report on the evaluation of the senior certificate examination, August, 2004. Pretoria: Umalusi.
- Van der Slik, F. 2005. Statistiese analise van die TAG/TALL 2004 uitslae. Seminaar aangebied by toetsontwikkelingsessie van 1-3 Junie 2005, Universiteit van Pretoria. Ongepubliseerde opmerkings en bespreking.
- Van der Slik, F. 2009. Gender bias and gender differences in two South African tests of academic literacy. *Southern African Linguistics and Applied Language Studies*, 27(3): 277-290.
- Van der Slik, F. & Weideman, A.J. 2005. The refinement of a test of academic literacy. *Per Linguam*, 21(1):23-35.

- Van der Slik, F. & Weideman, A.J. 2007. Testing academic literacy over time: is the academic literacy of first year students deteriorating? *Ensovoort*, 11(2):126-137.
- Van der Slik, F. & Weideman, A.J. 2008. Measures of improvement in academic literacy. *Southern African Linguistics and Applied Language Studies*, 26(3):363-378.
- Van der Slik, F. & Weideman, A.J. 2009. Revisiting test stability: further evidence relating to the measurement of difference in performance in a test of academic literacy. *Southern African Linguistics and Applied Language Studies*, 27(3):253-263.
- Van der Slik, F. & Weideman, A.J. 2010. Examining bias in a test of academic literacy: Does the *Test of Academic Literacy Levels (TALL)* treat students from English and African language backgrounds differently? Te verskyn.
- Van der Walt, J.L. & Steyn, H.S. (Jnr). 2007. Pragmatic validation of a test of academic literacy at tertiary level. *Ensovoort*, 11(2):138-153.
- Van Dyk, T.J. 2005. Towards providing effective academic literacy intervention. *Per Linguam*, 21(2):38-51.
- Van Dyk, T.J. & Kistner, L. 2008. *Verslag oor akademiese geletterdheid van eerstejaarstudente aan die Universiteit Stellenbosch: die 2005- tot 2008-kohorte*. Ongepubliseerde verslag. Stellenbosch: Universiteit Stellenbosch.
- Van Dyk, T.J., Van Dyk, L., Blanckenberg, H.C. & Blanckenberg, J. 2007. Van bevreemdende diskoers tot toegangsportaal: e-Leer as aanvulling tot 'n akademiese geletterdheidskursus. *Ensovoort*, 11(2), 154-172.
- Van Dyk, T.J. & Weideman, A.J. 2004a. Switching constructs: On the selection of an appropriate blueprint for academic literacy assessment. *Journal for Language Teaching*, 38(1): 1-13.

- Van Dyk, T.J. & Weideman, A.J. 2004b. Finding the right measure: from blueprint to specification to item type. *Journal for Language Teaching*, 38(1):15-24.
- Van Dyk, T.J., Zybrands, H., Cillié, K., Coetzee, M. 2009. On being reflective practitioners: the evaluation of a writing module for first-year students in the Health Sciences. *Southern African Linguistics and Applied Language Studies*, 27(3):333-344.
- Van Els, T., Bongaerts, T., Extra, G., Van Os, C. & Janssen-van Dieten, A. 1984. *Applied Linguistics and the Learning and Teaching of Foreign Languages*. Groningen: Wolters-Noordhof.
- Van Rensburg, C. & Weideman, A.J. 2002. Language proficiency: Current strategies, future remedies. *Journal for Language Teaching*, 36(1&2):152-163.
- Van Schalkwyk, S.C. 2008. *Acquiring academic literacy: A case of first-year extended degree programme students at Stellenbosch University*. Ongepubliseerde doktorsale tesis. Stellenbosch: Universiteit Stellenbosch.
- Van Wijk, T. 1987. Voëlvlug oor die opset en ontwikkeling van die Suid-Afrikaanse universiteit, Pp. 16-26 in *Instituut vir Reformatoriese Studie. Venster op die universiteit*. Potchefstroom: Potchefstroomse Universiteit vir Christelike Hoër Onderwys.
- Verhelst, N.D. 2000. *Estimating the reliability of a test from a single test administration*. Measurement and Research Department Reports 98-2. Arnhem: National Institute for Educational Measurement.
- Vermeulen, P. & Melck, A.P. 2004. The new funding framework: Challenges for the University of Pretoria. Aanbieding aan die Senaat van die Universiteit van Pretoria op 28 Januarie 2004, Universiteit van Pretoria. Ongepubliseerde opmerkings en bespreking.

