

Kritiek, Commentaar, Beschouwing

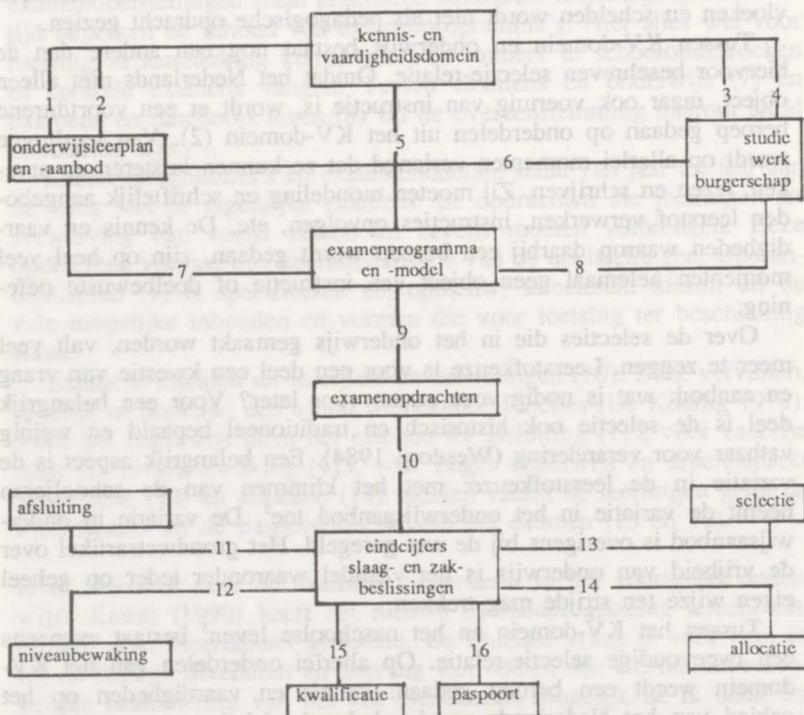
Kees de Glopper

Over eindexamens, validiteit en betrouwbaarheid

1. Over examenfuncties

Het vorige nummer van Spiegel bevatte een interessante bijdrage van Kreeft (1990) die uitnodigt tot nadere beschouwing over eindexamens, validiteit en betrouwbaarheid. Welke plaats nemen de eindexamens Nederlands in het voortgezet onderwijs in? Wat is hun positie ten opzichte van het voorafgaande onderwijs en het leven na het diploma? De onderstaande figuur probeert daarvan een beeld te geven.

Figuur 1. Examens en hun context



In het hart van de figuur staat het examenprogramma met als eventuele verdere specificatie het examenmodel (vgl. Kreeft 1990). De kop bevat het kennis- en vaardigheidsdomein. Onze verkenning begint daar.

Het kennis- en vaardigheidsdomein (KV-domein) is de verzameling van mogelijke leerinhouden, als het ware de catalogus van de mogelijke leerstof van een schoolvak¹. In ons geval gaat het om het vak Nederlands en daarmee is het domein beperkt tot mogelijke kennis en vaardigheden op het gebied van taal en taalgebruik (taalvaardigheid, taalbeschouwing en literatuur). Ik noem het een verzameling van *mogelijke* inhouden om aan te geven dat het een abstractie is. Er is geen schoolboek dat omvangrijk genoeg is om alle leerstof uit het domein te bevatten. Uit het KV-domein moet geselecteerd worden, in het onderwijs, in de examens en in het naschoolse leven.

In het onderwijs wordt een selectie uit de mogelijke leerstof aangeboden (1). Er moet gekozen worden omdat niet alles in de beschikbare tijd en met de beschikbare middelen onderwijsbaar is. Niet alles hoeft ook aan iedereen onderwezen te worden: veel leerlingen brengen van huis uit een aanzienlijke linguïstische competentie in het Nederlands mee. Ook is niet alle leerstof op ieder moment relevant. Wat in eerdere jaren onderwezen is, hoeft (gelukkig) niet allemaal herhaald te worden. Verder worden onderwijsbare kennis en vaardigheden niet altijd passend geacht. Om maar een voorbeeld te noemen: het leren vloeken en schelden wordt niet als pedagogische opdracht gezien.

