

Taaltoetsen. Waarom, wat en hoe?

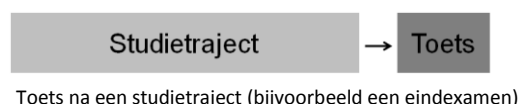
Bart Deygers & Sibo Kanobana

Instellingen van hoger onderwijs die een taalbeleid voeren, hanteren op een bepaald moment in het traject vaak een taaltoets. Die toets kan verschillende doelen dienen, maar heel vaak dient ze om taalzwakke studenten op te sporen om ze nadien te remediëren. In deze bijdrage gaan we dieper in op een aantal centrale vragen die iedere instelling van hoger onderwijs zich zou moeten stellen wanneer ze overweegt een taaltoets te gebruiken als diagnostisch instrument of als doorslaggevend argument in een deliberatie: waarom, wat en hoe toetsen?

Waarom?

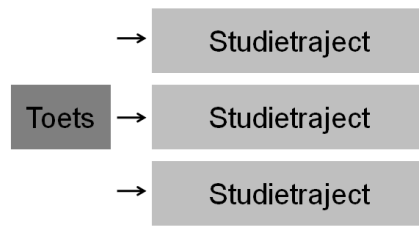
Waarom wil je als instelling een taaltoets gebruiken? Wat is het doel van de toets? Het zijn essentiële vragen die wel eens onbeantwoord blijven in het proces van toetsontwikkeling. Vaak wordt een toets afgenomen bij studenten omdat toetsen een gevoel van zekerheid bieden of omdat het zo lijkt te horen. Later in deze bijdrage zullen we zien dat de zekerheid van toetsen slechts relatief is en uit andere hoofdstukken in dit boek blijkt dat toetsen zelfs geen noodzakelijke component van een adequaat taalbeleid hoeven te zijn.

Het is mogelijk om toetsen ruwweg in twee categorieën in te delen, naargelang hun plaats in het leerproces. De meerderheid van de toetsen in het hoger onderwijs meten kennis en vaardigheden van studenten tijdens of na een studietraject. Er mag dus een min of meer gelijk studietraject bij de testpopulatie verondersteld worden en de te beheersen kennis en vaardigheden worden bepaald door wat er behandeld is in voorgaande lesmomenten.



Daartegenover staan toetsen die kennis of vaardigheden meten vóór een studietraject. Het doel hiervan kan zijn om studenten te stroomlijnen in bepaalde niveaugroepen (*placement tests*) of om de sterktes en zwakheden van studenten (groepen) in kaart te brengen (diagnostische toetsen¹). Het online toetsinstrument Dialang is een bekend voorbeeld van een op het ERK gemodelleerde diagnostische taalttest (Cf. Alderson & Huhta 2005). Er zijn opleidingen die Dialang gebruiken om een algemeen beeld te krijgen van het taalniveau van instromende studenten.

¹ Diagnostische toetsen kunnen natuurlijk ook tijdens het studietraject worden afgenomen om het onderwijs meer te kunnen afstemmen op de werkpunten van de individuele leerder.



Toets voor een studietraject

In het hoger onderwijs worden taaltoetsen vaak ingezet aan het begin van het academiejaar om te beslissen welke studenten nood hebben aan remediëring. Toch blijkt het niet evident te oordelen over de taalvaardigheid van studenten via plaatsingstoetsen, die een heel eigen aanpak en interpretatie vragen (Moussavi, 2002 en Bachman, 1990), al was het maar omdat deze toetsen een ander referentiekader hebben dan toetsen die tijdens of na een voor de hele testpopulatie uniform traject worden afgenomen. McNamara (2001) pleit, wegens de specifieke aanpak die diagnostische taaltoetsen vereisen voor verregaande voorzichtigheid bij het opstellen en afnemen ervan. Een studie van Green en Weir uit 2004 bevestigt dat oordeel. Hun onderzoek wijst uit dat de meerderheid van de door instellingen zelf gemaakte plaatsingstoetsen niet in staat zijn een accuraat oordeel te vellen over taalniveau van instromende studenten.

“[...] low-cost placement instruments may provide crude indications of students’ abilities but, unless supported by more extensive procedures, they are unlikely to provide the kind of detailed diagnostic information that is desirable for teachers and learners [...] practical considerations continue to dictate that placement tests do not provide the necessary information, even at the level of syntax.”

