

Validering van een academische taalvaardigheidstoets in de Surinaamse context: testcores en testpercepties

W. Matawlie, S. Pelgrim, J. Heeren en L. De Wachter

Samenvatting De studieduur van studenten aan de Anton de Kom Universiteit Suriname is vaak langer dan de nominale studieduur. Omdat onderzoek laat zien dat taal een noodzakelijke (maar geen voldoende) voorwaarde is voor studiesucces, lijkt het nuttig om de academische taalvaardigheid van eerstejaarsstudenten te testen en zo nodig remediërende maatregelen aan te bieden. Hiervoor is een in Vlaanderen (België) ontwikkelde digitale academische taaltest als uitgangspunt gebruikt. Om de test te valideren voor de Surinaamse context deden 109 eerstejaarsstudenten van de Faculteit der Maatschappijwetenschappen de test. Testresultaten laten een veel lagere score voor Surinaamse studenten zien in vergelijking met resultaten in België. Ondanks de talige overeenkomst, was duidelijk dat de test niet zonder meer transfereerbaar is. Daarom focust dit onderzoek expliciet op de validiteit van de test in een Surinaamse context vanuit het argumentatief kader van Knoch en Elder (2013). Deze studie gaat vanuit dat kader dieper in (1) op de psychometrische kwaliteiten van de testitems; (2) op de testperceptie door de studenten en (3) op de perceptie door de docenten. Deze triangulatie van perspectieven verschaft een duidelijker inzicht in de validiteit van de test voor de Surinaamse context.

Kernwoorden academische taalvaardigheid, validiteitsonderzoek, testperceptie

Artikelgeschiedenis

Ontvangen: 02 oktober 2021

Ontvangen in gereviseerde vorm:
19 januari 2022

Geaccepteerd: 10 oktober 2022

Online: 19 juni 2023

Contactpersoon

Widya Matawlie,
widya.matawlie@ua.aw

Copyright

© Author(s); licensed under
Creative Commons Attribution
4.0. This allows for unrestricted
use, as long as the author(s) and
source are credited.

Financiering onderzoek

-

Belangen

De auteurs hebben geen
belangen te vermelden.

147

PEDAGOGISCHE
STUDIËN

<https://doi.org/10.59302/ps.v100i2.14221>

org/10.59302/ps.v100i2.14221

ps.v100i2.14221

2023 (100) 147-171

1 Inleiding

Het taalgebruik in het hoger onderwijs is voor veel studenten wereldwijd een uitdaging. Ze worstelen met de taaleisen op dat niveau (Van Houtven, Peters, & El Morabit, 2010; Peters & Van Houtven, 2010; De Wachter, Heeren, Marx et al., 2013; Deygers, Peters & Van den Branden, 2017). Onder meer Van den Branden stelt dat in het hoger onderwijs 'gedecontextualiseerd taalgebruik' veel vaker de norm is dan in het dagelijks taalgebruik en dat impliciete relaties tussen zinnen, complexere zinsconstructies en minder frequente woorden daar deel van uitmaken (Van den Branden, 2010).

Omwille van die problematiek wordt de taalvaardigheid van studenten soms bij het begin van hun hoger onderwijs carrière getoetst, zeker in onderwijssystemen waar de toegang niet verkregen wordt door middel van een voorafgaande test. Voorbeelden hiervan zijn Van Dyk (2015), Read (2015), De Wachter et al. (2013) en Heeren, Speelman en De Wachter (2021a). Zij tonen aan dat taalvaardigheid niet de enige, maar wel een significante factor is voor de slaagkansen van studenten. Die 'post-entry language assessments' (PELA's) worden na inschrijving afgenomen en dienen dus niet om de toegang tot het hoger onderwijs te reguleren, maar helpen universiteiten om studenten die mogelijk een verhoogd risico lopen vroegtijdig te detecteren en ondersteuning aan te bieden (Knoch & Elder, 2013; Read, 2015; Read, 2016).

In die context heeft de KU Leuven in 2010 een digitale taalvaardigheidstest ontwikkeld met de naam 'Taalvaardig aan de Start' (TaalVaST). Die toetst de beheersing van de Nederlandse taal in het hoger onderwijs op het gebied van woordenschat (afleiden van betekenis van abstracte concepten) en leesvaardigheid, met een tijdslimiet van 30 minuten. De resultaten binnen de KU Leuven Associatie tonen dat de test, die wordt afgenomen bij eerstejaarsstudenten, een doelgroep van mogelijke risicostudenten selecteert die problemen hebben met die academische taalvaardigheid (De Wachter & Heeren 2013; De Wachter et al. 2013; Heeren, Speelman, & De Wachter, 2021a). Die risicostudenten worden in een aantal faculteiten ook opgevolgd door middel van taalworkshops en digitale leerpaden, in een poging om hun taalvaardigheid bij te spijkeren.

Ook in Suriname kampen studenten bij AdeKUS (Anton de Kom Universiteit Suriname) met de beheersing van academische taalvaardigheid. In Suriname wordt net als in Vlaanderen het Nederlands als officiële taal gebruikt, en zoals in Vlaanderen is het Nederlands niet voor iedere student de (eerste) taal die thuis gesproken wordt. Veel studenten zijn met andere woorden 'meertalig'. Uit diverse (zelf)evaluatierapporten van de afdeling Kwaliteitszorg van AdeKUS, in het kader van accreditatietrajecten, en uit vakevaluaties, blijkt dat docenten van de diverse opleidingen van de Faculteit der Maatschappijwetenschappen zich zorgen maken over het taalniveau van de studenten en de invloed die dit heeft op de resultaten op toetsen en ingeleverde papers. Een vroege begeleiding van studenten die een laag taalniveau hebben, lijkt dus ook in Suriname gewenst.

De Leuvense TaalVaST-toets kan daarom ook voor Suriname relevant zijn, zeker als de toets erin zou slagen om in een vroeg stadium studenten te identificeren die extra begeleiding nodig hebben op het gebied van taal. Er is bij AdeKUS een pilotafname van de test uitgevoerd die toonde dat de resultaten van de Surinaamse studenten gemiddeld veel lager lagen dan de resultaten van de Belgische studenten (Matawlie, Pelgrim, Heeren, & De Wachter, 2021). Om te weten op welke manier die lagere resultaten geïnterpreteerd kunnen worden, moet de validiteit van de toets verder bekeken worden. Dit artikel zal dan ook focussen op dat ‘valideringsvraagstuk’ vanuit het argumentatief kader van Knoch en Elder (2013). Vanuit validiteitsoogpunt is het namelijk essentieel om na te gaan of de KU Leuven-toets even inzetbaar is in de Surinaamse context. Dat is dan ook het centrale doel in dit artikel: achterhalen of de Leuvense TaalVaST-toets ook valide is in de Surinaamse context. Omdat de verschillende stappen in het validiteitsraamwerk op elkaar voortbouwen, focust dit artikel op de eerste stappen in het argument, waarin wordt onderzocht of de testscore een adequate weerspiegeling is van het geobserveerde testgedrag, of de scores consistent zijn en, hoewel in mindere mate, of de test aansluit bij het beoogde construct (Knoch & Elder, 2013). Daarom onderzoekt deze studie de psychometrische kwaliteiten van de test en de testperceptie door enerzijds studenten en anderzijds docenten. Op die manier moet blijken of de test op dezelfde manier werkt als aan de KU Leuven en of de resultaten op dezelfde manier geïnterpreteerd kunnen worden. Immers, zoals Chapelle, Jamieson en Hegelheimer (2003) terecht opmerken, valideringsvraagstukken moeten telkens worden afgestemd op de specifieke context waarin ze worden uitgevoerd.