Tussen KV-domein en onderwijs bestaat nog een andere dan de hiervoor beschreven selectie-relatie. Omdat het Nederlands niet alleen object, maar ook voertuig van instructie is, wordt er een voortdurend beroep gedaan op onderdelen uit het KV-domein (2). Van leerlingen wordt op allerlei momenten verlangd dat ze kunnen luisteren en spreken, lezen en schrijven. Zij moeten mondeling en schriftelijk aangeboden leerstof verwerken, instructies opvolgen, etc. De kennis en vaardigheden waarop daarbij een beroep wordt gedaan, zijn op heel veel momenten helemaal geen object van instructie of doelbewuste oefening.

Over de selecties die in het onderwijs gemaakt worden, valt veel meer te zeggen. Leerstofkeuze is voor een deel een kwestie van vraag en aanbod: wat is nodig voor nu en voor later? Voor een belangrijk deel is de selectie ook historisch en traditioneel bepaald en weinig vatbaar voor verandering (Wesdorp 1984). Een belangrijk aspect is de variatie in de leerstofkeuze: met het klimmen van de schooljaren neemt de variatie in het onderwijsaanbod toe². De variatie in onderwijsaanbod is overigens bij de wet geregeld. Het grondwetsartikel over de vrijheid van onderwijs is het vaandel waaronder ieder op geheel eigen wijze ten strijde mag trekken.

Tussen het KV-domein en het naschoolse leven³ bestaat eveneens een tweevoudige selectie-relatie. Op allerlei onderdelen van het KV-domein wordt een beroep gedaan: kennis en vaardigheden op het gebied van het Nederlands worden bekend of beheerst verondersteld (3). Bepaalde onderdelen worden gewaardeerd, andere verguisd.

In het naschoolse leven wordt meestal geen regulier onderwijs Nederlands meer genoten; toch vindt er wel op niet te verwaarlozen schaal training en leren plaats. Bij cursussen vergadertechnieken of

schriftelijk rapporteren worden bepaalde onderdelen van het KV-domein als object van instructie geselecteerd (4).

De selecties die voor studie, werk en burgerschap gemaakt worden, zijn voor een belangrijk deel afwijkend van de keuzes die het onderwijs maakt (6)⁴. Het onderwijs is zeker niet naadloos toegesneden op de eisen die het 'alledaagse leven' aan (jong-)volwassenen stelt. Het hoeft en dient dat op grond van zijn pedagogisch-normatieve opdracht overigens ook niet te zijn. De spanning tussen hetgeen het onderwijs biedt (1) en de samenleving vraagt (3) kan door beter onderwijs Nederlands wel verminderd worden. En gelukkig is later compensatie (zie 4) mogelijk, al wordt die vaak duur betaald.

De eindexamens voortgezet onderwijs Nederlands staan tussen twee vuren. Ze moeten met het verleden en de toekomst rekening houden. Zolang de spanning tussen onderwijsaanbod (1) en naschoolse eisen (3) niet te groot is, kan de selectie die het examenprogramma uit het KV-domein vormt deze twee heren dienen: het onderwijs en de samenleving. Toch zullen examens ook bij overeenstemming van (1) en (3) volgens eigen wetmatigheid selectief zijn (5). Tussen domein en examendoelstellingen staan praktische bezwaren: toetsmogelijkheden en tijd beperken de inhoud van examenprogramma's. Niet alles wat voor toetsing in aanmerking komt, laat zich immers in schoolonderzoeken of centrale examens vangen. Tussen examens en onderwijs (7) en examens en naschools leven (8) zal de overeenstemming daarom altijd onvolledig zijn.

Examenprogramma's en examenmodellen staan van jaar tot jaar aan de wieg van de eigenlijke examens: de opdrachten die leerlingen ter verwerving van veelal zessen en zevens moeten volbrengen. Deze opdrachten vormen een operationalisatie van de doelstellingen. Operationaliseren (9) is specificeren en, opnieuw, selecteren: kiezen uit de vele mogelijke inhouden en vormen die voor toetsing ter beschikking staan.

Examens monden uit in cijfers en beslissingen (10). Deze vervullen, feitelijk of hopelijk, een aantal welbekende functies (De Koning 1978). Voor het onderwijs dient het examen ter afsluiting (11), voor vadertje staat ter niveaubewaking (12), voor hoger onderwijs en arbeidsmarkt ter selectie (13) en allocatie (14). Bezien vanuit de leerlingen geldt de kwalificatie- en de paspoortfunctie van het examen (15 en 16).