- Viljoen, H. 2005. Constraints in the teaching and learning of humanities at South African universities: A Delphi-study. *South African Journal of Higher Education*, 19(1):31-43.
- Visser, A. & Hanslo, M. 2005. Approaches to predictive studies: Possibilities and challenges. *South African Journal of Higher Education*, 19(6):1160-1176.
- Vorster, J. & Reagan, T. 1990. On the lexical development of L1 and L2 speakers. *South African Journal of Linguistics*, 9(3):80-84.
- Wainer, H. & Braun, I.H. (reds.) 1988. *Test validity*. Hillside, New Jersey: Lawrence Erlbaum.
- Webb, V.N. 2002. English as a second language in South Africa's tertiary institutions: a case study at the University of Pretoria. *World Englishes*, 21(1):49-69.
- Weideman, A.J. 1981. *Systematic concepts in linguistics*. Unpublished MA thesis. Bloemfontein: University of the Free State.
- Weideman, A.J. 1988. *Linguistics: A crash course for students*. Bloemfontein: Patmos.
- Weideman, A.J. 1999. Five generations of applied linguistics: Some framework issues. *Acta Academica*, 31(1):77-98.
- Weideman, A.J. 2003a. *Academic literacy: Prepare to learn*. Pretoria: Van Schaik.
- Weideman, A.J. 2003b. Justifying course and task construction: Design considerations for language teaching. *Acta Academica*, 35(3):26-48.
- Weideman, A.J. 2003c. Assessing and developing academic literacy. *Per Linguam*, 19(1&2):55-65.
- Weideman, A.J. 2003d. Towards accountability: a point of orientation for post-modern applied linguistics in the third millennium. *Literator*, 24(1):1-20.

- Weideman, A.J. 2006. Transparency and accountability in applied linguistics. *Southern African Linguistics and Applied Language Studies*, 24(1):71-86.
- Weideman, A.J. 2007a. The redefinition of applied linguistics: Modernist and postmodernist views. *Southern African Linguistics and Applied Language Studies*, 25(4):589-605.
- Weideman, A.J. 2007b. A responsible agenda for applied linguistics: Confessions of a philosopher. *Per Linguam*, 23(2):29-53.
- Weideman, A.J. 2007c. *Self-evaluation: Unit for Academic Literacy 2007*. Ongepubliseerde verslag. Pretoria: Universiteit van Pretoria.
- Weideman, A.J. 2009a. Constitutive and regulative conditions for the assessment of academic literacy. *Southern African Linguistics and Applied Language Studies*, 27(3): 235-251.
- Weideman, A.J. 2009b. Uncharted territory: A complex systems approach as an emerging paradigm in applied linguistics. *Per Linguam*, 25(1):61-75.
- Weideman, A.J. 2009c. *Beyond expression: A systematic study of the foundations of linguistics*. Grand Rapids, Michigan: Paideia Press.
- Weideman, A.J. 2010. *A framework for the study of linguistics*. Ongepubliseerde manuskrip.
- Weideman, A.J.(red.), 1994. Redefining applied linguistics. *Proceedings of the 14th annual conference of the Southern African Applied Linguistics Association*. Bloemfontein: SAALA.
- Weideman, A.J. & Van der Slik, F. 2008. The stability of test design: Measuring difference in performance across several administrations of a test of academic literacy. *Acta Academica*, 40(1):161-182.
- Weir, C.J. 1990. *Communicative language testing*. Londen: Prentice Hall.

- Weir, C.J. 1993. *Understanding and developing language tests*. Hertfordshire: Prentice Hall Europe.
- Weir, C.J. 2005. *Language testing and validation*. Houndmills, Basingstoke: Palgrave Macmillan.
- Widdowson, H.G. 1979. *Explorations in applied linguistics*. Oxford: Oxford University Press.
- Widdowson, H.G. 2000. On the limitations of linguistics applied. *Applied Linguistics*, 21(1):3-25.
- Widdowson, H.G. 2001. Coming to terms with reality: Applied linguistics in perspective. Pp.2-17 in Graddol, D. (red.), 2001.
- Wilkins, D.A. 1976. *Notional syllabuses: A taxonomy and its relevance to foreign language curriculum development*. Oxford: Oxford University Press.
- Williams, E. 2004. Literacy studies. Pp.576-603 in Davies, A. & Elder, C. (reds.), 2004.
- Wood, T. 1998. Issues relating to the cognitive development of students at historically disadvantaged institutions. *South African Journal of Higher Education*, 12(1):87-94.
- Yeld, N. 2001. *Assessment, equity and language of learning: Key issues for higher education selection in South Africa*. Ongepubliseerde doktorsale tesis. Kaapstad: University of Cape Town.
- Yeld, N. & Hendry, J. 2002. *Grade inflation in the Senior Certificate?* Ongepubliseerde voorlopige verslag oor die problematiek van simbole-inflasie. Kaapstad: University of Cape Town.
- Zamel, V. & Spack, R. (reds.). 1998. *Negotiating academic literacies: Teaching and learning across languages and cultures*. Mahwah, New Jersey: Lawrence Erlbaum.