2 Theoretisch kader

2.1 Validiteit als complex begrip

Chapelle (2011) geeft aan dat validiteit als begrip oorspronkelijk werd gedefinieerd als de mate waarin een test meet wat hij beoogt te meten. Hoewel dat zeker klopt, is het moeilijk om dit in de praktijk om te zetten en te bepalen wanneer er voldoende indicaties zijn om te concluderen dat de test meet wat hij beweert te meten (Chapelle, 2011). Een invloedrijk validiteitskader dat deze vraag probeert te stroomlijnen is het argumentgebaseerde valideringsdenken geïntroduceerd door Kane (1992). Bij argumentgebaseerde validering worden de geldigheid van de interpretatie van testresultaten en de acties die daaraan worden verbonden, getoetst door na te gaan of de claims onderliggend aan de test en het gebruik ervan te rechtvaardigen zijn. Bekende argumentgebaseerde validiteitskaders gebruiken het argumentatief model van Toulmin (2003). Voorbeelden van dergelijke argumentatieve kaders die gebruikmaken van Toulmin

raamwerk zijn Bachman en Palmer (2010), alsook Kane (1992; 2006; 2013). In Toulmin's model worden claims die gemaakt worden op basis van data, gegeven feiten in de werkelijkheid, expliciet uitgewerkt door onderzoekers. De opdracht is dan om bewijsmateriaal te verzamelen die de claims rechtvaardigen, of die deze net onderuit halen of nuanceren. Kane (1992) geeft aan dat een onderzoeker of testgebruiker start met het lineair uitleggen van de verschillende claims in een argument dat dan vervolgens kan worden gestaafd met bewijsmateriaal en uiteindelijk kan worden getoetst. Vaak wordt teruggesproken op de visuele voorstelling van Kane, Crooks en Cohen (1999).

Figuur 1

De bruganalogie naar Kane, Crooks en Cohen (1999)



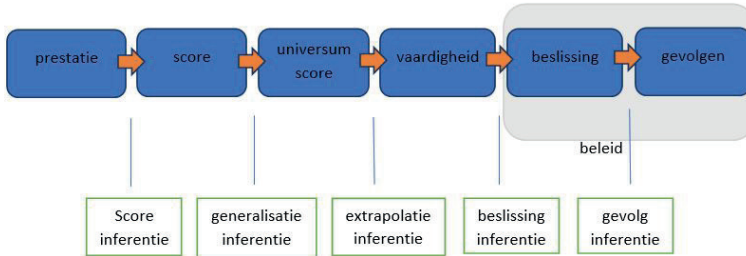
Zij zien de verschillende claims als bouwblokken die verbonden worden met inferenties die een soort brug vormen tussen de bouwblokken en die moeten worden gestaafd met bewijsmateriaal. Bewijs voor de geldigheid van een inferentie vormt dan de data waarop de volgende claim rust. Als een claim moet worden genuanceerd of zelfs onderuit wordt gehaald, dan heeft dat gevolgen voor alle daarop volgende bouwblokken. De eerste brug onderbouwt de geldigheid van de toekenning van scores op basis van de taken die studenten uitvoeren, de tweede bekijkt de veralgemening van die score en de derde de mate waarin de score geëxtrapoleerd kan worden naar het reële doeldomein.

Het raamwerk van Knoch en Elder (2013), dat ontwikkeld werd specifiek voor taaltoetsen na instroom, baseert zich met de eerste vier 'blokken' op het model van Kane (1992), dat zich meer toespitst op de interpretatie van testcores. Knoch en Elder koppelen daaraan de laatste twee blokken, geïnspireerd vanuit het model van Bachman en Palmer (2010), dat zich ankert in testgebruik en -gevolgen. Figuur 2 geeft het model van Knoch en Elder (2013) weer.

De pijlen tussen de bouwstenen staan voor de te staven inferenties. Als alle rechtvaardigingen voor de claims onderliggend aan de inferenties degelijk gestaafd zijn, kan er gesproken worden van een coherent argument. Als het bewijs voor een bepaalde brug zwak is of zelfs onderuit wordt gehaald, dan heeft dat, zoals eerder aangegeven, gevolgen voor de geldigheid van de daarop volgende inferenties (Kane et al., 1999). De brug van de prestatie van studenten op een test naar de score wordt ook wel de score-inferentie of evaluatie-inferentie genoemd. De claim die moet worden ondersteund is dan dat de score op de test een geldige weergave is van de prestatie van de studenten. De volgende brug

Figuur 2

Inferenties en claims: stappenplan voor validering van Knoch en Elder (2013)



staat dan voor de ‘generalisatie’-inferentie. Daarin wordt gekeken of de score die een student krijgt, losstaat van het moment waarop die test wordt afgenomen of, in het geval dat er verschillende testversies naast elkaar bestaan, van de testversie die wordt gehanteerd. In de volgende stap moet de extrapolatie-inferentie worden ondersteund. Daarbij wordt de samenhang met de taken die studenten moeten uitvoeren in de effectieve taalcontext onderzocht. Knoch en Elder (2013) vermelden hierbij ook de theoretische dimensie, wat wil zeggen dat ook kan worden onderzocht of de testresultaten overeenkomen met de verwachtingen op basis van een theoretisch taalmodel. De beslissing-inferentie gaat na of de beslissingen op basis van het testresultaat duidelijk zijn en de gevolg-inferentie staft uiteindelijk of de gevolgen positief zijn voor alle betrokken partijen. Knoch en Elder (2013) plaatsen deze laatste twee in de beleidscontext van de instelling. Net als bij Kane, Crooks en Cohen (1999) wijzen ook Knoch en Elder (2013) op het feit dat de argumentatieve bouwblokken sequentieel op elkaar volgen.

Knoch en Elder (2013) geven daarnaast ook aan dat bij de dataverzameling de input van betrokkenen bij het testproces waardevolle bewijspistes kan vormen. In navolging van het idee van meervoudige bewijslijnen geopperd door Messick (1989), speelt de perceptie van gebruikers een cruciale rol in het validiteitsargument (Fox & Cheng, 2015; Iwashita & Elder, 1997). Problemen met de testprocedures, de testomstandigheden of een slecht begrip van de instructies, kunnen ervoor zorgen dat de scores geen correcte weerspiegeling zijn van het gewenste testgedrag, maar van construct-irrelevante factoren. Het onderzoeken van testpercepties laat ook toe om te polsen naar de representativiteit van scores en het nut ervan voor de betrokkenen. Representativiteit staat dan voor de mate waarin het taalgebruik en de taken in de test die uit het doeltaaldomein weerspiegelen. Weigle en Malone (2016) wijzen erop dat taaltesten voor academische doeleinden, ondanks de beperkingen inherent aan het testformat, toch de authentieke taal en taaltaken van de academische context moeten weerspiegelen. Het ‘nut’, wat Bachman en Palmer (1996) ‘impact’

noemen, doelt op de mate waarin testers de informatie die de test hun geeft als een meerwaarde ervaren voor hun educatieve carrière. Hoewel deze perceptie-aspecten verder gaan dan de eerste bouwblokken van het argument, kunnen ze een waardevol bijkomend perspectief geven bovenop de psychometrische data (Iwashita & Elder, 1997). Ze kunnen bijvoorbeeld mogelijke verklaringen bieden voor bevindingen die naar voor komen in de psychometrische analyses. Zo kan het zijn dat studenten de test niet aanvaarden als representatief voor hun context, waardoor ze hem niet oprecht invullen. Ook als het moeilijkheidsniveau van vragen niet overeenkomt met het vaardigheidsniveau van de studenten, kunnen interpretaties van betrokkenen inzicht bieden in de manier waarop de items al dan niet aansluiten bij de vaardigheid van de studenten en waar de mogelijke discrepanties zitten.