Is er toekomst voor de examens Nederlands in het voortgezet onderwijs? Kreeft (1990) heeft het nieuwe examenprogramma terecht een opgave voor Neerlandici genoemd. De cruciale vraag is voor hem of bij de huidige diversiteit en omvang van doelen een curriculum-afhankelijke examinering van het vak Nederlands mogelijk is. Ik denk dat dat alleen maar kan als de inhoudelijke diversiteit een halt toe wordt geroepen. Dat kan in Nederland, zoals ik elders heb geschreven (De Gloppe 1990), niet langs de koninklijke weg, want pogingen tot het vaststellen van uitgewerkte leerplannen en specifieke onderwijsdoelstellingen zijn door de heersende interpretatie van grondwetsartikel 23 tot mislukken gedoemd. Wie de inhoud van het onderwijs wil stroomlij-

nen, moet door de achterdeur naar binnen. Met haar opdracht aan de CVEN heeft de voormalige staatssecretaris Ginjaar-Maas daarvoor gekozen: voor sturing van het onderwijs via het examen. Of dat zal lukken is de vraag. Het zou mij verbazen als de uiteindelijke voorstellen van de CVEN in onverdunde vorm tot ministeriële beschikking worden verheven. Bij de inhoudelijke onderwijsvrijheid passen centrale eindexamens eigenlijk heel slecht. Decentrale, schoolgebonden examens kunnen de afsluitingsfunctie beter vervullen. Zulke examens heten nu schoolonderzoek.

Bieden vernieuwde examens Nederlands betere mogelijkheden tot niveaubewaking? Door hun aard --ieder jaar andere opgaven, die per schooltype door alle leerlingen gemaakt moeten worden-- zijn eindexamens zeer gebrekkige instrumenten voor bewaking van het onderwijsniveau. Veranderingen in het onderwijsniveau kunnen met examens niet vastgesteld worden, omdat opdrachten niet periodiek herhaald worden afgenomen. Door de beperkte toetstijd en -middelen kunnen ook niet alle relevante onderdelen geëxamineerd worden. Voor niveaubewaking is periodiek peilingsonderzoek een veel nuttiger instrument. Er valt bij voorbeeld te denken aan jaarlijks steekproefonderzoek waaraan scholen en leerlingen verplicht zijn mee te doen. Omdat niet alle te onderzoeken leerlingen dezelfde opdrachten hoeven maken, kan in zulk onderzoek een breed scala van kennis en vaardigheden aan bod komen.

Selectie en allocatie zijn onder de heersende omstandigheden vooral functies van schooltype- en vakkenpakketkeuze. De examens spelen hierin nauwelijks een rol (Kreeft 1990). Zullen vernieuwde examens Nederlands deze functies beter vervullen? Ik verwacht niet dat dat het geval zal zijn. Het vak Nederlands draagt niet of nauwelijks bij aan selectiebeslissingen: het regent zessen en zevens. Taalvaardigheidsproblemen in de toekomst, in studie, werk en burgerschap, worden door betere examens niet verholpen. Alleen een drastische heroriëntering van de onderwijsinhoud kan hier iets opleveren, maar dat valt voorlopig (vrijheid van onderwijs!) niet te verwachten. En al zou de inhoud van het onderwijs welbewust op naschoolse eisen worden afgesteld, dan mag met het beperkte aantal uren Nederlands geen wonder verwacht worden. Vaardigheidsonderwijs is intensief en tijdrovend. Het is niet voor niks dat het zo belangrijke onderwijs in spreken en luisteren tot op heden nauwelijks uit de verf komt.

Ligt het niet meer voor de hand om selectie en allocatie te regelen op de plaats waar feitelijk geselecteerd wordt: bij de intree in het hoger onderwijs en op de arbeidsmarkt? In werkelijkheid gebeurt dit naar alle waarschijnlijkheid al. De samenhang tussen taalvaardigheid en doorstroming in het hoger onderwijs (Baltzer e.a. 1988) vormt daar een aanwijzing voor.

Zullen vernieuwde examens een verbeterd paspoort opleveren? En zal de kwalificatiefunctie beter gediend worden? Ook dat is zeer de vraag. Bij doorvoering van de voorstellen van de CVEN zal het paspoort wel veel beter ogen, maar geen toegang bieden tot terreinen die voorheen verboden gebied waren. Ook een verblijfsvergunning in

hoger onderwijs of werk valt niet te verdienen. Zullen nieuwe examencijfers aantonen dat leerlingen essentiële kennis en vaardigheden verworven hebben? De bewijskracht van toekomstige vijven en zessen zal afhangen van de meettechnische kwaliteit van nieuwe examens. Daarover wil ik in de volgende paragraaf nog een paar opmerkingen maken. Ter afsluiting van de bovenstaande, wat pessimistische bespiegeling, wil ik de volgende keuze ter overweging geven.