(Green & Weir 2004)

Er zijn verschillende redenen waarom een instaptoets kan falen. Er kunnen problemen zijn op het vlak van betrouwbaarheid. Mogelijk zijn de toetsdoelen niet adequaat gedefinieerd. Toetsen die worden opgesteld met een fout doel voor ogen, kunnen niet effectief zijn. Plaatsingstoetsen die worden opgesteld als een traditioneel terugblikkend examen zijn dat evenmin. Daarom is een plaatsingstoets in het hoger onderwijs op basis van de eindtermen van het secundair onderwijs mogelijk ineffectief. Door twee doelen te vermengen of door één type toets op te stellen volgens de regels van een ander type, weet men niet meer wat men nu eigenlijk aan het meten is.

Een probleem dat zich mogelijk kan voordoen is dat een geïntegreerde taaltoets wordt opgesteld en beoordeeld volgens de methodologie die verbonden is aan een elementgerichte methode. Iedere taaltoets vertrekt immers vanuit een theoretisch model met betrekking tot taal, ook construct genoemd. Het ene model ziet taal als een verzameling onderdelen (denk aan het model van Bachman & Palmer uit 1996 waarbij taalvaardigheid wordt opgesplitst in veertien deelaspecten), terwijl anderen (Weir 1990) een eerder geïntegreerd taalmodel propageren. De keuze voor een bepaald model heeft repercussies voor de toetsinhoud, die dan focust op onderdelen van taal (elementgericht, via woordenschat, grammatica, leesvaardigheid etc.) dan wel op taal als geheel (via taken die bijvoorbeeld verschillende vaardigheden integreren).

Of een taaltoets nu vooruit of achteruit blik en of ze nu elementgericht is dan wel geïntegreerd; de reden waarom getoetst wordt is meestal dezelfde. Via een taaltoets wil men conclusies kunnen trekken over iemands niveau van taalvaardigheid. Plaatsingstoetsen dienen de examinerator zo goed mogelijk in staat te stellen om te beslissen of de taalvaardigheid van de kandidaat goed genoeg is om te functioneren op een bepaald minimumniveau. Bij taaltoetsen in het kader van een taalbeleid zou dat minimumniveau vooraf vastgelegd moeten zijn. Zonder duidelijke definitie van het vereiste minimumniveau, kan men onmogelijk bepalen of iemand al dan niet het vereiste taalniveau haalt en wanneer men dat niet kan, schiet de test zijn doel voorbij.

Onderstaand beoordelingsinstrument van de *Occupational English Test* (OET) wordt gebruikt om te beslissen of mensen al dan niet kunnen instromen in een schakelprogramma ("bridging programme"). Zich al dan niet kunnen redden, is de minimumcompetentie waaraan iedere student wordt afgetoetst. De beoordelaars van de toets weten wat deze minimumcompetentie precies inhoudt.

Overall communicative effectiveness

- 1 Elementary level of communicative effectiveness
- 2 Clearly could not cope in a bridging programme in a clinical setting involving interactions with patients and colleagues
- 3 Just below minimum competence needed to cope in a bridging programme in a clinical setting involving interactions with patients and colleagues.
- 4 Has minimum competence needed to cope in a bridging programme in a clinical setting involving interactions with patients and colleagues.
- 5 Could easily cope in a bridging programme in a clinical setting involving interactions with patients and colleagues
- 6 Near native communicative effectiveness

Bron: McNamara, 2000

Het beoordelingsmodel van de OET wordt gebruikt in een placement test voor een voortgezette opleiding en dient dus een gelijksoortig doel als veel toetsen in het kader van een taalbeleid: het niveau van instromende studenten bepalen. Een antwoord op de waarom-vraag hangt dus samen met een hele reeks keuzes die de toets zelf zullen bepalen. Waarom men toetst heeft invloed op de inhoud van de toets (op basis van behandelde leerstof of leerstofonafhankelijk, zeer specifiek taalgebruik of algemeen taalgebruik) en op de manier waarop die toets wordt afgenomen (geïntegreerd en authentiek of via inferenties op basis van *discrete-point testing*). Hieronder gaan we dieper in op de volgende vraag die zich stelt: wat testen?