2.2. Onderzoeksvragen

De afname van de taaltoets in de Surinaamse context wordt geëvalueerd vanuit de eerste stappen uit het argumentatief validiteitsraamwerk van Knoch en Elder (2013). Het gaat hier immers om een pilot, waarbij eerst de geldigheid van de eerste bouwstenen moet worden onderzocht, vooraleer volgende stappen aan bod kunnen komen. Desalniettemin wordt er via de testperceptie ook al gepolst naar het nut en de representativiteit van de test. Om de eerste inferenties te kunnen onderbouwen, moet de stelling dat de score een valide reflectie is van de taken die de studenten uitvoeren, gestaafd worden met bewijzen. We volgen hierbij Knoch en Elder (2013) door verschillende perspectieven te combineren. De overkoepelende onderzoeksvraag is daarbij

In welke mate is de testscore een betrouwbare en correcte afspiegeling van de testprestatie van de studenten?

Om deze stelling te staven, geven Knoch en Elder (2013) drie pistes aan waarvoor bewijsmateriaal moet worden gezocht om de hoofdclaim te kunnen rechtvaardigen. Ten eerste moet worden aangetoond dat de administratieve en logistieke procedures correct verlopen zijn. Ten tweede moet worden aangetoond dat de instructies bij de taken duidelijk zijn en ten derde moet worden aangetoond dat de test op een gepast moeilijkheidsniveau functioneert. Dat laatste houdt ook in dat de test een duidelijk onderscheid maakt tussen sterke en minder sterke studenten. Naast deze pistes wordt ook bekeken in welke mate de test betrouwbaar is en tot slot ook in welke mate de testgebruikers het materiaal en de test als representatief en nuttig beschouwen. Hoewel deze laatste vragen strikt genomen bij de extrapolatie- en zelfs beleidsinferentie horen, zoals hierboven reeds is toegelicht, is het toch nuttig om deze aspecten al vanaf het begin van het valideringsproces mee te nemen omdat ze bijkomende informatie kunnen leveren die ook voor de interpretatie van bevindingen in de eerste bouwstenen van belang kunnen zijn.

We delen de hoofdvraag daarbij op in vier subvragen die verbonden zijn aan drie deelstudies:

1. Hoe is de samenhang tussen de items en maken de items voldoende onderscheid tussen sterkere en minder sterke studenten?
2. Is de test betrouwbaar?
3. Zijn de testprocedures en instructies duidelijk voor alle studenten en beschouwen zij de test als representatief en nuttig?
4. Beschouwen de docenten de testitems als representatief en nuttig?

In de discussie trianguleren we de resultaten van de itemanalyses, de studentenbevraging en de bevraging van een representatieve groep docenten om zo een genuanceerder beeld te krijgen van de technische kwaliteit van de test in combinatie met de perceptie van de test en de testafname door de verschillende betrokken partijen (Fox & Cheng, 2015).

3 Participanten, materiaal en methode

3.1 Participanten

De participanten zijn onderverdeeld in twee groepen, te weten de studenten en de docenten. Er is besloten vanuit beide perspectieven te kijken naar de test, om een genuanceerder beeld te krijgen van hoe de test gepercipieerd wordt. Eerst zal de groep van de studenten besproken worden, daarna de docenten.

Studenten in Suriname konden zich vrijwillig aanmelden voor een testafname. 112 eerstejaarsstudenten van diverse opleidingen binnen de Faculteit der Maatschappijwetenschappen hebben de taalvaardigheidstest online afgelegd in een gesurveilleerde omgeving in de computerfaciliteit van de universiteit. Het betreft een gelegenheidssample van opleidingen binnen de Faculteit der Maatschappijwetenschappen, waarbij taal een belangrijke factor is. 109 van de studenten hadden nog niet eerder een opleiding op HBO- of WO-niveau gevolgd; zij vormen de populatie van dit onderzoek. Het aantal eerstejaarsstudenten aan de Faculteit der Maatschappijwetenschappen bedroeg 278 studenten. De steekproef bedroeg dus ruim 39% van de totale populatie. Een overzicht van de achtergrondgegevens van deze studenten is terug te vinden in Tabel 1.

Alle voltijdse docenten van de Anton de Kom Universiteit in Suriname zijn aangeschreven voor het invullen van een online vragenlijst over de test en de verschillende testitems. Er hebben 12 docenten meegedaan, waarvan 9 van de faculteit der Maatschappijwetenschappen, 2 van de Medische faculteit en 1 van de Technische faculteit. Hun mening over verschillende aspecten van de test en de perceptie van de moeilijkheidsgraad voor de student kan ook meer inzicht geven in de constructvaliditeit van de test. Alle docenten op een na hebben het Neder-

Tabel 1

Achtergrondgegevens eerstejaars studenten Faculteit der Maatschappijwetenschappen

Achtergrondgegevens studenten		N
Opleiding	Bedrijfskunde	11
	Psychologie	29
	Agogiek & Onderwijs	15
	Economie	12
	Public Administration	44
	Sociologie	1
Thuis taal enkel Nederlands	Ja	32
	Nee	80
Geslacht	Man	20
	Vrouw	92
Leeftijd	18-19	29
	20-21	48
	22+	35

Tabel 2

Achtergrondgegevens docenten Anton de Kom Universiteit Suriname

Achtergrondgegevens docenten		N
Verbonden aan Faculteit	Maatschappij Wetenschappen	9
	Medische Wetenschappen	2
	Technologische Wetenschappen	1
Jaren ervaring	0 tot 5	2
	5 tot 10	7
	10 tot 15	2
	15+	1
Eigen opleiding genoten in	Amerika	1
	België	1
	India & Costa Rica	1
	Nederland	3
	Suriname	6
Thuis taal NL	Ja	11
	Nee	1

154

PEDAGOGISCHE
STUDIËN
<https://doi.org/10.59302/ps.v100i2.14221>

lands als thuistaal en 50% heeft in Suriname de opleiding genoten en 50% in het buitenland. Van de docenten die in het buitenland studeerden, heeft de helft haar opleiding in Nederland gedaan (25% van het totaal) en de helft elders. Een overzicht van de achtergrondgegevens van deze docenten is terug te vinden in Tabel 2.

3.2 Materiaal

De gebruikte taaltest meet de bekwaamheid van starters om binnen een beperkte tijdspanne een set items te beantwoorden over algemene academische woordenschat en academische leesteksten die representatief zijn voor het eerste jaar aan de universiteit. Naast taalkennis op basis van complexe taalconstructies worden ook strategische vaardigheden gemeten (De Wachter & Heeren, 2013). Strategische vaardigheden omvatten het kunnen leggen van verbanden en het kunnen onderscheiden van hoofd- en bijzaken. Met behulp van woordfrequenties en leesbaarheidsindexen zijn de teksten en woorden afgestemd op het taalniveau van een eerstejaarsstudent aan de universiteit. De test is al uitvoerig onderzocht in de Belgische context (De Wachter et al., 2013; Heeren et al., 2021a; Heeren, Speelman, & De Wachter, 2021b).

De test gebruikt zeven verschillende itemtypes, te weten: het vinden van de correcte synoniemen van een woord in een of meerdere zinnen (synoniem); het selecteren van het correcte woord in een langere tekst (dropdown woord in tekst); het selecteren van een woord dat past in drie verschillende zinnen (1 woord voor 3 contexten); het neerschrijven van een gegeven woord in een andere woordvorm (woordvorm); ‘scrambled’ tekst die in de logische volgorde moet worden geplaatst (alineaoopbouw drag and drop); lezen op tekststructuur (lezen op structuur); en lezen op betekenis waarbij studenten de correcte of incorrecte uitspraak bij een tekstfragment moeten aanduiden (lezen op betekenis). In Bijlage 1 zijn enkele voorbeelditems opgenomen.