Opsomming en sleutelsterme

Opsomming

Deurvloeikoerse aan universiteite in Suid-Afrika is laag en dit dra onder andere by tot finansiële verliese vir 'n verskeidenheid van belanghebbendes: die studente self, hul ouers/voogde, beursgewers, universiteite, en die staat. Dit noop instellings derhalwe om ondersoek in te stel na daardie faktore wat 'n invloed het op studiesukses. Verskeie internasionale en plaaslike ondersoeke toon aan dat akademiese taalvermoë een van 'n aantal faktore is wat wel 'n beduidende invloed op akademiese sukses het. Ten einde ingeligte besluite oor laebelang kwessies soos taalondersteuning te maak, benodig universiteite egter meganismes wat hul hiertoe in staat sal stel. Een sodanige meganisme is die gebruik van meetinstrumente, soos die Toets van Akademiese Geletterdheid (TAG), om diagnoses te maak oor studente se akademiese taalvaardighede (ook akademiese geletterdheid genoem) en hul dan op gepaste taalondersteuningsprogramme te plaas. Toetse het egter die vanselfsprekende vermoë om lewens op die een of ander wyse te raak, en daarom bestaan daar 'n besondere behoefte aan 'n verantwoordbare en deursigtige aanpak tot die ontwerp en implementering van hierdie tipe toetse, en die interpretasie van toetsresultate. Dit word normaalweg gedoen aan die hand van ondersoeke na veral die betroubaarheid en geldigheid van 'n toets; laasgenoemde aan die hand van 'n proses van validering. Hierdie studie is 'n voorbeeld van so 'n ondersoek waar TAG aan deeglike en sistematiese skrutineringsonderwerp is. 'n Raamwerk wat die toegepaste taalkunde as dissipline van ontwerp beskou, is gevolg waarteen die konstruk (teoretiese onderbou), die toetsspesifikasies (die bloudruk) en taaktipes, sowel as die betroubaarheid en geldigheid van TAG ondersoek is. Die raamwerk hou in dat 'n toets 'n tegniese ontwerp is wat as oplossing vir 'n (taal)probleem voorgelê kan word. Hierdie tegniese ontwerp moet aan bepaalde konstitutiewe beginsels, wat die funderende funksie van die ontwerp is, voldoen en dit word weer

deur regulatiewe beginsels ontsluit. Eersgenoemde sluit aspekte soos betroubaarheid en geldigheid in en laasgenoemde implementeerbaarheid, bruikbaarheid, deursigtigheid, verantwoordbaarheid en sorgsaamheid. Hoewel die studie op die konstitutiewe voorwaardes vir die ontwerp en ontwikkeling van TAG gefokus het, is daar ook telkens verwys na die belang van 'n wisselwerking tussen die konstitutiewe en die regulatiewe voorwaardes en die feit dat 'n meetinstrument soos TAG nie net vanuit een perspektief ondersoek kan word nie. In hierdie studie word daar dus 'n argument gevoer vir 'n tegnies-verantwoordbare aanpak tot die ontwikkeling en implementering van, en ondersoeke na meetinstrumente soos TAG ten einde dit op regverdig en verantwoordelike wyse te kan gebruik. Die slotsom van hierdie studie is dat TAG 'n toets van hoë kwaliteit is wat sonder huiwering deur universiteite vir plasingsdoeleindes aangewend kan word, omdat dit op verantwoordelike wyse en met besondere sorgsaamheid ontwerp, geïmplementeer en deurlopend ondersoek word.

Abstract

Throughput rates at South African universities are low and contribute, among others, to financial losses for a number of stakeholders: students, their parents/guardians, donors of scholarships, universities and the state. This compels institutions to investigate those factors influencing study success. Various international and local investigations indicate that academic language ability is one of several factors that has a significant influence on academic success. In order to make informed decisions about low stakes issues such as language support, universities need mechanisms to enable them to do so. One such mechanism is the use of measuring instruments, such as the Toets van Akademiese Geletterdheidsvlakke (TAG), to make a diagnosis of students' academic language ability, and to then place them on appropriate language support programmes. Tests, however, have the obvious power to touch the lives of people in some way. It is therefore necessary that an accountable and transparent approach is followed when designing and implementing tests such as TAG, and when interpreting test results.

This is usually done by investigating the reliability and validity of a test; the latter by means of a process of validation. This study is an example of such an investigation where TAG is subjected to thorough and systematic scrutiny. A framework that considers applied linguistics as a discipline of design was followed, against which the construct (underlying theoretical framework), test specifications (the blueprint) and task types, as well as the reliability and validity of TAG were investigated. The framework suggests that a test is a technical design that can be used as solution to a (language) problem. This technical design is grounded upon certain constitutive conditions underlying it, and disclosed by regulative conditions. The former include aspects such as reliability and validity, and the latter utility, transparency, accountability and care. Although the focus of this study was on the constitutive conditions for designing and developing TAG, it repeatedly refers to the importance of the interplay between constitutive and regulative conditions, as well as the fact that a test such as TAG cannot be investigated from one perspective only. The argument of this study is thus that a technically accountable approach to the development and implementation of, and investigations into tests such as TAG is necessary in order to use them in a justifiable and responsible manner. The conclusion of this study is that TAG is a test of high quality and that it can without a doubt be used by universities for purposes of placement, because it was designed and implemented with care and its results are constantly being scrutinised.