Het taalbeleid 'Academisch Nederlands' aan de UGent hanteert een oriënterende en niet dwingende toets. De universiteit kiest voor een aanpak waarbij de studenten zelf initiatief moeten nemen. De onafhankelijkheid van de student wordt hierbij gezien als een inherent onderdeel van een universitaire opleiding.

Ondanks het vrijblijvende karakter van de toets, wordt er wel voor gezorgd dat studenten zoveel mogelijk op de hoogte zijn van het bestaan van de toets. Hiervoor werden verschillende acties ondernomen.

- De studiebegeleiders van de deelnemende faculteiten fungeren als aanspreekpunt en maken de toets bekend via hun website

- Alle studenten van 1^{ste} bachelor kregen eveneens een email om hen tot deelname te motiveren.
- Vakdocenten en projectmedewerkers maakten de taaltoets bekend in de lessen.
- Er werd een intensieve poster- en stickeractie gevoerd.

In 2009-2010 nam ongeveer één derde van de aangeschreven studenten deel aan de screening.

Omdat aansluitend aan de toets verschillende remediëringstrajecten worden aangeboden, werd de toets aangeboden tijdens de eerste drie weken van het academiejaar. In de loop van de vierde week van het academiejaar werden de resultaten geanalyseerd, waarna studenten werden uitgenodigd op feedbacksessies.

De taaltoets heeft een signaalfunctie en dient niet als gatekeeper. Studenten nemen zelf het initiatief om hun vaardigheden te testen en zijn vrij om het advies te volgen. Het doel is immers niet de toets op zich, maar de bewustwording van het belang van taal in een academische context. Studenten met een minder goed resultaat worden uitgenodigd voor workshops die gebaseerd zijn op hun problemen en hun studiemateriaal. Ook studenten die goed scoorden of niet deelnamen aan de toets zijn echter welkom in de workshops.

WAAROM?

De waarom-vraag is van essentieel belang bij toetsontwikkeling. Een aantal vaak voorkomende redenen om te toetsen zijn

- Kennis en/of vaardigheden meten na een studietraject om na te gaan of een student de aangebrachte materie beheerst.
- Kennis en/of vaardigheden meten tijdens een studietraject met als doel na te gaan waar een student zich bevindt in een studietraject.
- Kennis en/of vaardigheden meten in een diagnostische test die bedoeld is om de sterktes en de werkpunten van een student te identificeren.
- Kennis en/of vaardigheden meten vóór een studietraject met als doel de begincompetenties in kaart te brengen en eventueel bepaalde studenten extra te ondersteunen.

(Cf. Alderson & Huhta 2005)

Toetsen in het kader van een taalbeleid academisch Nederlands behoort tot de laatste categorie. De toetsontwikkelaar wil dus nagaan of student x voldoende taalcompetenties bezit om aan de studierichting van zijn/haar keuze te beginnen. Om daarover uitsluitsel te kunnen geven dient men te weten over welke kennis en/of vaardigheden men minimaal moet beschikken om in een bepaalde opleiding te kunnen functioneren.

Wat?

Wanneer vast ligt waarom men een toets wil afnemen, dringt een tweede vraag zich op. *Wat* moet er getoetst worden? Het spreekt voor zich dat deze vragen elkaar in de realiteit niet strikt lineair opvolgen. Testconstructie is per definitie een cyclisch proces waarbij elke fase de vorige en de volgende kan beïnvloeden. Toch is het belangrijk om eerst de waarom-vraag beantwoord te hebben,

omdat die vraag, zoals hierboven geargumenteed werd een grote invloed heeft op het hele ontwikkelingsproces.

In de praktijk worden veel toetsen gemaakt met vooral de toetsinhoud voor ogen omdat men vaak een scherper beeld meent te hebben van wat men wil toetsen dan van waarom men dat wil. Toch houdt het een risico in om de wat-vraag louter intuïtief te beantwoorden (Douglas 2000, Alderson 2005). Als men immers wil weten of een student voldoende competenties bezit om te kunnen deelnemen aan de richting in het hoger onderwijs van zijn of haar keuze, volstaat het niet steeds om 'academisch Nederlands' te toetsen? Er bestaan immers oneindig veel variaties en specificaties binnen academisch taalgebruik. Zo beantwoordt juridisch Nederlands aan heel andere linguïstische wetmatigheden dan eerder literair of strikt wetenschappelijk Nederlands. Bovendien kennen bepaalde opleidingen een heel eigen begrippenkader waarvan bepaalde aspecten misschien al bij aanvang van de studie verondersteld gekend zijn.