Na de testafname kregen de studenten nog een vragenlijst over hun ervaring van de testprocedure en hun perceptie van de representativiteit en het nut van de test. De vragenlijst werd ontwikkeld door Heeren, Speelman en De Wachter en beschreven door dezelfde auteurs (2021c). De vragen werden grotendeels gebaseerd op eerdere perceptieonderzoeken naar PELA's (Knoch, Elder en O'Hagan, 2016; Yan Thirakunkovit, Kauper, & Ginther, 2016). Hiermee wordt de perceptie van de student in kaart gebracht. De vragenlijst bevat een totaal van 25 vragen (zie Bijlage 2), die onderverdeeld zijn in vragen die betrekking hebben op de periode voor de testafname, tijdens de testafname en na de testafname. De antwoordopties bestaan telkens uit een Likertschaal met vier opties. De eerste Likertschaal is gebruikt voor de stellingen over de testafname en de test (helemaal eens, eerder eens, eerder niet eens, helemaal niet eens) en de tweede voor de bevraging van de moeilijkheid van de test en de items (zeer moeilijk, eerder moeilijk, eerder gemakkelijk, zeer gemakkelijk).

De docenten van de Anton de Kom Universiteit Suriname zijn aangeschreven voor het invullen van een online vragenlijst over de test en de verschillende test-items. Hun mening over verschillende aspecten van de test en de perceptie van de moeilijkheidsgraad voor de student kan ook meer inzicht bieden in hoe de tekst gepercipieerd wordt door verschillende stakeholders. De vragenlijst (Google forms) werd ontwikkeld door de auteurs en bevat eerst een aantal stellingen per itemtype telkens met een concreet voorbeeld uit de test, daarna volgen drie algemene vragen. De docenten kunnen daarbij antwoorden via een Likertschaal met vier opties (helemaal eens, eerder eens, eerder niet eens, helemaal niet eens). De vragenlijst stelt de docenten in de gelegenheid een oordeel te geven over hun verwachting over het succesvol kunnen afleggen van de test door de studenten. Verder kunnen de docenten feedback geven over eventuele door hen gewenste aanpassingen. De volledige vragenlijst is opgenomen in Bijlage 3.

3.3 Methode

Methodologisch valt dit onderzoek uiteen in drie deelstudies, namelijk de test-interne kenmerken, de perceptie van de studenten en de perceptie van de docenten. De eerste is gebaseerd op een itemanalyse, de tweede op een bevraging van de studenten en de laatste op een bevraging van de docenten.

Met het bekijken van de testinterne kenmerken wordt geanalyseerd of de items op het juiste moeilijkheidsniveau zijn en of ze voldoende onderscheid maken tussen sterkere en minder sterke studenten. De moeilijkheid van items, gemeten als het percentage correcte antwoorden, schommelt bij dit soort testen bij voorkeur tussen 20% en 80% (Weideman, 2011). Het discriminerend vermogen, gemeten als de item-totaal correlatie, valt bij voorkeur niet onder de 25% (Fulcher & Davidson, 2007).

Bij de perceptie van de test door de studenten wordt gekeken naar hoe ze de test en de afname ervan hebben ervaren. De vragenlijst betreffende de perceptie van de studenten is afgenomen nadat de studenten de taalvaardigheidstest hadden afgerond en hun resultaten hadden gezien. Hun oordeel kan immers inzicht geven in construct-irrelevante factoren die het testresultaat beïnvloeden (zoals achtergrondlawaai, technische problemen en onduidelijke instructies), wat de validiteit van de test mee kan beïnvloeden (Knoch & Elder, 2013). Daarnaast is het ook belangrijk dat de studenten de test als representatief en nuttig beschouwen en dat hun score betekenisvol is (Read, 2008). Voor de analyses worden de frequenties van de antwoorden van de studenten gebruikt.

Door middel van een aan de docenten voorgelegde vragenlijst, wordt de perceptie van de docenten bekeken. Docenten hebben voor iedere itemcategorie uit de taaltoets een voorbeeld gekregen, vervolgens is gevraagd naar hun oordeel over de test (via Google forms).

Aspecten als duidelijkheid van de instructie, passend bij het bestaande en gewenste niveau van taalgebruik van de studenten zijn bevroegd, alsook of de taak passend was voor het gewenste en het huidige niveau van de studenten. Door de kleine aantallen worden de resultaten van deze studie descriptief besproken.

4 Resultaten

4.1 Itemanalyse en betrouwbaarheid

De studenten scoorden gemiddeld 38.4% op de test, waarbij de laagste score 12% was en de hoogste 76%. De itemanalyses geven verder aan dat op itemniveau het aantal studenten dat de items goed kon maken, varieerde van 1.9% tot 84%. Een vijfde van de items (5 van de 25) bleek door minder dan 20% van de studenten correct gemaakt te zijn. Twee items uit het itemtype '1 woord voor 3 contexten' werden het slechtst gemaakt. Slechts 1.9% en 2.8% scoorde deze twee items correct.

Om de verschillende itemtypes met elkaar te kunnen vergelijken, in tegenstelling tot de losse items, zijn ook de gemiddelde scores per itemtype berekend. Hieruit bleek dat de studenten met name moeite hadden met alineaoopbouw (12.3% had dit goed) en 1 woord voor meerdere contexten invullen (17.7% had dit goed). Lezen op structuur (61.8% goed) en synoniemen (52.6% had dit goed) gingen hen aanzienlijk makkelijker af. Een overzicht per itemtype is te vinden in Tabel 3.

157

PEDAGOGISCHE
STUDIËN
<https://doi.org/10.59302/ps.v100i2.14221>

Tabel 3

Overzicht type vragen met gemiddeld percentage antwoorden goed en perceptie moeilijkheidsgraad door student

Itemtype	Gemiddeld percentage goed	Percentage moeilijk
Lezen op structuur	61.8	44.6
Synoniem	52.6	60.2
Woordvorm	43.1	49.4
Lezen op betekenis	35.9	69.9
Dropdown woord in tekst	34.7	81.9
1 woord voor 3 contexten	17.7	83.1
Alineaoopbouw drag and drop	12.3	51.8

De Cronbach's alpha voor de hele vragenlijst was 0.60. Dat is dus een relatief lage betrouwbaarheid aangezien 0.70 vaak als ondergrens wordt gezien voor een aanvaardbare betrouwbaarheid (Fulcher & Davidson, 2007). Om het discriminerend vermogen van de items nader te bekijken, is naar de correlatie van de items met de totale score gekeken. Die geeft aan hoe goed een item onder-

Validering van een academische taalvaardigheidstoets in de Surinaamse context: test scores en testpercepties

W. Matawlie, S. Pelgrim, J. Heeren en L. De Wachter

scheid maakt tussen studenten met een hoge score en een lage score. Twee items hadden een negatieve samenhang met de totale score, terwijl deze items wel door bijna de helft van de deelnemers goed werd gemaakt (lezen op structuur en synoniemen, respectievelijk -0.171 en -.037). Bij het verwijderen van deze 2 items stijgt de Cronbach alpha licht naar respectievelijk 0.635 en 0.618. Dit is nog steeds laag. Nog een item (dropdown woord in tekst) heeft een zeer lage samenhang met de totaalscore (0.018). Het verwijderen van dit item leidt tot een Cronbach alpha van 0.610. De overige 22 items laten een redelijke samenhang en betrouwbaarheid zien.