Sleuteltermen

Academic literacy

Applied linguistics

Concurrent validity

Consequential validity

Constitutive conditions

Construct validity

Content validity

Design principles

Equivalent form reliability

Face validity

Internal consistency

Interpretation

Justification

Predictive validity

Regulative conditions

Reliability

Task types

Test

Test Retest Reliability

Testing

Transparency

Utility

Validation

Validity

Bylaag

A

**Toetsafnemershandleiding vir die
Noordwes Universiteit, die
Universiteit van Pretoria, en die
Universiteit van die Vrystaat**

1. Inleiding

Die Toets van Akademiese Geletterdheidsvlakke (TAG) is die akademiese geletterdheidstoets van die Universiteite van Stellenbosch, Pretoria en die Vrystaat, sowel as die Noordwes Universiteit. Dit word gebruik om studente vroegtydig te identifiseer vir wie akademiese taalvaardigheid (ook akademiese geletterdheid genoem) 'n struikelblok in die suksesvolle afhandeling van hul studie kan wees. Op grond van die toetsuitslag word aanbevelings vir plasing op gepaste ondersteuningskursusse gemaak. In sommige gevalle kan die toets ook as deel van 'n toelatingstoetsbattery aangewend word en vervul dan benewens bogenoemde funksie ook ten dele die rol van 'n toelatingstoets. Die skryf van die toets is verpligtend vir alle nuweling eerstejaarstudente en word óf tydens die verwelkomingsweke van die onderskeie instellings geskryf, óf gedurende die voorafgaande jaar, soos per instelling vereis. Geen student kan benadeel word deur hierdie toets af te lê nie, hetsy dit as plasingstoets of as deel van 'n toelatingsmeganisme gebruik word.

TAG is 'n betroubare, geldige en gestandaardiseerde meetinstrument:

'n Toets wat **betroubaar** is, is 'n toets wat op enige ander plek, op enige ander tydstop, in min of meer dieselfde omstandighede, met min of meer dieselfde toetspopulasie afgeneem word, en wat ongeveer dieselfde uitslag behoort op te lewer. Dit dui dus op die interne konsekwentheid van die toets.

Indien 'n toets **geldig** is, meet dit wat dit behoort te meet (in hierdie geval, die vlak van akademiese geletterdheid). Geldigheid word op kwalitatiewe en op kwantitatiewe vlak bepaal deur empiriese navorsing.

'n **Gestandaardiseerde** toets voorveronderstel bepaalde standaard-kriteria wat konstant gehou word van een toetsgeleentheid na 'n ander. Hierdie kriteria dikteer standaard-prosedures vir die opstel en ontwikkeling van toetsinhoud, toetsafname en bepunte. Daar is 'n verband tussen die betroubaarheid en geldigheid van 'n toets en die standaardisering daarvan - indien die standaard-prosedures nie nagekom word nie, beïnvloed dit die betroubaarheid en geldigheid van die toets en kan daar teen sekere studente gediskrimineer word.

2. Rolomskrywing

In hierdie dokument word daar telkens na drie rolle verwys, naamlik die van toetskoördineerder, toetsafnemer en toesighouer. Hoewel elkeen van die persone wat vir hierdie rolle aangewys is, unieke funksies vervul funksioneer hulle tog saam as 'n span en om dié rede is hierdie enkele dokument opgestel waarin duidelik sal blyk hoe die rolle in wisselwerking tot mekaar staan en wedersyds aanvul. Dit is daarom belangrik dat alle betrokkenes van hierdie volledige dokument kennis sal neem.

2.1 Toetskoördineerder

Die toetskoördineerder is in beheer van die volledige toetsproses en alle personeel betrokke. Hy/sy is daarom verantwoordelik vir alle logistieke beplanning en organisasie rondom die toetsingsgeleentheid. Die opleiding en begeleiding van toetsafnemers en toesighouers resorteer ook onder hierdie persoon en almal rapporteer aan hom/haar op die dag van toetsing. Alle navrae kan ook aan hom/haar gerig word.

2.2 Toetsafnemer

Die toetsafnemer is daardie persoon wat in die onderskeie lokale in beheer is van die toets. Hy/sy is in die breë verantwoordelik vir die bestuur van die toetssessie (die logistiek, sowel as die toetsafnemers) in 'n betrokke lokaal en moet toesien dat daar streng by die riglyne in hierdie handleiding gehou word. Alle toesighouers wat aan 'n spesifieke lokaal toegedeel is, rapporteer aan hom/haar tydens die toetssessie.

2.3 Toesighouer

'n Toesighouer is daardie persoon wat verantwoordelik is vir die implementering van alle riglyne in hierdie handleiding vervat. Hy/sy rapporteer aan die toetsafnemer in 'n spesifieke lokaal en moet deeglik vertrou wees met die inhoud van hierdie handleiding, aangesien die sukses van 'n toetssessie grotendeels toe te skryf is aan die gestandaardiseerde administrasie daarvan.