Hoe specifiek moet de inhoud van de taaltoets dan zijn? In de literatuur over het testen van academische taalvaardigheid bestaan er zowel fervente voorstanders als hevige tegenstanders van verregaande specificiteit in taaltoetsen (Hyland 2001, Widdowson 2001, Douglas 2001). Toch bestaat er consensus over het feit dat aangepast inputmateriaal (zoals leesteksten met betrekking tot het studiegebied) positieve effecten heeft op de kwaliteit van een toets, of minstens op de perceptie ervan (Davies 2001). Omdat voorkennis in het eerste jaar hoger onderwijs niet echt een rol mag spelen bij het evalueren van taalvaardigheid, hoeft het inputmateriaal niet al te specifiek te zijn, maar een zekere alfa-, bèta of gamma-orëntatie in de onderwerpkeuze van de inputteksten kan zeker geen kwaad. Te veel specificiteit kan er immers voor zorgen dat voorkennis een te grote rol zal spelen bij het tekstbegrip en bovendien is het moeilijk om de exacte grens tussen twee taalspecifieke variëteiten af te bakenen. Waar eindigt Nederlands voor de biomedische wetenschappen en begint Nederlands voor diergeneeskunde? Beide vakgebieden delen bepaalde onderdelen, dus in welke mate moeten die richtingen ander inputmateriaal krijgen? Een pilootonderzoek aan de faculteit diergeneeskunde van de UGent wees in 2008 alleszins uit dat een taaltest die put uit het brede veld van de biomedische wetenschappen de voorkeur wegdraagt (Cf. Deygers & Neyts 2008). Net zomin als er een 'algemeen academisch Nederlands' bestaat, bestaat er immers ook geen Nederlands voor diergeneeskunde, een vakgebied dat veel specialisaties met eigen begrippenkaders kent.

Hoe specifiek moet bijvoorbeeld een leestekst dan zijn? Een testontwikkelaar mag niet al te veel onderwerpspecifieke voorkennis verwachten van instromende studenten, maar streeft toch een vorm van specificiteit na zodat de test als geloofwaardig gepercipieerd wordt door de studenten. Representatief toetsmateriaal wordt best niet door het ontwikkelingsteam alleen uitgewerkt. Een goed ontwikkelingsteam weet waarom er getest wordt en kent het doelpubliek. Op basis van die informatie kunnen er domeinspecialisten aangesproken worden. Domeinspecialisten zijn mensen die ervaring hebben binnen het domein waarvoor er getoetst wordt. Zo kan een collega die lesgeeft in de eerste Bachelor Office Management een domeinspecialist zijn bij het maken van een plaatsingstoets voor die richting. Voor een algemene plaatsingstoets die in de hele instelling gebruikt wordt, zijn misschien collega's nodig uit verschillende opleidingsonderdelen, terwijl voor andere toetsen ongetwijfeld ook professionals van buiten het onderwijsveld aangesproken zullen moeten worden. Ook studenten zelf zijn een nuttige bron van informatie. Een tweedejaarsstudent weet waarschijnlijk welke taalvereisten er zijn voor eerstejaarsstudenten en welke taaltaken potentieel probleem kunnen opleveren.

De taaltoets academisch Nederlands aan de UGent is ontwikkeld op maat van de deelnemende faculteit. Na overleg met domeinspecialisten (docenten) werd er een toets opgesteld op basis van cursussen en syllabi van de eerste bachelor, evenals proefexamens die werden afgenomen in vorige jaren. Het is echter belangrijk dat er niet te veel voorkennis wordt verwacht (bijvoorbeeld vakjargon dat in de loop van de studies behandeld zal worden). Daarom kregen de domeinspecialisten regelmatig de kans om feedback te geven op de toets tijdens het ontwikkelingsproces.

