

# Taaltoetsen getoetst

## Een empirisch onderzoek van toetsen voor Engels en Duits in het secundair onderwijs in Vlaanderen

EEF DECLOEDT EN LIES SERCU

De auteurs rapporteren over een onderzoek naar het toetsgedrag van leerkrachten vreemde talen in Vlaanderen, uitgevoerd op basis van een corpus toetsen. De resultaten laten zien dat de klemtoon in toetsen weliswaar op transfervragen ligt en niet op zuivere kennisvragen, maar dat reproductievragen en niet-gecontextualiseerde vragen het authentieke karakter van de toetsen onderuithalen. Toetsen voor de opeenvolgende jaren van het secundair onderwijs verschillen ook niet significant van elkaar qua kennis-, transfer- en communicatiegehalte.

In de literatuur over testen en evalueren valt de laatste jaren een groeiend bewustzijn te noteren over de impact die testresultaten kunnen hebben op wie de test aflegt (McNamara, 2000; McNamara, 2004). Het is dan ook uitermate belangrijk dat diegenen die toetsen ontwikkelen en gebruiken, beschikken over de nodige professionele vaardigheden en inzichten. Slecht geredigeerde of gebruikte toetsen zullen immers een onjuist beeld geven van wat iemand kan en kent.

In dit artikel rapporteren we over een onderzoek naar het testgedrag van leerkrachten moderne vreemde talen. Op basis van

de analyse van een corpus hebben we willen nagaan of zich in het evaluatiegedrag van Vlaamse leerkrachten Duits en Engels bepaalde algemene tendensen laten onderscheiden en of die een toetsaanpak weerspiegelen die tot een correcte weergave van het taalvaardigheidsniveau in de vreemde taal leidt. Concreet stelden we ons onder andere de volgende vragen: Laat zich uit de toetsen een algemeen basisconcept van taalvaardigheid en/of kennis bij de leerkrachten afleiden? Verschillen toetsen voor beginners en gevorderde leerders substantieel? Verschillen toetsen voor leerlingen uit het TSO (havo) en het ASO (vwo)?

In wat volgt schetsen we kort het theoretische kader van het onderzoek. Daarna wordt de onderzoeksmethode gepresenteerd en met enkele voorbeelden geïllustreerd. Vervolgens geven we aan wat de belangrijkste resultaten zijn en gaan we na wat de implicaties zijn voor de evaluatiepraktijk in het vreemdetalenonderwijs.

### *Communicatieve taalvaardigheid als uitgangspunt voor taaltoetsen*

In het vreemdetalenonderwijs in Vlaanderen, zoals in de rest van Europa en belangrijke

andere delen van de wereld, staat de verwerking van communicatieve vaardigheid centraal. In Vlaanderen blijkt dit uit de geldende leerplannen voor Frans, Engels, Duits en Spaans voor het secundair onderwijs, maar evengoed uit die voor het Russisch of het Chinees in de volwasseneneducatie. Zij weerspiegelen het construct 'communicatieve taalvaardigheid' zoals dat werd uitgewerkt in het Common European Framework of Reference for Modern Languages (2001).

Bachman en Palmer (1996) onderscheiden in hun definitie van communicatieve vaardigheid een kenniscomponent en een strategische component. Het is niet voldoende te beschikken over grammaticale, tekstuele, functionele of sociolinguïstische kennis; taalleerders moeten die kennis ook kunnen gebruiken in authentieke communicatieve situaties.

Willen evaluatie-instrumenten, zoals schriftelijke summatieve toetsen, valide zijn, dan moeten zij dit uitgangspunt weerspiegelen (Sercu, Vyncke & Peters, 2003). Vragen die kennis peilen van bepaalde aspecten van de woordenschat of de grammatica worden aangevuld met functionele communicatieve opgaven die nagaan of leerlingen die kennis ook kunnen gebruiken in geziene contexten en transfereren naar nieuwe contexten. Zowel kennis (van de vorm en de betekenis) als vaardigheid (met de vorm en de betekenis) moeten getoetst worden in een authentieke talige context, want enkel zo kan de toets een idee geven van de mate waarin de taalleerder in de vreemde taal zal kunnen functioneren in de 'echte' wereld (Purpura, 2005). Over de jaren heen verhogen ook de eisen die aan taalvaardigheid worden gesteld. Eenvoudige vragen op het woord- of zinsniveau worden aangevuld met vragen op het tekstniveau; productievragen en communicatieve taken vervangen reproductievragen of aanvulvragen (Alderson, 2000; Read, 2000).