Op basis van het voorgaande kan dus wel gesteld worden dat het eerder lage discriminerende vermogen van een aantal items leidt tot een relatief lage betrouwbaarheid van de test in zijn geheel in de Surinaamse context. Het percentage correcte antwoorden, hun item-totaal correlatie en Cronbach's alpha wanneer dit item wordt weggelaten, kunnen in Tabel 4 afgelezen worden.

Tabel 4

Overzicht itemnummers uit de databank, met percentage goede antwoorden, item-totaal correlatie en itemtypes

Itemnummer	Itemtype	Percentage goed	Item-totaal correlatie	Cronbach's alpha
139.1	Woordvorm	84.0	0.156	0.591
97.2	Woordvorm	74.5	0.224	0.584
156.1	Lezen op structuur	68.9	0.126	0.596
153.2	Dropdown woord in tekst	62.3	0.245	0.580
121.1	Synoniem	60.4	0.194	0.587
73.1	Synoniem	54.7	0.177	0.589
118.1	Lezen op structuur	54.7	0.200	0.586
159.1	Synoniem	50.9	-0.037	0.618
76.1	Lezen op structuur	46.2	-0.171	0.635
91.1	Synoniem	44.3	0.078	0.603
115.1	1 woord voor 3 contexten	43.3	0.327	0.569
67.1	Lezen op betekenis	42.5	0.186	0.588
153.3	Dropdown woord in tekst	40.6	0.018	0.610
97.1	Woordvorm	38.7	0.339	0.567
153.1	Dropdown woord in tekst	30.2	0.087	0.600
103.1	Lezen op betekenis	29.2	0.152	0.592
127.1	Woordvorm	26.4	0.302	0.574
124.1	1 woord voor 3 contexten	22.6	0.228	0.583
151.1	Woordvorm	22.6	0.485	0.553
168.1	Dropdown woord in tekst	20.8	0.308	0.574
85.1	Dropdown woord in tekst	19.8	0.427	0.561

Itemnummer	Itemtype	Percentage goed	Item-totaal correlatie	Cronbach's alpha
38.1	Woordvorm	12.3	0.311	0.577
109.1	Alineaopbouw drag and drop	12.3	0.217	0.586
64.1	1 woord voor 3 contexten	2.8	0.053	0.599
165.1	1 woord voor 3 contexten	1.9	0.159	0.595

4.2 Studentenperceptie

Als we kijken naar de perceptie van de studenten, dan komt dat overeen met de itemanalyse wat de moeilijkheid van de items betreft. Wanneer studenten de moeilijkheid van de verschillende itemtypes aangeven, dan werd het itemtype '1 woord voor 3 contexten' als 'moeilijk tot redelijk moeilijk' beoordeeld door meer dan driekwart (83.1%) van de studenten. Ook voor het item-type 'dropdown woord in tekst' geeft een groot deel (81.9%) aan dit moeilijk te vinden. Minder moeite hebben ze met lezen op structuur en woordvorm: respectievelijk minder dan de helft (44.6%) en de helft (49.4%) geeft aan hier moeite mee te hebben.

Tabel 5

Overzicht stellingen en het percentage studenten dat het eens is (helemaal eens en eens) met deze stellingen
(n = 112)

Vraag	Percentage eens
Test moeilijk	89.2
Geen technische problemen	88.0
Nut test duidelijk	85.5
Voldoende computerkennis	83.1
Inzicht academische leesvaardigheid	83.1
Instructies duidelijk	81.9
Weinig achtergrondlawaai	81.9
Oprachten weerspiegelen vaardigheden opleiding	77.1
Feedback score nuttig	74.7
Feedback score duidelijk	73.5
Taalgebruik representatief voor opleiding	69.9
Testscore makkelijk te interpreteren	65.1
Correcte weerspiegeling academische leesvaardigheid	56.6
Duidelijk hoe test zou verlopen	44.6
Zenuwachtig voor test	30.1

Uit de bevraging van de studenten blijkt ook dat de meesten voorafgaand aan de test niet zenuwachtig waren (slechts 30.1% zenuwachtig) en tijdens de test niet veel hinder ondervonden van technische problemen (88%), gebrekkige computervaardigheden (83.1%) of achtergrondlawaai (81.9%). Een ruime meerderheid vond de instructies bij de vragen duidelijk (81.9%). Hoe de testafname zou verlopen bij de start van de testsessie was echter minder duidelijk (44.6%). Zoals te verwachten op basis van de itemanalyses vonden de studenten de test zeer moeilijk (89.2%). Toch geeft meer dan driekwart aan dat het instrument inzicht geeft in hun academische taalvaardigheid (83.1%) en dat de opdrachten in de test het taalgebruik van de opleiding weerspiegelen (77.1%). De meeste studenten gaven ook aan dat de taal representatief was voor de taal in de opleiding (69.9%). Slechts iets meer dan de helft (56.6%) geeft aan dat de test een correcte weergave geeft van hun eigen academische taalvaardigheid. In Tabel 5 staat een overzicht van de stellingen met de bijbehorende percentages van studenten, die het eens en eerder eens waren met deze stellingen.

4.3 Docentenperceptie

Ook de docenten hebben een beoordeling gegeven over de verschillende itemtypes. Hierbij is steeds gekeken naar de duidelijkheid van de instructie, passendheid van het taalgebruik en de taak voor het gewenste en huidige niveau van de studenten. De docenten beoordelen het itemtype 'juiste synoniem' op alle ondervraagde aspecten positief in tegenstelling tot het itemtype 'woordvorm'. De antwoorden lopen iets meer uit elkaar bij 'duidelijkheid van de instructie' en 'representativiteit van de taak'.

Meer concreet lijken de docenten de instructie bij de vragen wel duidelijk te vinden, het taalgebruik wordt wisselend representatief gezien voor het huidige niveau, maar zeker wel representatief voor het gewenste niveau. Ook bij de taak wordt hetzelfde beeld gezien; men vindt het wisselend representatief voor het huidige niveau van de eerstejaarsstudenten, maar wel representatief voor het gewenste niveau. De enige uitzondering hierbij is het itemtype alineaopbouw: hier ziet men het de taak wel als representatief voor het huidige niveau, maar vindt men de taak minder gepast voor het gewenste niveau. In Tabel 6 is de docentenperceptie per itemtype opgenomen.

Bij de algemene vragen geeft driekwart van de docenten aan dat de test een correcte weerspiegeling is van de te verwachten taalvaardigheid van de studenten. Opvallend genoeg vindt geen enkele ondervraagde docent dat het academisch taalniveau van de eerstejaarsstudenten op AdeKUS voldoende is. Voor het overzicht in tabelvorm wordt verwezen naar Tabel 7.