Een centrale vraag bij dit alles heeft betrekking op de constructvaliditeit van de toets. Meet de toets wat ze wil meten? Als een toets tot doel heeft na te gaan of een student talig zal kunnen functioneren in de komende opleiding, is het belangrijk na te gaan of dat ook effectief gemeten wordt. Het is dus van belang de domeinspecialisten (iedereen die nuttige input kan geven over de taalvereisten in de realiteit) te identificeren. Uit de gesprekken met de domeinspecialisten kunnen de essentiële begincompetenties van studenten gedestilleerd worden. Achterhalen of iemand zal kunnen functioneren in een bepaalde richting impliceert dat er op de voor een bepaalde studierichting essentiële zaken gefocust wordt terwijl andere taalelementen misschien onderbelicht blijven.

Via focusgroepsgesprekken, semigestructureerde interviews of via een andere onderzoeksmethode kan nagegaan worden wat de talige verwachtingen zijn in de opleiding die volgt. Het ene ontwikkelingsteam kan ervoor kiezen om deze verwachtingen zo rechtstreeks en letterlijk mogelijk om te zetten in toetstaken en dus de kaart van de authenticiteit te trekken. Een ander team kan ervoor kiezen om niet met geïntegreerde taaltaken te werken en kan proberen om met aangepast inputmateriaal te werken. Zoals eerder aangehaald zijn er sterke voor- en tegenstanders van authentieke, geïntegreerde taaltoetsen. Bij sommige ontwikkelaars zal een voorkeur voor één van beide visies de toetsinhoud en –vorm bepalen. Bij anderen zullen praktische overwegingen (zoals de noodzaak om online te testen) bepaalde mogelijkheden uitsluiten, zodat “hoe” en “wat” elkaar beïnvloeden.

Afhankelijk van de mate waarin een toets de realiteit wil simuleren worden het inputmateriaal en de vragen dus realistischer en specifiek. Grotere aandacht voor het toekomstige taalgebruik en vakgebonden taalvereisten betekent ook dat de domeinspecialisten een stem kunnen hebben in het opstellen van een beoordelingsmodel van bepaalde taken. Toetsen die juridisch Nederlands testen bijvoorbeeld kunnen niet opgesteld of beoordeeld worden zonder juridische feedback. Strikt talige regels zijn immers niet steeds van toepassing in wetteksten en juridische argumentatie. We hebben in dergelijke gevallen een domeinspecialist nodig die mee bepaalt welk antwoord adequaat is in een context van juridisch Nederlands. Het citaat van Jacoby en McNamara (1999) hieronder geeft aan dat bij het toetsen van specifiek academische vaardigheden een gedegen talenkennis niet steeds voldoende is. Het is aan de toetsontwikkelaar om via de domeinspecialisten te achterhalen welke intuïtieve criteria domeinspecialisten toepassen bij het beoordelen van bepaalde taken.

“Rather than relying on external rating scales or handbooks, professionals typically call upon a rich inventory of tacitly known criteria in order to determine whether and to what extent some particular performance is competent or falls short of the mark”

(Jacoby & McNamara 1999)

WAT?

Weten waarom er getoetst wordt is een essentiële voorwaarde om goed te kunnen toetsen. Weten wat er getoetst moet worden is dat ook. Zeker wanneer er een opleidings specifieke toets wordt ontworpen waarbij de noodzakelijke begincompetenties voor een opleiding worden bevestigd, worden er best domeinspecialisten bij het ontwikkelingsproces betrokken. Deze domeinspecialisten kunnen nuttige input geven over de vereiste competenties en over de kenmerken waaraan antwoorden moeten voldoen.

Vanzelfsprekend is het niet zo dat de wat-vraag pas beantwoord mag of kan worden nadat het antwoord op de waarom-vraag geëxpliciteerd is. Toetsontwikkeling is een cyclisch en dialectisch proces waarbij elke fase beïnvloed wordt door wat volgt of voorafgegaan is.

Hoe?

In een ideale wereld is een taaltoets zo *gevarieerd* mogelijk en bestaat ze uit verschillende soorten taken die aan alle noodzakelijke vaardigheden appelleren. Ze is zo *betrouwbaar* mogelijk doordat verschillende parallelle beoordelaars hun oordeel geven aan de hand van een nagenoeg waterdicht beoordelingsinstrument. De ideale toets is ook zo motiverend mogelijk doordat ze taken met een geïndividualiseerde inhoud op een betekenisvolle manier operationaliseert.