Het construct 'communicatieve taal-

vaardigheid' zoals het hierboven kort werd omschreven en de implicaties voor taaltoetsing die eruit volgen, vormden het raamwerk voor het empirische onderzoek waarover we hier rapporteren en weerspiegelen zich in ons analyse-instrument.

### *Schriftelijke toetsen geëvalueerd: een empirisch onderzoek*

#### ONDERZOEKSGROEP EN CORPUS

Eind januari 2004 nodigden we de directies van alle Vlaamse secundaire scholen per e-mail uit om mee te werken aan ons onderzoek. In totaal gingen 64 leerkrachten Engels en 26 leerkrachten Duits op onze uitnodiging in, een relatief klein aantal, dat wellicht ten dele verklaard kan worden vanuit het specifieke voorwerp van ons onderzoek: scholen noch docenten zijn bereid hun toetsmateriaal ter beschikking te stellen uit angst dat dit in verkeerde handen terecht zou kunnen komen.

Van de 500 toetsen die we elektronisch of per post ontvingen, werden er 226 geselecteerd voor verdere analyse: 156 toetsen voor Engels en 70 toetsen voor Duits. De geanalyseerde toetsen lagen verspreid over onderwijstypes (ASO, TSO en BSO<sup>1</sup>) en jaren (2de tot 7de jaar).

#### *Data-analyse*

In een Accesdatabase kreeg elke toets een aantal labels mee: een label voor taal (Engels of Duits), onderwijsvorm (ASO, TSO, BSO), jaar (2 - 7), en aantal uren onderwijs in die taal per week (1-3)<sup>2</sup>. Deze informatie liet ons toe bijvoorbeeld toetsen voor Duits met toetsen voor Engels te vergelijken of toetsen voor ASO met toetsen voor TSO.

Elke toetsvraag werd geïdentificeerd als een leesvaardigheids-, schrijfvaardigheids- of spreekvaardigheidsvraag. Als spreekvaardig-

heid werden die toetsvragen gecodeerd die de leerlingen vroegen een dialoog te vervolledigen of bijvoorbeeld een tekst te schrijven die in de echte wereld uitgesproken zou worden. ‘Woordenschat’ en ‘grammatica’ worden niet apart gecodeerd, maar geïntegreerd in de vaardigheden, geheel in overeenstemming met ons theoretische kader.

Per vaardigheid werden een aantal inhoudelijke aspecten in kaart gebracht. Er werd bekeken of een toetsvraag betrekking had op het woord-, zins- of tekstniveau. Het verschil tussen zinnen en tekst (aangezien tekst ook uit zinnen bestaat) had te maken met de samenhang van de zinnen in kwestie: waren het losse zinnen, zonder inhoudelijke samenhang of was het een samenhangende tekst waarmee gewerkt werd? We geven een voorbeeld voor een vraag op woord-, zins- en tekstniveau.

Woordniveau: Was gebraucht man zum... (z.B. Schreiben in ein Heft (drei Sachen)? Setze die richtigen Wörter ein und achte auf die Akkusativformen. Gebrauch in deiner Antwort eine Form von, ein’.

Zinsniveau: Bilde Vergleichssätze. Zu jedem Adjektiv soll es einen Satz mit Komparativ und einen mit Superlativ geben.

Tekstniveau (opgave bij nieuwe leestekst): Welche dieser Aussagen könnte zu wem passen?

Ten tweede werd nagegaan of het verwachte taalgedrag van de leerlingen te situeren viel op het kennis- (K), transfer- (T) of communicatieniveau (C) volgens de taxonomie van doelstellingen voor het talenonderwijs, zoals uitgewerkt door Valette en Disick (1972). Met deze taxonomie is het mogelijk vraagtypes en toetsvragen binnen een bepaalde moeilijkheidsgraad te categoriseren. Grof genomen stonden kennisvragen in onze codering voor letterlijke vragen, in een vooraf geziene context. ‘Transfer’ definieerden we als het

toepassen van kennis in een nieuwe situatie of context. Bij een communicatieve oefening, tot slot, produceert/begrijpt de leerling zelf nauwkeurig en vlot de inhoud en/of vorm van een nieuwe, authentieke boodschap. We geven voorbeelden van elk niveau.

Kennisniveau: Translate the following words: de leerling geeft de vertaling van geziene woorden.

Transfervniveau: Wortfelder. Sieh dir diese Schulklasse an. Beschreibe, was du (nicht) siehst. Schreibe wenigstens 10 gute Sätze mit möglichst viel Wortschatz zum Thema: de leerlingen schrijven nieuwe zinnen op basis van de informatie op de foto.