Tabel 6

Docentenperceptie per itemtype: percentage docenten eens (helemaal eens en eens) met de stellingen
(n = 12)

Itemtype	Duidelijke instructie	Taalgebruik representatief huidig niveau	Taalgebruik representatief gewenst niveau	Taak representatief huidig niveau	Taak representatief gewenst niveau
Woordvorm	50.0	66.6	75.0	58.3	83.4
1 woord 3 contexten	83.4	75.0	83.3	66.7	83.4
Lezen op betekenis	91.7	58.3	83.4	66.7	91.7
Juiste synoniem	91.7	91.7	91.7	91.7	91.7
Lezen op structuur	91.7	76.0	91.7	91.7	91.7
Juiste woord	83.3	83.3	91.7	83.3	91.7
Alinea opbouw	75.0	83.3	91.7	83.3	75.0

Tabel 7

Overzicht stellingen betreffende de taalvaardigheid en het percentage docenten dat het eens (helemaal eens en eens) is met deze stellingen (n = 12)

Vraag	Percentage eens
Test correcte weerspiegeling te verwachten taalvaardigheid	75.0
Test correcte weerspiegeling huidig niveau taalvaardigheid	66.7
Academische taalvaardigheid eerstejaars voldoende	0.0

5 Conclusie en discussie

Voordat de resultaten geduid kunnen worden in een besluit, is het belangrijk als beperking aan te halen dat de studenten zich vrijwillig konden aanmelden en dat de test geen onderdeel was van een examen. Dat kan immers betekenen dat er een zekere selectiebias is op de populatie. Om deze mogelijke bias te vermijden, zou in een vervolgstudie de test bij de gehele groep studenten moeten worden afgenomen, bijvoorbeeld als vast onderdeel van de start van het curriculum. Om de overkoepelende onderzoeksvraag en tevens de eerste claims in het validiteitsraamwerk te beantwoorden, namelijk dat de score op de test een

Validering van een academische taalvaardigheidstoets in de Surinaamse context: testcores en testpercepties

W. Matawlie, S. Pelgrim, J. Heeren en L. De Wachter

betrouwbare weerspiegeling vormt van de prestaties van de studenten, werden drie deelstudies opgezet: (1) een studie naar de testinterne eigenschappen, namelijk de moeilijkheid en het discriminerend vermogen van de items en de betrouwbaarheid van de test; (2) de perceptie van de test en de testafname door de studenten; (3) de perceptie van de docenten.

De drie studies vertonen gelijklopende resultaten. De eerste studie naar de kenmerken van de testitems toont dat verschillende items niet correct gescoord worden door een meerderheid van de Surinaamse studenten en dat sommige items een kleine of zelfs licht negatieve samenhang met de testcores vertonen. Daarmee samenhangend bleek de betrouwbaarheid van de test als geheel, gemeten met Cronbach's alpha, ook eerder laag. Dat de test en de items moeilijk zijn, wordt bevestigd in de resultaten van de vragenlijst van de studenten en de docenten. Bij die laatste groep doet zich een interessante vaststelling voor: de docenten merken namelijk op dat de items te moeilijk zijn voor de groep studenten, maar dat ze wel aansluiten bij het niveau dat ze zouden verwachten van een eerstejaarsstudent. Ook de studenten zien de taal en taken in de test als representatief voor wat ze moeten doen, maar tegelijkertijd geven ze ook aan dat de test moeilijk is.

Verder blijkt uit de studie naar de perceptie van studenten dat er geen sterke indicaties zijn voor logistieke problemen of andere construct-irrelevante factoren zoals gebrekkige computervaardigheid of achtergrondlawaai bij de testadministratie die het testresultaat mogelijk beïnvloed hebben. Hoewel bijna de helft van de studenten aangeeft dat er meer toelichting over het testverloop nodig is bij de start van de test, geven ze wel aan, net als de docenten, dat de instructies bij de vragen duidelijk zijn. Wanneer de representativiteit van de test wordt bevraagd, blijken de studenten de taken en taal als representatief te beschouwen. Studenten vinden echter niet dat de test veel inzicht biedt in hun academische leesvaardigheid.

De bevraging van studenten en docenten geeft aan dat perspectieven van testgebruikers interessante data kunnen opleveren die complementair en verhelderend kunnen zijn samen met meer technische analyses (Fox & Cheng, 2015). Hoewel de inzichten van de bij de test betrokken partijen op zichzelf niet volstaan als geldig bewijs van testvaliditeit, kunnen ze wel meer inzicht geven in combinatie met andere bewijslijnen (Cheng & DeLuca, 2011). Zoals Bradshaw (1990, 27) aanhaalt over dit onderwerp: "These groups appear to have opinions to offer, and are willing to offer them". Ze kunnen mogelijk verklaringen vormen voor uitkomsten van technische en psychometrische analyses of, zoals in deze studie, mogelijke verklaringen voor patronen uitsluiten. Direct betrokkenen, en in het bijzonder de geteste personen zelf, zijn daarenboven de enigen die inzicht kunnen geven in hoe de testprocedure ervaren wordt (Bachman & Palmer, 1996), daarnaast kan het integreren van de stem van de direct betrokkenen beschouwd worden als een democratisch gebeuren (Shohamy, 2001).

Hoewel de testresultaten interessante inzichten bieden, blijkt de test niet zonder meer transfereerbaar of inzetbaar te zijn in een Surinaamse context. De studies geven aan dat aanpassingen aan de test nodig zullen zijn om de eerste inferentiebrug te kunnen nemen. De lokale context is inderdaad altijd belangrijk wanneer testen gevalideerd worden (Chapelle et al., 2003; Davies & Elder, 2005). De claim dat de resultaten een geldige weerspiegeling zijn, wordt vooral ondergraven door de moeilijkheidsgraad van de test die nu duidelijk te hoog ligt. Daaraan gekoppeld is ook de lage betrouwbaarheid van de test in de Surinaamse context opvallend en dat vereist nader onderzoek. Verder valt op dat de Surinaamse studenten, net als de docenten, de items weliswaar representatief vinden voor het academische taalgebruik binnen hun opleiding, maar dat ze het resultaat toch minder representatief vinden voor hun eigen academische taalvaardigheid. Ook dat verdient nadere introspectie, bijvoorbeeld door focusgesprekken te organiseren na de afname van de vragenlijst (zie bijvoorbeeld Bright & von Randow, 2004).

Zoals in het theoretisch kader is aangegeven: als het bewijs voor een bepaalde brug zwak is of zelfs onderuit wordt gehaald, dan heeft dat gevolgen voor de geldigheid van de daarop volgende inferenties (Kane et al., 1999). De claim die moet worden ondersteund is dan dat de score op de test een geldige weergave is van de prestatie van de studenten. Dat is dus op dit moment nog moeilijk hard te maken. Dat toont wel aan dat een argumentatieve aanpak, waarbij de validatiestappen expliciet gemaakt worden en op elkaar voortbouwen, een praktisch inzetbare methode is met duidelijke stappen. Het geeft testontwikkelaars met andere woorden "a place to start, guidance on how to proceed, or criteria for gauging the progress and deciding when to stop." (Kane, 2011, 8).

Dit onderzoek geeft duidelijk aan dat er nood is aan een verdere optimalisatie van de taalvaardigheidstest in de Surinaamse context, zodat op basis daarvan verdere stappen in het valideringsargument kunnen worden genomen om uiteindelijk het verband met studiesucces en de daarop volgende beleidsstappen te kunnen nemen. Ondanks gepercipieerde taal, cultuur en institutionele overeenkomsten, is de test dus niet zonder meer transfereerbaar of inzetbaar. Dat bevestigt de stelling van Chapelle et al. (2003) dat een verschil in context steeds moet worden meegenomen bij het valideren van een test. Op dit moment discrimineren de items niet voldoende, is hun samenhang te laag en is de test niet betrouwbaar genoeg. Er zullen nieuwe items moeten worden opgesteld, gebaseerd op Surinaamse academische teksten. Het doel van een PELA is immers dat bij de juiste groep studenten een gepaste interventie ingezet kan worden om zo het studiesucces van de instromende studenten te verhogen. Toekomstig onderzoek zal moeten uitwijzen of een aanpassing van de testitems een ander beeld oplevert. De aanpassingen aan de test kunnen het best gebeuren in samenspraak met de verschillende opleidingen van AdeKUS. Ook zou het interessant zijn om de test in te zetten in andere delen van het Caribisch gebied met een vergelijkbare onderwijscontext.