3. Toetsadministrasie

Aangesien TAG 'n gestandaardiseerde toets is, is dit noodsaaklik dat jy jousef vooraf deeglik voorberei. Hier volg 'n uiteensetting van die toetsprosedure en die take wat uitgevoer moet word.

3.1 Voor die toets

- Raak vertrouwd met die toets, enige ander toetsmateriaal, en die prosedure vir die afneem daarvan.
- Bring die nodige veranderinge aan die toetslokaal aan: verskuiwing van meubels, verstelling van lugversorging en beligting, fasiliteite vir gestremde studente, ens.
- Maak seker dat die volgende in die lokaal beskikbaar is:
 - Voldoende toets-/eksamen-registerstrokie
 - Voldoende toetsboeke en antwoordblaaie
 - Voldoende potlode, skerpmakers en uitveërs
 - 'n Oorhoofse projektor en/of 'n dataprojektor in 'n werkende toestand

3.2 Tydens die toets

Fase 1

- By elke ingang van die toetslokaal moet twee toesighouers wees wat toets-/eksamen-registerstrokie en potlode uitdeel soos wat die studente die toetslokaal binnegaan.
- Studente neem hul sitplekke in soos vir enige toets/eksamen: een sitplek oop tussen kandidate; alle kandidate in rye agter mekaar. Indien die lokaal te klein is vir so 'n sitplan, moet alle leë sitplekke gevul word – begin voor in die lokaal met die vul van sitplekke.
- Die res van die toesighouers moet deur die lokaal beweeg en sekermaak dat studente niks anders as die potlood en die toets-/eksamen-registerstrokie by hulle het nie, en nie die toets-/eksamen-registerstrokie voltooi nie.

Fase 2

- Bespreek die doel van die toets (bl. 1 van die toetsboek) met die studente.
- Verduidelik dat die vrae van die toets meervoudige keuse is, en dat dit op die los antwoordblad beantwoord moet word, behalwe vir die laaste afdeling wat in die toetsboek self beantwoord moet word (slegs indien van toepassing).
- Verduidelik dat die skryf van die toets niemand kan benadeel nie, hetsy dit as plasingstoets of as deel van 'n toelatingsmeganisme gebruik word – verskaf die nodige redes hiervoor.
- Ongeveer een uur tyd word vir die aflê van die toets toegelaat (kyk op bl. 1 van die toets vir 'n presiese tydsaanduiding).
- Verduidelik waar en wanneer die toetsresultate bekendgemaak sal word, en in watter formaat.
- Kondig aan dat geen rofwerkpapier (daar is genoeg ruimte in die toetsboeke om rofwerk op te doen) of sakrekenaars/selfone toegelaat word nie en derhalwe weggesit moet word.

Fase 3

- Voltooi die toets-/eksamen-registerstrokies saam met die studente deur 'n voorbeeld daarvan op die oorhoofse projektor/dataprojektor te plaas. Toesighouers beweeg deur die lokaal en sien toe dat studente dit korrek voltooi.
- Wanneer die toets-/eksamen-registerstrokies voltooi is, word dit na die sykkante (links, en/of regs, afhangend van die lokaaluitensetting) van die lokaal gestuur.
- Toesighouers beweeg van ry tot ry en tel of die aantal toets-/eksamen-registerstrokies ooreenkom met die aantal persone wat in die ry sit. Sodra al die toets-/eksamen-registerstrokies ingeneem en getel is, moet enkele toesighouers dit weer tel en sekermaak dat die totale aantal persone in die lokaal ooreenstem met die totale aantal toets-/eksamen-registerstrokies.
- Voltooi die toepaslike kontroleform.
- Kondig aan dat die toetsboeke nou uitgedeel sal word, dat daar absolute stilte moet heers en dat die toetsboeke nie oopgemaak mag word nie. Toesighouers deel die toetsboeke met die ingevoegde los antwoordblaaie ry vir ry uit (slegs die aantal wat per ry benodig word).