Communicatieniveau: Read the text and answer the following questions: de leerlingen lezen een nieuwe tekst en proberen die te begrijpen.

Bovendien gingen we na of de toetsvraag een relatief directe relatie kon hebben met taalgebruik in een authentieke situatie. Daarvoor maakten we het onderscheid tussen wel en niet-gecontextualiseerde vragen. Een gecontextualiseerde vraag moest voldoen aan vier criteria: ze heeft betrekking op een functionele, authentieke context (1), de leerling decodeert die authentieke context of ongeziene tekst (2), er is sprake van (communicatieve) toepassing van kennis (3) en de leerling kan zich in de nieuwe situatie inleven (4). Een niet-gecontextualiseerde vraag wordt dan een vraag om de vraag, die (formele) kennis of transfer test.

Gecontextualiseerde vraag: Annika will Karl besuchen. Sie reist von Köln nach Leipzig. Beschreibe ihren Reisweg mit dem Zug. (vertrekkur, aankomstuur, duur reis, spoor, overstappen: waar? wanneer? hoeveel tijd?) Benutze dazu die Reiseinformationen auf der nächsten Seite. Schreibe gute deutsche Sätze. Benutze möglichst viel Wortschatz

und Redemittel zum Thema Reisen. De leerlingen leven zich dus in een nieuwe, authentieke context in en passen hun kennis communicatief toe.

Niet-gecontextualiseerde vraag: *What do we call...? E.g. a glass which has been coloured and cut into various shapes to form pictures (usually found in churches).* Dit is een formele oefening: de leerlingen reproduceren woordenschat.

### Resultaten

Wij geven hieronder een overzicht van de belangrijkste resultaten van de corpusanalyse. Daarbij schetsen we eerst enkele tendensen die gelden voor het volledige corpus om nadien te gaan kijken naar een aantal opmerkelijke verschillen tussen groepen toetsen (ASO versus TSO/BSO, Duits versus Engels, beginjaren versus gevorderden).

#### BELANGRIJKE TENDENSEN GELDEND VOOR HET HELE CORPUS

Ten eerste laten onze data zien dat, zoals te verwachten, schrijfvaardigheid (ongeveer 82% van alle vragen)<sup>3</sup> verreweg het vaakst aan bod komt in de schriftelijke toetsen die we analyseerden. Daarbij gaat het echter meestal om schrijven als ondersteunende en functionele vaardigheid en dus niet om ‘echte’ communicatieve schrijfvaardigheid. Voor lezen is dit ook zo, maar minder, aangezien in heel wat toetsen ook een apart onderdeel gericht op leesvaardigheid is opgenomen. 12% van de vragen werd gecodeerd als ‘spreekvaardigheidsvraag’. Dit is een relatief hoog aandeel, dat mede verklaard kan worden vanuit het feit dat we woordenschat- en grammaticakennisvragen in ‘gesproken’ teksten ook als spreekvaardigheid gecodeerd hebben.

Wat nu de vraag betreft of de toetsen ‘echte’ vaardigheid in semi-authentieke situaties kunnen testen, daar konden we vaststellen

Vaardigheid	Niveau		
	Kennis	Transfer	Communicatie
Lezen	11	73	15
Spreken	5	77	18
Schrijven	29	66	5

Tabel 1: Taxonomisch niveau van de vragen (aantallen in percentages, afgerond)

Vaardigheid	context JA	context NEEN
Lezen	42	58
Spreken	37	63
Schrijven	23	77

Tabel 2: Contextualisering (aantallen in percentages, afgerond)

Vaardigheid	Woord	Zin	Tekst	Totaal
Lezen	13	44	44	100
Spreken	9	70	21	100
Schrijven	54	33	13	100

Tabel 3: Aandeel vragen op woord-, zins- en tekstniveau per vaardigheid

dat de meeste vragen voor lezen, spreken en schrijven zich op het transfervniveau bevinden (zie Tabel 1) en dat de leerlingen zich veelal niet in een authentieke situatie moeten inleven (zie Tabel 2).

Uit Tabel 3 blijkt de verdeling van de vragen over het woord-, zins- en tekstniveau. Doordat de leerlingen bij het evalueren van communicatieve leesvaardigheid een ongeziene tekst krijgen voorgeschoteld, is het geen verrassende vaststelling dat het belang van tekstvragen het grootst is bij de vaardigheid lezen (44%), vergeleken met de andere vaardigheden (21% en 13%). De woordvragen

ten slotte komen het vaakst voor bij schrijfvaardigheid en zijn er groter in aantal dan vragen op zinsniveau. Dit heeft te maken met het veel meer functionele dan communicatieve karakter van schrijfvaardigheidsvragen.