Literatuur

- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Bradshaw, J. (1990). Test-takers' reactions to a placement test. *Language Testing*, 7(1), 13-30.
- Bright, C., & von Randow, J. (2004). Tracking language test consequences: The student perspective. 18th IDP Australian International Education Conference, Sydney, Australia.
- Chapelle, C. A., Jamieson, J., & Hegelheimer, V. (2003). Validation of a web-based ESL test. *Language Testing*, 20(4), 409-439.
- Chapelle, C. A. (2011). Validation in Language Assessment. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning: Volume II* (pp. 717-730). London: Routledge.
- Cheng, L., & DeLuca, C. (2011). Voices from test-takers: Further evidence for language assessment validation and use. *Educational Assessment*, 16(2), 104-122.
- De Wachter, L., & Heeren, J. (2013). Een taaltest als signaal. De ontwikkeling en implementatie van een strategische taalvaardigheidstoets aan de KU Leuven. *Levende Talen Tijdschrift*, 14(1), 19-27.
- De Wachter, L., Heeren, J., Marx, S., & Huyghe, S. (2013). Taal: een noodzakelijke, maar niet de enige voorwaarde tot studiesucces. De correlatie tussen de resultaten van een taalvaardigheidstoets en de slaagcijfers bij eerstejaarsstudenten aan de KU Leuven. *Levende Talen Tijdschrift*, 14(4), 28-36.
- Deygers, B., Peters, E., & Van Den Branden, K. (2017). Checking assumed proficiency: Comparing L1 and L2 performance on a university entrance test. *Assessing Writing*, 32, 43-56.
- Elder, C., Bright, C., & Bennet, S. (2007). The role of language proficiency in academic success: Perspectives from a New Zealand university. *Melbourne Papers in Language Testing*, 12(1), 24-58.
- Fox, J., & Cheng, L. (2015). Walk a mile in my shoes: Stakeholder accounts of testing experience with a computer-administered test. *TESL Canada Journal*, 32, 65-86. <https://doi.org/10.18806/tesl.v32i0.1218>
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. London: Routledge.
- Heeren, J., Speelman, D., & De Wachter, L. (2021a). A practical academic reading and vocabulary screening test as a predictor of achievement in first-year university students: Implications for test purpose and use. *International Journal of Bilingual Education and Bilingualism*, 24(10), 1458-1473. DOI: 10.1080/13670050.2019.1709411
- Heeren, J., Speelman, D., & De Wachter, L. (2021b). Bepaalt taal wie het haalt? De samenhang tussen een academische taalvaardigheidscreening en het behalen van een bachelordiploma aan de universiteit. *Tijdschrift voor Hoger Onderwijs*, 39(1), 39-54.
- Heeren, J., Speelman, D., & De Wachter, L. (2021c). Post-entry language assessment in high-

- er education: Ensuring fairness by including the voice of the test-taker. *Collated papers for the ALTE 7th International Conference*, 45-48.
- Iwashita, N., & Elder, C. (1997). Expert feedback? Assessing the role of test-taker reactions to a proficiency test for teachers of Japanese. *Melbourne Papers in Language Testing*, 6(1), 53-67.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17-64). Westport: American Council on Education/Praeger.
- Kane, M. T. (2011). Validating score interpretations and uses: Messick lecture, Language Testing Research Colloquium, Cambridge, April 2010. *Language Testing* 29(1), 3-17.
- Kane, M. T. (2013). Validating the interpretation and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kane, M. T., Crooks, T., & Cohen, A. (1999). Validating Measures of Performance. *Educational Measurement: Issues and Practice*, 18(2), 5-17.
- Knoch, U., & Elder, C. (2013). A framework for validating post-entry language assessments (PELAs). *Papers in Language Testing and Assessment*, 2(2), 48-66.
- Knoch, U., Elder, C., & O'Hagan, S. (2016). Examining the validity of a post-entry screening tool embedded in a specific policy context. In J. Read (Red.), *Post-admission language assessment of university students* (pp. 23-42), Zwitserland: Springer.
- Matawlie, W., Pelgrim, S., Heeren, J., De Wachter, L. (2021). De inzet van een academische taalvaardigheidstoets voor eerstejaarsstudenten in het hoger onderwijs in Suriname: eerste resultaten en uitdagingen. In H. Neus, B. Bérénos et al. (Red.), *Taalkunde, literatuur en onderwijs: Nederlandse taal in contact (Vijfde CARAN-conferentie in Suriname)* (pp. 57-69). Paramaribo: Anton de Kom Universiteit.
- Peters, E., Van Houtven, T., & El Morabit, Z. (2010). Is meten echt meer weten? Taalvaardigheid van instromende studenten in het hoger onderwijs in kaart gebracht. In E. Peters & T. Van Houtven (Eds.), *Taalbeleid in het hoger onderwijs: de hype voorbij?* Leuven: Acco.
- Read, J. (2008). Identifying academic language needs through diagnostic assessment. *Journal of English for Academic Purposes*, 7, 180-190.
- Read, J. (2015). *Assessing English proficiency for university study*. London: Palgrave Macmillan.
- Read, J. (2016). Some key issues in post-admission language assessment. In J. Read (Ed.), *Post-admission language assessment* (pp. 3-20). Switzerland: Springer.
- Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18(4), 373-391.
- Toulmin, S.E (2003). *The uses of argument: Updated edition*. Cambridge University Press.
- Van den Branden, K. (2010). Taalbeleid in het hoger onderwijs. In E. Peters & T. Van Houtven (Eds.), *Taalbeleid in het hoger onderwijs: de hype voorbij?* (pp. 213-223). Leuven/Den Haag: Acco.
- Van Dyk, T. (2015). Tried and tested: Academic literacy tests as predictors of academic success. *Tijdschrift voor Taalbeheersing*, 37(2), 159-186. <https://doi.org/10.5117/>

TVT2015.2.VAND

- Van Houtven, T., Peters, E., & El Morabit, Z. (2010). Hoe staat het met de taal van studenten? Exploratieve studie naar begrijpend lezen en samenvatten bij instromende studenten in het Vlaamse hoger onderwijs. *Levende Talen Tijdschrift*, 11(3), 29-44.
- Weideman, A. (2011). Academic literacy tests: Design, development, piloting and refinement. *Journal for Language Teaching*, 45(2), 100-113.

Abstract

Validation of an academic language proficiency test in the Surinamese context: test scores and test perceptions

The study duration of students at Anton de Kom University Suriname is often longer than the nominal length of study. As research shows that language proficiency is a necessary (but not sufficient) condition for study success, it seems useful to test the academic language skills of first-year students to detect possible at-risk students. A digital academic language test developed at KU Leuven in Flanders (Belgium) was used as a starting point for this. To validate the test for the Surinamese context, 109 first-year students from the Faculty of Social Sciences took the test.

Test results show a much lower score for Surinamese students compared to results in Belgium. Therefore, it seems useful to investigate the validity of the test. This study uses the argumentative framework of Knoch and Elder (2013) as an overarching validation framework and investigates (1) the psychometric qualities of the test items; (2) test perception by the students and (3) perception by the teachers. By triangulating different methods and perspectives a clearer insight into the validity of the test for the Surinamese context can be gained. The results of the investigation show that even though the language of instruction in the Surinamese educational system is the same as in Flanders, it is clear that the test is not transferable as such and further research is needed.