In de realiteit zorgen financiële en praktische beperkingen maar ook technologische vereisten er wel eens voor dat er keuzes moeten gemaakt worden. En hoewel keuzes een vast onderdeel uitmaken van het ontwikkelingsproces van een taaltoets, zorgen negatieve keuzes vaak voor een verlies aan kwaliteit. Praktische redenen, eerder dan een weloverwogen methodologische argumentatie, kunnen de keuze voor een digitale dan wel papieren test bovenmatig beïnvloeden. Technologische vereisten eerder dan toetstheoretische overwegingen kunnen ervoor zorgen dat een test elementgericht in plaats van geïntegreerd opgebouwd is.

Op zich zijn er heel wat valide argumenten te bedenken voor of tegen digitale toetsen en zijn er tal van wetenschappelijke publicaties die pleiten voor of tegen een elementgerichte aanpak (Van den Branden 2006, Widdowson 2001). Het is dus niet zo dat één bepaalde aanpak in alle situaties intrinsiek beter is dan een andere. Het punt is dat een aanpak een hele rits consequenties meebrengt. Wanneer ook die consequenties aangepast, verzwakt of uitgefilterd worden, blijft er geen concept meer over en wordt het bijzonder moeilijk om nog op een betrouwbare manier te toetsen.

De taaltoets academisch Nederlands van de Universiteit Gent wordt digitaal aangeboden via Curios, een online toetsplatform. De taaltoets bevestigt

woordenschat, grammatica, verwijs- en verbindingswoorden, spelling, lezen, syntactisch inzicht, tekststructuur en samenvatten.

Studenten kunnen drie weken lang overal en altijd de taaltoets maken. Op die manier is de toets maximaal beschikbaar, maar zijn de resultaten vanzelfsprekend minder betrouwbaar. Omdat het hoofddoel van het taalbeleid niet de toets is, maar eerder bewustmaking en remediëring, was een verminderde betrouwbaarheid een aanvaardbare toegift.

Een digitale toets heeft echter beperkingen. Niet alle taaktypes zijn mogelijk, zoals eerder aangegeven is er geen controle op de omstandigheden waarin de toets wordt afgelegd. Verder is het van essentieel belang dat de toets niet te lang duurt. Er wordt voor gezorgd dat de toets in 30 à 60 minuten gemaakt kan worden. Indien de toets langer is, wordt gevreesd dat studenten halverwege afhaken en dat de toetsresultaten onbetrouwbaar zijn.

Toch is het mogelijk om een kwaliteitsstandaard te handhaven met beperkte middelen. Onderstaand voorbeeld beschrijft ontwikkelingsprocessen van een toets algemene academische taalvaardigheid aan een universiteit. Het voorbeeld toont aan dat het mogelijk is een taaltoets uit te werken die – ondanks beperkingen – met een potentieel aanvaardbare graad van betrouwbaarheid en validiteit.

FASE 1: PLANNING

Waarom?	Academische taalvaardigheid instromende studenten nagaan.	Motivatie De universiteit wil een beeld krijgen van de evolutie van academische taalvaardigheid en ze wil potentiële probleemstudenten een hulpmiddel aanreiken.
Wat?	Algemene academische taalvaardigheid.	Opleidings specifieke taalvaardigheid komt later. De universiteit wil nagaan of studenten een basiswoordenschat bezitten en of ze algemene academische taalvaardigheden bezitten.
Hoe?	Er wordt gekozen voor een elementgerichte digitale taaltest met meerkeuzevragen en invuloefeningen.	Er zijn geen personeelsmiddelen om authentieke taken zoals samenvattingen en essays te beoordelen.

FASE 2: ONTWERP

Eerste versie	Het ontwikkelingsteam schrijft een eerste versie van de toets.	Op basis van input van domeinspecialisten en na een literatuuronderzoek wordt een eerste versie geschreven.
---------------	--	---

Feedback	Resonansgroep geeft feedback	De resonansgroep (bestaande uit leden van het ontwikkelingsteam en domeinspecialisten) geeft feedback.
Tweede versie	Het ontwikkelingsteam schrijft een tweede versie van de toets.	Op basis van de feedback wordt de eerste versie herschreven.