### *Verschillen tussen toetsen*

In dit onderdeel gaan we kort in op een aantal opmerkelijke verschillen tussen toetsen voor ASO- versus TSO-leerlingen, toetsen voor beginners versus gevorderden en toetsen voor Engels versus Duits. Hierbij beperken wij ons tot onze belangrijkste conclusies; een weergave van de onderzoeksresultaten valt buiten het bestek van dit artikel. De resultaten kunnen opgevraagd worden bij de auteurs.

Ten eerste valt het op dat leerlingen in het TSO significant<sup>4</sup> meer transfervragen voorgeschoteld krijgen dan leerlingen in het ASO. Van de leerlingen in het ASO wordt vaker verwacht dat zij gekende leerstof reproduceren (kennisniveau). Dit suggereert dat van TSO-leerlingen meer vaardigheid dan kennis wordt verwacht en van ASO-leerlingen meer kennis dan vaardigheid. Het verschil in kennis/transfer tussen het ASO en TSO is duidelijker voor Duits dan voor Engels. In toetsen voor Duits reproduceren ASO-leerlingen duidelijk meer kennis dan TSO-leerlingen.

Tussen de opeenvolgende jaren zagen we een stuk minder statistisch significante verschillen dan verwacht. We vonden wel een stijging in de graad van contextualisering van de vragen met de jaren, maar die verschillen waren niet significant. Voor het 6de jaar Engels meten we zelfs een terugval in contextualisering. Uit onze analyse blijkt verder dat leerlingen in het 4de jaar, vergeleken met die uit het 2de jaar, meer met tekstvragen worden geconfronteerd en minder met vragen gericht op het zinsniveau. Deze algemene tendens wordt echter niet verder doorgetrokken in het 5de en 6de jaar. Wat het taxonomische

niveau van de vragen betreft, hebben we geen statistisch significante verschillen kunnen waarnemen tussen de opeenvolgende jaren. We hadden een significante stijging van het aandeel communicatie en een daling van het aandeel kennis verwacht voor zowel Engels als Duits over de jaren heen, maar vonden het omgekeerde: het aantal kennisvragen neemt toe van het 2de tot het 5de jaar. Ten opzichte van het 5de jaar daalt in het 6de jaar Duits het aantal vragen gericht op kennis en er worden iets meer communicatieve en transfervaardigheden van de leerlingen verwacht. Voor Engels is dit niet het geval.

### *Conclusies*

Op basis van de resultaten verkregen uit de analyse kunnen we concluderen dat de huidige evaluatiepraktijk op een aantal punten voor verbetering vatbaar is. Over het algemeen scoren de toetsen niet goed op het vlak van authenticiteit. Het feit dat het testen van leesvaardigheid, in tegenstelling tot pakweg 10-15 jaar geleden, deel is gaan uitmaken van toetsen, maakt dat toetsen met betrekking tot authenticiteit wel beter zijn gaan scoren. Vaak wordt namelijk aan de leerlingen een nieuwe, authentieke tekst aangeboden die zij zelfstandig moeten verwerken. Wat schrijfvaardigheid betreft kan de authenticiteit echter nog in belangrijke mate verbeterd worden. Zich schriftelijk adequaat kunnen uitdrukken in een nieuwe authentieke situatie wordt veel minder geëvalueerd in schriftelijke toetsen dan authentieke leesvaardigheid.

Een verwonderlijke vaststelling voor ons is het ontbreken van significante verschillen tussen de opeenvolgende jaren voor wat het kennis-, transfer- en communicatiegehalte van de toetsen betreft. Vanuit het cyclische proces van taalontwikkeling, zoals vooropgesteld in het Common European Framework of Reference for Languages (2001), gingen

we uit van de veronderstelling dat leerlingen naarmate ze in een hoger jaar les volgen de vreemde taal beter zouden beheersen en dat die betere beheersing ook gereflecteerd zou worden in het type vragen dat ze op toetsen voorgeschoteld krijgen. Dit is niet het geval, wel integendeel. Als we dan toch een heel minimale (zij het niet statistisch significante) trend opmerken, is het er één in tegengestelde richting met meer kennisvragen per opeenvolgend leerjaar. Ook het uitblijven van een hogere graad van contextualisering voor leerlingen die in een hoger jaar les volgen, komt niet overeen met onze verwachting dat meer communicatieve vaardigheden in de hogere jaren gepaard zouden gaan met een groter gewicht voor authentieke gebruiksccontexten. Daarnaast verwachtten we een groter wordend belang van de tekstvragen. Dit is zo voor het 4de ten opzichte van het 2de jaar, maar wordt niet verder doorgetrokken naar het 5de en 6de jaar. Deze drie vaststellingen tonen ons dat er op dat vlak geen evolutie ligt in de toetsen: Er wordt niet meer echte communicatie verwacht in de hogere jaren.