Keywords academic language skills, validity research, test perception

166

PEDAGOGISCHE
STUDIËN

[https://doi.](https://doi.org/10.59302/ps.v100i2.14221)

[org/10.59302/](https://doi.org/10.59302/ps.v100i2.14221)

[ps.v100i2.14221](https://doi.org/10.59302/ps.v100i2.14221)

Auteurs

Widya Matawlie was voorheen universitair docent bij de studierichting Bedrijfskunde aan de Anton de Kom Universiteit Suriname en thans is zij verbonden als universitair docent aan de Financieel Economische Faculteit van de University of Aruba. **Suzanne Pelgrim** is universitair docent en promovendus bij de studierichting Psychologie aan de Anton de Kom Universiteit Suriname. **Jordi Heeren** is universitair docent academisch Nederlands aan KU Leuven in België. **Lieve de Wachter** is hoogleraar academisch Nederlands, Nederlandse taalbeheersing aan KU Leuven in België.

Correspondentieadres: Widya Matawlie, University of Aruba, J. Irausquin Plein 4, Oranjestad, Aruba. E-Mail: widya.matawlie@ua.aw

167

PEDAGOGISCHE
STUDIËN

<https://doi.org/10.59302/ps.v100i2.14221>

Bijlagen

Bijlage 1. Voorbeelditems van de taaltest

Voorbeelditem 1: Het juiste synoniem

Lees de twee zinnen. Wat is het meest correcte synoniem van het aangeduide woord?

“Het geheel van regels en technieken dat hier wordt gebruikt, is het verwijs-systeem dat momenteel aan het departement geschiedenis van de KU Leuven wordt gehanteerd en waarover een brede consensus bestaat.”

“De vergadering tussen vertegenwoordigers van de aftredende meerderheid over het bankgeheim is gisteren constructief verlopen. Er tekent zich een con-sensus af rond de opheffing van het bankgeheim.”

- A. verantwoordelijkheid
- B. gevoeligheid
- C. overeenstemming
- D. tevredenheid

Voorbeelditem 2: Alinea-opbouw

Lees de zinnen van een tekst. Ze staan niet in de juiste volgorde. Sleep ze naar hun juiste plaats. Twee zinnen staan al juist.

Titel: Hersenen verwerken 100.500 woorden per dag

A Zij zijn van mening dat er nieuwe hersencellen ontstaan wanneer er steeds meer informatie binnenkomt.

B Dat is maar liefst 34 gigabytes per dag, hetgeen overeenkomt met een vijfde van de opslagcapaciteit van een computer.

C Die toevloed van informatie zorgt er volgens veel wetenschappers voor dat onze hersenen overbelast worden.

D Dat aantal vinden we in de resultaten van een nieuw onderzoek terug.

E Daardoor zou zelfs de structuur van de hersenen veranderen. Dat alles kan tot concentratieproblemen leiden.

1 Gemiddeld hoort of leest een volwassene 100.500 woorden per dag.

2 ...

3 ...

4 ...

5 ...

6 ...

7 Al zien sommigen de structuurveranderingen wel als iets positiefs.

168

PEDAGOGISCHE
STUDIËN

[https://doi.](https://doi.org/10.59302/ps.v100i2.14221)

[org/10.59302/](https://doi.org/10.59302/ps.v100i2.14221)

[ps.v100i2.14221](https://doi.org/10.59302/ps.v100i2.14221)

Bijlage 2. Vragenlijst studenten: hoe ervaar jij de taaltest?

Voor de testafname

	Helemaal eens	Eerder eens	Eerder niet eens	Helemaal niet eens
1. Het was duidelijk hoe de test zou verlopen (duur, type vragen)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Het nut van deze test was voor mij duidelijk.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Voor de start van de test was ik zenuwachtig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Tijdens de testafname

	Zeer moeilijk	Eerder moeilijk	Eerder gemakkelijk	Zeer gemakkelijk
4. Ik vond de test	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Helemaal eens	Eerder eens	Eerder niet eens	Helemaal niet eens
5. De instructies bij de vragen waren duidelijk.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Mijn kennis van het gebruik van een computer was voldoende voor deze test.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Ik heb de manier waarop ik de test invulde aangepast doordat er een tijdlimiet was.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Het testsysteem vertoonde geen technische problemen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Er was weinig achtergrondlawaai waardoor ik me goed kon concentreren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. Ik heb vaak naar de resterende tijd gekeken.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Geef aan hoe moeilijk of gemakkelijk je de verschillende types vragen in de test vond

	Zeer moeilijk	Eerder moeilijk	Eerder gemakkelijk	Zeer gemakkelijk
11. Academische woorden: selecteer het woord dat past in de tekst uit een 'drop-down'-menu	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. Academische woorden: geef een synoniem voor het woord in het vet (multiple choice)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. Academische woorden: schrijf een woord in een andere vorm in een gegeven zin	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Validering van een academische taalvaardigheidstoets in de Surinaamse context: test scores en testpercepties

W. Matawie, S. Pelgrim, J. Heeren en L. De Wachter

14. Academische woorden: selecteer het woord dat in de drie contexten past	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15. Lezen op betekenis: Welke uitspraak is fout? (multiple choice)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16. Lezen op structuur: Welke structuur heeft de tekst? (multiple choice)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17. Lezen: sleep de zinnen tot een logische tekst	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Na de testafname

	Helemaal eens	Eerder eens	Eerder niet eens	Helemaal niet eens
18. Mijn testscore was gemakkelijk te interpreteren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19. Het resultaat dat ik behaalde op de test is volgens mij een correcte weerspiegeling van mijn academische leesvaardigheid.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20. Het taalgebruik in de test was volgens mij representatief voor het taalgebruik in de opleiding.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
21. De opdrachten die ik moest doen in de test weerspiegelen volgens mij dingen die ik ook zal moeten doen in mijn opleiding.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22. De tijdslimiet geeft volgens mij de verwachte tijdsdruk tijdens het academiejaar correct weer.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
23. De feedback bij mijn score was duidelijk.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24. De feedback bij mijn score was nuttig.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
25. De test heeft me inzicht gegeven in mijn academische leesvaardigheid.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Wil je nog iets kwijt over deze test?

170

PEDAGOGISCHE
STUDIËN

<https://doi.org/10.59302/ps.v100i2.14221>

Bijlage 3. Vragenlijst docenten

Algemene vragen aan begin:

1. Hoe lang geeft u al les op de universiteit?
2. Bij welke studierichting?
3. Wat is uw thuistaal (of welke taal wordt bij u thuis het meest gesproken)?
4. In welk land heeft u gestudeerd?
5. Aan welke instelling heeft u gestudeerd?

Vragen per item:

Kunt u aangeven in hoeverre u het met de volgende stellingen eens bent?

Vragen

1. De instructie bij de vraag is duidelijk.
2. Het taalgebruik bij de vraag is volgens mij representatief voor het huidige taalgebruik in de opleiding.
3. Het taalgebruik bij de vraag is volgens mij representatief voor het gewenste taalgebruik in de opleiding.
4. De taak die studenten moeten uitvoeren in dit item is volgens mij representatief voor het huidige academisch niveau van studenten in de opleiding.
5. De taak die studenten moeten uitvoeren in dit item is volgens mij representatief voor het academisch niveau dat verwacht wordt van studenten in de opleiding.

Antwoordmogelijkheden

Helemaal eens Eerder eens Eerder oneens Helemaal oneens
Toelichting mogelijk, Geef eventueel aanpassing

Algemene vragen aan eind:

Kunt u aangeven in hoeverre u het met de volgende stellingen eens bent?

Vragen

1. Deze test geeft een correcte weerspiegeling van de te verwachten academische taalvaardigheid
2. Deze test geeft een correcte weerspiegeling van de huidige academische taalvaardigheid van de studenten
3. De academische taalvaardigheid van eerstejaarsstudenten is voldoende

Antwoordmogelijkheden

Helemaal eens Eerder eens Eerder oneens Helemaal oneens
Toelichting mogelijk,

Heeft u nog opmerkingen over de taalvaardigheidstest, verbeterpunten, zaken die u mist?