FASE 3: ONTWERP

Pilot	De voorlopige toetsversie wordt voorgelegd aan 100 scholieren uit het zesde jaar secundair.	Door de toets te proefdraaien voor een honderdtal personen die de uiteindelijke testpopulatie zo goed mogelijk benaderen kunnen toekomstige problemen vermeden worden.
Analyse & revisie	Op basis van de analyse worden dubbelzinnige vragen aangepast en worden te moeilijke en te makkelijke items weggelaten.	Relatief eenvoudige berekeningen (gemiddelde, standaardafwijking) identificeren te moeilijke en te makkelijke vragen. Interviews of enquêtes kunnen gebruikt worden om te peilen naar de mening over de toets.

FASE 4: TOETS

De toets wordt afgenomen. De resultaten worden bijgehouden en geanalyseerd om de test nog verder te optimaliseren.

Bovenstaand voorbeeld schetst een ontwikkelingsproces van een elementgerichte digitale toets. Vanzelfsprekend is het ook mogelijk om meer authentieke, productieve taken te gebruiken in een taaltoets, zolang de consequenties van die beslissing op de taakopbouw, het beoordelingsmodel en het gehele ontwikkelingsproces in acht genomen worden.

Het zal duidelijk zijn dat de hoe-vraag een grote impact heeft op de uiteindelijke vorm van een taaltoets. De waarom- en de wat-vragen leggen de fundamenten waarop het hoe steunt, maar geen enkele vraag kan bevredigend beantwoord worden zonder aandacht te hebben voor de andere vragen.

Hoe?

Hoe een toets opgesteld wordt, welke taken er gebruikt worden, is misschien de meest concrete van de drie tot nog toe gestelde vragen. Toch is het zo dat het hoe niet adequaat beantwoord kan worden wanneer het waarom en het wat niet gestroomlijnd zijn.

Hoe is immers de uitwerking van het theoretische concept van de toetsontwikkelaars in combinatie met het doel en de inhoud van de toets.

En Daarna?

Het ontwikkelingsproces van een taaltoets dat in deze bijdrage kort beschreven is, heeft tot doel een aantal essentiële kwaliteitsparameters van taaltoetsen te garanderen. Zonder aandacht voor het waarom en het wat van een toets, verwatert misschien de aandacht voor de validiteit. Aandacht voor het hoe kan de betrouwbaarheid versterken.

Er zijn verschillende soorten validiteit, maar we hebben het hier vooral over de construct validity: meet de toets wat hij zegt te meten. Meet deze luistertaak effectief luistervaardigheid of eerder kennis van de wereld? Is die spreektaak niet eerder een leesopdracht? Via kritische reflectie tijdens het ontwikkelingsproces en via bevraging van domeinspecialisten kunnen validiteitsproblemen vaak voor een groot stuk worden opgevangen.

Een tweede potentiële valkuil heeft betrekking op de accuraatheid waarmee een toets meet. Betrouwbaarheid gaat over de accuraatheid, objectiviteit en standvastigheid van een beoordelingsmodel. Krijgt de kandidaat een score die hij of zij verdient?

Betrouwbaarheid verzekeren is een grotendeels technische, kwantitatieve kwestie. Kwalitatief onderzoek helpt bij het verzekeren van validiteit, maar zonder statistische data kan er weinig zinnigs gezegd worden over de betrouwbaarheid van een test. Via een proefsessie of pilot met een honderdtal proefpersonen die sterk lijken op de doelgroep van de toets kunnen de verschillende testitems geanalyseerd worden en kan betrouwbaarheid verzekerd worden.

Tijdens de ontwikkelfase van de taaltoets academisch Nederlands aan de UGent werd er een pilot afgenomen bij 234 laatstejaarsstudenten in ASO-richtingen van 3 scholen.

De leerlingen kregen de opdracht de voorlopige versie van de toets te maken. In elke school werd een lesuur (50 minuten) ingeschakeld om de pilot af te nemen via Curios, het online toetsplatform van de UGent. Deze tijdsbeperking was de voornaamste reden om de toets in te korten. Na een korte introductie bleven er nog een veertigtal minuten over om de test te maken en een motivatietekst te schrijven. De afname is zonder noemenswaardige problemen verlopen en de meeste leerlingen schreven een representatieve tekst.