Welke aanbevelingen voor de evaluatiepraktijk laten zich uit deze bevindingen afleiden? Het lijkt ons dat de evaluatiepraktijk in vergelijking met 15 jaar geleden in belangrijke mate communicatiever is geworden, maar dat er nog mogelijkheden voor verbetering zijn. Wanneer leerkrachten meer in termen van een cyclisch taalontwikkelingsproces zouden gaan denken en in hun toetsen een progressie zouden realiseren op het vlak van moeilijkheidsgraad (kennis/transfer/communicatie, woord/zin/tekst), en anderzijds kennis- en transfervragen nog meer zouden durven vervangen door communicatieve toetstaken, dan zou het construct 'communicatieve vaardigheid' ongetwijfeld beter weerspiegeld worden in de gebruikte toetsen.

Een grondige hertekening van de wijze waarop toetsen en permanente evaluatie

samen het effectieve taalvaardigheidsniveau van elke leerling adequaat kunnen weerspiegelen, zal leiden tot een evaluatiepraktijk die toelaat het leerproces optimaal te sturen en die bovendien geen onterechte gevolgen heeft voor degene die geëvalueerd wordt.

#### NOTEN

1. ASO staat voor Algemeen Secundair Onderwijs en is vergelijkbaar met het Nederlandse VWO. TSO staat voor Technisch Secundair Onderwijs en is vergelijkbaar met het Nederlandse HAVO of VMBO-t. BSO, tot slot, is vergelijkbaar met de overige afdelingen van het VMBO.
2. Deze informatie kon in de verdere analyse niet worden meegenomen omdat niet alle leerkrachten deze informatie vermeldden.
3. De percentages zijn gemiddelden voor alle toetsen en alle leerkrachten.
4. De gerapporteerde verschillen zijn alle gebaseerd op t-test resultaten. Wanneer een resultaat statistisch significant is bij  $p < .05$ , dan betekent dit dat er slechts 5% kans is dat het vastgestelde verschil te wijten is aan de toevallige samenstelling van de proefgroep en dat er 95% kans is dat het resultaat een werkelijk bestaand verschil weerspiegelt. De t-testresultaten werden niet opgenomen in de tekst, maar kunnen opgevraagd worden.

#### LITERATUUR

- Alderson, J. C. (2000). *Assessing Reading. The Cambridge language assessment series*. Cambridge: Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Canale, M., & Swain, M. (1980) *Theoretical bases of communicative approaches to*

second language teaching and testing. *Applied Linguistics*, 1, 1-47.

Council of Europe (2001). *Common European Framework of Reference for Languages. Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Kasper, G., & Kellerman, E. (Eds.) (1997). *Communication Strategies. Psycholinguistic and Sociolinguistic Perspectives*. London and New York: Longman.

Kramsch, C. (1998). *Language and Culture*. Oxford: OUP.

McNamara, T. F. (2000). *Language Testing*. Oxford: Oxford University Press.

McNamara, T. F. (2004) *Language Testing*. In A. Davies and C. Elder (Eds.), *Handbook of Applied Linguistics* (pp. 763-783). Oxford: Basil Blackwell.

Purpura, J. E. (2005). *Assessing Grammar*. The Cambridge language assessment series. Cambridge: Cambridge University Press.

Read, J. (2000). *Assessing Vocabulary*. The Cambridge language assessment series. Cambridge: Cambridge University Press.

Sercu L., Vyncke C., & Peters E. (2003). *Testen en evalueren in het vreemdetalenonderwijs. Cahiers voor didactiek*. Mechelen: Wolters Plantyn.

Sercu, L. (2002). Autonomous learning and the acquisition of intercultural communicative competence. Implications for course development. *Language culture and curriculum* 15(1), 61-74.

Valette, R. M. & Disick, R. S. (1972). *Modern Language. Performance Objectives and Individualization: A Handbook*. Harcourt: New York.

Weigle, S. (2002). *Assessing Writing*. Cambridge: CUP.