- Kontroleer dat daar nie per abuis te veel toetsboeke per ry uitgedeel is nie en dat niemand oorgeslaan is nie.
- Laat die studente die toetsboeke op die eerste bladsy oopmaak. Herinner hulle daaraan dat jy reeds *Algemene inligting* (bl. 1 van die toetsboek) met hulle bespreek het – verwys Fase 2 hierbo. Hulle moet nou die toetsboek op die aangewese plek (ook bl. 1) onderteken.
- Blaai vervolgens deur die toetsboek en laat die studente ook deur hul boeke blaai om te verseker dat alle bladsye gedruk is (alle afdelings en vrae), en dat daar 'n los antwoordblad in die boek gevoeg is. Versoek die studente om die toetsboek en die los antwoordblad van mekaar te skei.
- Begin om die buiteblad van die toetsboek saam met die studente te voltooi deur 'n voorbeeld daarvan op die oorhoofse projektor/dataprojektor te plaas. Toesighouers beweeg deur die lokaal en sien toe dat studente dit korrek voltooi.
- Voltooi vervolgens die biografiese besonderhede van die los antwoordblad saam met die studente deur 'n voorbeeld daarvan op die oorhoofse projektor/dataprojektor te plaas – onnodige merke op die los antwoordblad word nie toegelaat nie. Toesighouers beweeg deur die lokaal en sorg dat studente dit korrek voltooi. NB: Herinner studente daaraan dat hulle nie die los antwoordblaai bo-op die gekleurde buiteblad van die toetsboeke moet plaas/daarop druk nie, aangesien die kleur van die buiteblad op die antwoordblad kan afgee. Toesighouers moet veral sekermaak dat studente hul geboortedatums en/of studentennommers korrek kodeer.

Fase 4

- Herinner studente daaraan dat hulle beide die toetsboeke en die antwoordblaai na afloop van die toets moet inhandig en dat geen sakrekenaars/selfone of rofwerkpapier toegelaat word nie.
- Kondig die presiese tyd wat die studente het om die toets te voltooi, aan (inligting op bl. 1 van toetsboek). Indien hulle sukkel met 'n afdeling/vraag moet hulle aangaan na die volgende afdeling/vraag. Hulle moet vinnig en akkuraat werk. Niemand mag die lokaal tydens die toetstyd verlaat of laat betree nie.

- Die toets neem 'n aanvang. Die toegelate tyd vir voltooiing van die toets begin dus van hierdie punt af.
- Toesighouers moet stil deur die lokaal beweeg en toesien dat daar nie ongerymdhede plaasvind nie. Indien wel, moet die toetsafnemer onmiddellik daarvan in kennis gestel word.
- Kondig aan wanneer die helfte van die tyd verstreke is. Nadat die toegelate tyd verstreke is, kondig aan dat die toegelate tyd verstreke is en dat potlode neergesit moet word.

3.3 Na afloop van die toets

- Nadat die toetstyd verstreke is, moet studente hul antwoordblaaie voor in hul toetsboeke plaas (ongeveer 2 cm. moet aan die bokant uitsteek vir kontroledoeleindes). Die toetsboeke word dan na die sykant (links, en/of regs, afhangend van die lokaaluitensetting) van die lokaal gestuur.
- Toesighouers beweeg van ry tot ry en tel of die aantal toetsboeke, met ingeslote antwoordblaaie, ooreenkom met die aantal persone wat in die ry sit. Intussen word alle potlode ook na die sykant gestuur. Sodra al die toetsboeke met ingeslote antwoordblaaie ingeneem en getel is, kan studente die lokaal verlaat.
- Toesighouers tel die totale aantal toetsboeke met ingeslote antwoordblaaie weer en kontroleer dat dit ooreenstem met die totale aantal persone wat die toets afgelê het in die in die spesifieke lokaal. Hierdie getal moet ook ooreenstem met die totale aantal toets-/eksamen-registerstrokies wat in Fase 3 hierbo voltooi is.
- Alle toetse word vervolgens grof gealfabetiseer ('n streng alfabetiese volgorde word nie vereis nie; slegs alle A's, B's, C's, ens. word bymekaar geplaas).
- Toesighouers plaas die toetsboeke en antwoordblaaie in die toepaslike houer.
- Toesighouers bondel die toetsboeke in pakke van 50 saam en bind 'n rekkie om die onderskeie pakke. Sorg dat elke pak met boeke deeglik gemerk is (1-50, 51-100, 101-150, ens.) en dui die lokaal, toetstyd en toetsafnemer se naam duidelik daarop aan. Die toetsafnemer teken die genoemde kontroleform sodra hy/sy tevrede is dat alles volgens standaard-prosedure hanteer is.

- Handel met antwoordblaaie, toetsboeke en toets-/eksamen-registerstrokie soos met die toetskoördineerder ooreengekom.
- Toesighouers maak potlode skerp en kry lokaal reg vir volgende toetsessie.

AKADEMIESE GELETTERDHEIDSTOETSE KONTROLEVORM

FAKULTEIT: _____

LOKAAL: _____

DATUM²:

1 FEBRUARIE	6 FEBRUARIE	7 FEBRUARIE
-------------	-------------	-------------

TYD:

08:00	10:00	12:00
-------	-------	-------

AANTAL TOETSE ONTVANG: _____

AANTAL TOETSE GEBRUIK: _____

AANTAL TOETSE FOUTIEF: _____

AANTAL TOETSE ONGEBRUIK: _____

NAAM EN HANDTEKENING VAN TOETSAFNEMER:

NAAM

HANDTEKENING

² Let daarop dat 'datum' en 'tyd' hier slegs vir illustratiewe doeleindes ingevul is. Dit word van jaar tot jaar en van instelling tot instelling aangepas en toetsafnemers/toesighouers merk slegs die toepaslike blokkie af.