Op basis van deze pilot werden bepaalde items weggelaten omdat ze onbetrouwbaar bleken. Er werd tevens beslist om meer actieve en productieve taken op te nemen in de digitale taaltoets voor 1^{ste} bachelor.

Afzien van een onderzoek naar betrouwbaarheid en validiteit, betekent dat er geen echte conclusies kunnen getrokken worden op basis van een toets. Men is immers niet zeker of de toets op een accurate manier meet wat hij zegt te meten. Geloof in de resultaten van een toets waarvan betrouwbaarheid of validiteit niet nagegaan zijn, is dus zeer broos en de zekerheid die deze resultaten bieden is slechts een valse zekerheid.

De taaltoets academisch Nederlands aan de UGent kan altijd en overal gemaakt

worden, wat gevolgen heeft voor de betrouwbaarheid. Daarom is er voor gekozen om geen numerieke score te geven aan de studenten.

Na post-analyse van de toets werden de deelnemende studenten per email uitgenodigd voor persoonlijke feedback. Ook studenten die een hoge score behalen, werden uitgenodigd omdat een hoge score niet noodzakelijk impliceert dat studenten geen problemen ondervinden met andere aspecten van taalvaardigheid die via een digitale toets niet gemeten kunnen worden.

Tijdens de feedback werden de beperkingen van de toets aan de studenten duidelijk gemaakt. Het is belangrijk dat de taaltoets niet gezien wordt als een instrument dat een grens bepaalt. Het is echter wel een instrument om studenten bewust te maken van het belang van taal in hun opleiding. Op die manier willen we hun aandacht vestigen op eventuele talige problemen die ze kunnen ondervinden in een academische context. Studenten die ook niet hebben deelgenomen aan de toets worden warm uitgenodigd om de test te bekijken en deel te nemen aan eventuele remediëringssessies tijdens het academiejaar.

Jammer genoeg beschikken de ontwikkelaars van taaltoetsen in het kader van een taalbeleid niet over grote budgetten en wordt er zelden beroep gedaan op psychometrici of statistici om betrouwbaarheid te garanderen. Het is daarom sterk aan te raden om voorzichtig om te springen met resultaten waarvan niet zeker is of ze accuraat zijn en om geen rechtstreekse en onomkeerbare conclusies te verbinden aan meetinstrumenten waarvan de correctheid niet buiten kijf staat. Anders gezegd: zonder het waarom, wat en hoe van een taaltoets ten volle te bevragen, is het gevaarlijk er een beleid op te bouwen.

Bibliografie

- Alderson, Charles J. & Ari Huhta (2006). "The development of a suite of computer-based diagnostic tests based on the Common European Framework". *Language Testing*. 22(3)
- Alderson, Charles, Caroline Clapham en Diane Wall (2005). *Language Test Construction and Evaluation*. Cambridge University Press.
- Bachman, Lyle F (2000). "Modern Language Testing at the Turn of the Century: assuring that what we count counts". *Language Testing*. 17(1)
- Davies, Alan (2001). "The Logic of testing languages for Specific Purposes". *Language Testing*. 18(2)
- Douglas, Dan (2000). *Assessing Languages for Specific Purposes*. Cambridge University Press
- Douglas, Dan (2001). "Language for Specific Purposes assessment criteria: where do they come from?". *Language Testing*. 18(2)
- Green, Anthony B. and Cyril J. Weir (2004). "Can placement tests inform instructional decisions?". In *Language Testing*. 21(4)
- Hyland, Ken (2001). "Specificity Revisited: how far should we go now?". *English for Specific Purposes*. 21(4)
- Krekeler, Christian (2006). "Languages for special academic purposes (LSAP) testing: the effect of background knowledge revised". In *Language Testing*. 23(1)
- McNamara, Tim (2000). *Language testing*. Oxford University Press
- Van den Branden, K., M. Bygate en J Norris (2009). *Task-based language teaching: a reader*. John Benjamins Publishing.
- Van den Branden, K. (Ed.) (2006). *Task-based language teaching: from theory to practice*. Cambridge University Press.
- Widdowson, Henry G (2001). "Communicative Language Testing: The Art of the Possible". *Studies in Language Testing* 11. Cambridge University Press.
- Wu, Weipeng en Charles W. Stansfield (2001). "Towards Authenticity of Task in Test Development". *Language Testing*. 18(2)