Bylaag

B

Toetsspesifikasies vir voornemende studente aan die Universiteit Stel- lenbosch



1. Inleidend

Die doel van die toegangstoetsbattery (TGT-battery) van die Universiteit Stellenbosch is om 'n universiteitseie maatstaf te bied om vlakke van gereedheid in Akademiese Geletterdheid (akademiese taalvermoë), Syfervaardigheid (kwantitatiewe geletterdheid), Wiskunde, Fisiese Wetenskappe en Denkvaardighede te meet, belangrike en aanvullende inligting ten opsigte van toegangskwessies te gee en die NSS-uitslae te yk. Die toegangstoetse word primêr gebruik om toegang te bepaal. Aansoekers se TGT-uitslae dra 'n gewig van 40% vir toelating tot die Universiteit, en hulle finale Graad 12-uitslae 'n gewig van 60%.

2. Oorsig

Toetsnaam	Akademiese Geletterdheid in Afrikaans en/of Engels
Kode(s)	
Doelgroep	Alle voornemende eerstejaarstudente aan die Universiteit Stellenbosch
Tyd	Ongeveer 60 minute per toets, afhangend van die spesifieke toets
Aantal vrae	Ongeveer 60 vrae per toets, afhangend van die spesifieke toets
Formaat	Volledig multikeuse, en by uitsondering aangevul met 'n skryfstuk
Voorwaardes	Geen sakrekenaar word toegelaat nie
Gewig	'n TGT-gemiddelde word bereken op grond van aansoekers se beste taalresultaat (Afrikaans óf Engels) saam met die punte van die twee ander toetse wat hulle moes skryf (Wiskunde, Fisiese Wetenskappe, Syfervaardighede en/of Denkvaardighede). Slegs die beste uitslag van die twee Akademiese Geletterdheidstoetse tel dus een derde van die totale toegangstoetsbattery en 13% van die uiteindelijke toelatingspunt in die 40:60-verhouding. Om hierdie rede kan die Akademiese Geletterdheidstoets op sy eie niemand benadeel nie.

3. Doel en konstruk

Teoretiese vertrekpunte en doel	Die teoretiese onderbou (die konstruk) van die Akademiese Geletterdheidstoets is gesetel in die navorsing wat gedoen is deur Bachman en Palmer (2006), Blanton (1994), Boughey (2000), Gee, 1990; 2003), Kress (1989), Lea en Street (1998), McKenna (2010), Van Dyk en Weideman (2004), Van Schalkwyk (2008)
--	---

Konstruk

en Yeld *et al.* (2000).

Op grond van bostaande literatuur, konsultasie met vakkundiges op nasionale en internasionale vlak, en die inwerkingstelling van 'n nasionale werkgroep wat twee keer per jaar byeenkom om toetse te ontwikkel en te verfyn, het die toets in hoofsaak ten doel om te bepaal of voornemende studente oor die vermoë beskik om taal as 'n sosiale instrument korrek en gepas aan te wend in spesifieke kontekste waar menslike interaksie (gesproke en geskrewe) deurlopend bemiddel en beding moet word. Die onderbou van die konstruk is gesetel in taalkennis (wat organisatoriese en pragmatiese kennis insluit), sowel as strategiese bekwaamheid (wat metakognitiewe strategieë insluit).

'n Bedreweheid in kommunikasie waar die proses van taalgebruik, eerder as taal opsigself, sentraal staan, word dus van studente verwag. So 'n kommunikatiewe en interaktiewe opvatting oor akademiese diskoers heg dus tipies nie meer nêr waarde aan studente se vermoë om nuwe woordeskat en taalkennis te verkry en taal korrek in verskillende kontekste en vir verskillende doeleindes aan te wend nie, maar ook na die mate waartoe studente die volgende akademiese aksies, in verhouding tot hulself en tot ander (beide skrywers en sprekers), suksesvol kan uitvoer:

- tekste (ook gesproke teks) in die lig van hul eie ervaring te kan vertolk, en ook eie ervarings in die lig van die tekste te vertolk;
- in die lig van daardie ervaring met tekste (ook gesproke teks) saam te stem of te verskil;
- tekste (ook gesproke teks) met mekaar in verband te bring;
- tekste (ook gesproke teks) saam te voeg, en die nuutgeskepte eenheid te gebruik om nuwe menings te vorm of argumente te bou;
- tot opinies te kan kom deur van inligting uit tekste (ook gesproke teks) gebruik te maak - ekstrapolasie;
- hul eie tekste (ook gesproke teks) deur middel van enige of al bogenoemde te skep;
- oor die uitvoering van enige of al bogenoemde te kan praat en te kan skryf; en
- laasgenoemde twee punte ter bevrediging van die betrokke gehoor uit te voer.

4. Van konstruks na spesifikasies

**Belyning
van
konstruk
met
spesifikasies**

Tyd

Daar sal ongeveer 60 minute per toets vir die aflê van die Akademiese Geletterdheidstoetse toegestaan word.

Voorwaardes

Geen sakrekenaars sal toegelaat word nie.

Gewig

Slegs die beste uitslag van die twee Akademiese Geletterdheidstoetse sal een derde van die totale TGT-berekening uitmaak. Dit kom neer op 13% van die finale toelatingspunt – verwys na die toelatingsbeleid van die Universiteit Stellenbosch by www.maties.com vir meer inligting hieroor. Niemand kan dus benadeel word deur 'n Akademiese Geletterdheidstoets op sy eie nie.

Aantal vrae

Die toets bestaan uit ongeveer 60 vrae, afhangend van die spesifieke toets.

Formaat

Vrae kom slegs in multikeuse-formaat voor en kan by uitsondering met 'n skryfstuk aangevul word.

Inhoud van vrae

Die volgende word van studente in die Toets van Akademiese Geletterdheidsvlakke vereis:

- begrip binne konteks van 'n omvattende akademiese woordeskatreeks;
- die vertolking en gebruik van metafore en idiome, en begrip van bybetekenis, woordspeling en dubbelsinnigheid;
- begrip van verhoudinge tussen verskillende teksdele, bewustheid van die logiese ontwikkeling van 'n (akademiese) teks vanaf inleiding tot gevolgtrekking, en kennis van taalgebruik wat die verskillende teksdele samehangend aanmekaar bind;
- die vertolking van verskillende tekssoorte (genre), en sensitiwiteit vir die betekenis wat daardeur oorgedra word en die gehoor op wie dit gemik is;
- die vertolking, gebruik en skep van inligting in grafiese of visuele formaat;
- onderskeid tussen belangrike en nie-belangrike inligting, feit en opinie, stellings en argumente, en oorsaak en gevolg, en die indeling en hantering van data wat vergelykings tref;
- die raaksien van opeenvolging en volgorde, en die doen van eenvoudige berekeninge met betrekking tot akademiese inligting waardeur vergelykings getref en 'n argument gevoer kan word;
- kennis van wat as staving vir 'n argument dien, en die vermoë om deur middel van afleidings tot opinies te kan kom / dit te kan ekstrapoleer, en dit, of die implikasies daarvan, op ander gevalle as die een voor hande toe te pas;
- begrip van die kommunikatiewe funksie van verskillende uitdrukkingswyses in akademiese taal (soos omskrywing, die verskaffing van voorbeelde, redenering,

definiëring); en

- die soeke na betekenis (van byvoorbeeld 'n akademiese teks) buite die grense van die sin, asook die vergelyking van een deel van 'n teks met 'n ander, en een teks met 'n ander.

5. Bronnelys

- Alderson, C.; Clapham, C.; Wall, D. 1995. *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Bachman, L.F. & Palmer, A.S. 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Blanton, L.L. 1994. Discourse, artefacts and the Ozarks: understanding academic literacy. *Journal of Second Language Writing*, 3 (1):1-16. Herdruk (as Hoofstuk 17: 219-235) in Zamel, V. & Spack, R. (reds.), 1998.
- Boughey, C. 2000. Multiple metaphors in an understanding of academic literacy. *Teachers and Teaching: Theory and practice*, 6(3):279-290.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. & McNamara, T. 1999. *Dictionary of language testing*. (Series editor: Milanovic, M.). Cambridge: Cambridge University Press.
- Davidson, F. & Lynch, B.K. 2002. *Testcraft*. New Haven: Yale University Press.
- Gee, J.P. 1990. *Social linguistics and literacies: Ideology in discourses*. Londen: Falmer Press.
- Gee, J.P. 2003. Opportunity to learn: A language-based perspective on assessment. *Assessment in Education*, 10(1):27-46.
- Gronlund, N.E. 1981. *Measurement and evaluation in teaching, 4th ed.* New York: MacMillan Publishing.
- Kress, G. 1989. *Linguistic processes in sociocultural practice*. Oxford: Oxford University Press.
- Lea, M.R. & Street, B.V. 1998. Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2):157-173.
- McKenna, S. 2010. Cracking the code of academic literacy: An ideological task. Pp. 8-15 in *Beyond the university gates: Provision of Extended Curriculum Programmes in South Africa*.
- Van Dyk, T.J. & Weideman, A.J. 2004. Finding the right measure: from blueprint to specification to item type. *Journal for Language Teaching*, 38(1): 15-24.
- Van Schalkwyk, S.C. 2008. *Acquiring academic literacy: a case of first-year extended degree programme students at Stellenbosch University*. Ongepubliseerde

	<p>doktorale tesis. Stellenbosch: Universiteit Stellenbosch.</p> <p>Yeld, N. <i>et al.</i> 2000. The construct of the academic literacy test (PTEEP). Mimeograaf. Kaapstad: Alternative Admissions Research Project, University of Cape Town.</p>
--	---

Opgestel deur: T. van Dyk

Augustus 2010