- a. Schoolgebonden curriculum-afhankelijke examens die dienen ter afsluiting van het gegeven onderwijs, gecombineerd met periodiek peilingsonderzoek aan het einde van het voortgezet onderwijs als middel voor niveaubewaking, en aangevuld met entree- of instaptoetsen voor het hoger onderwijs.
- b. Terugdringing van de inhoudelijke vrijheid van onderwijs, centrale afstemming van de onderwijsinhoud op de eisen die het naschoolse leven aan taalkennis en -vaardigheden stelt, en centrale eindexamens als natuurlijk scharnierpunt tussen voortgezet en hoger onderwijs.

2. Over validiteit en betrouwbaarheid

Aan examens moeten eisen betreffende communiceerbaarheid, validiteit en objectiviteit gesteld worden (Kreeft 1990). In reactie op Kreeft wil ik het hier hebben over validiteit en betrouwbaarheid. Objectiviteit lijkt me niet de juiste term: toetsen en examens kunnen immers tegelijk objectief en onbetrouwbaar zijn, of betrouwbaar en niet-objectief.

Validiteit of geldigheid is in het geding als we aan een toetsscore een interpretatie willen hechten. Of een examenopdracht inderdaad (bepaalde aspecten van de) schrijfvaardigheid meet, is een validiteitskwestie. Over de validiteit van het examenprogramma kan niet zinvol gesproken worden, wel over de relevantie of representativiteit. Hierbij is de relatie tussen (1), (3) en (5) in figuur 1 in het geding. Het is denkbaar dat het examenprogramma doelstellingen bevat waarvoor geen valide meetinstrumenten bestaan, of doelstellingen die zo vaag geformuleerd zijn, dat de validiteit van examenscores niet goed te evalueren valt. Dat alles betekent echter niet dat het validiteitsprobleem alleen of in de eerste plaats in de examenprogramma commissie thuishoort. Toetsconstructeurs kunnen door hun operationalisatie van examendoelstellingen (9) de validiteit maken en breken. In directe zin valt de validiteit van examenopdrachten te evalueren door de feitelijke opdrachten en meetresultaten te evalueren in het licht van de doelstellingen in het examenprogramma. Examenopdrachten mogen inderdaad valide genoemd worden als ze de doelstellingen uit het examenprogramma adequaat representeren (vgl. Kreeft 1990).

Deze directe vorm van 'interne' validiteit is heel belangrijk, maar uiteindelijk niet de meest interessante. Belangwekkender zijn vragen die zich alleen indirect laten beantwoorden. Is het examen constructvalide: zijn de examenopdrachten valide operationalisaties van onderde-

len van het KV-domein? Is het examen curriculumvalide en 'ecologisch' of 'maatschappelijk' valide: representeren de opdrachten de inhoud van het gegeven onderwijs en de eisen van het naschoolse leven? Bij deze verder reikende vragen is niet alleen de kwaliteit van de operationalisaties in het geding, in het antwoord moet ook betrokken worden in hoeverre onderwijs, examenprogramma en naschools leven op overeenkomstige wijze uit het KV-domein selecteren.

Betrouwbaarheid (liever dan objectiviteit) is een noodzakelijke voorwaarde voor validiteit. Dat wil echter niet zeggen dat de twee altijd in elkaars verlengde liggen. Wesdorp (1981) sprak in dit verband zelfs over een conflict tussen validiteit en betrouwbaarheid. Kreeft neemt een diametraal ander standpunt in, waar hij stelt dat globale beoordeling "vanuit het oogpunt van objectiviteit en *derhalve* ook van validiteit" (Kreeft 1990, 40; cursivering KdG) ongewenst is.

De kern van de problematiek is de mate van vrijheid van beoordelaars en beoordeelde. Door het opstellen van beoordelingsvoorschriften, correctiemodellen, schalen en wat dies meer zij, wordt de vrijheid van beoordelaars teruggedrongen, in de hoop dat dit leidt tot meer gewenste en minder ongewenste variatie in oordelen. De uiterste vrijheidsbeperking is te vinden in de objectieve toets met mechanische sleutels. Beperking van de vrijheid van de beoordeelde heeft als doel dat leerlingen gedwongen worden om te laten zien of zij een welomschreven iets weten of kunnen. Ook hier gaat het om het vergroten van gewenste en het terugdringen van ongewenste variatiebronnen.

De vrijheid van beoordelaar en beoordeelde kunnen niet onafhankelijk gevarieerd worden. Verkleining van de beoordelaarsvrijheid gaat in de regel gepaard met meer gebondenheid voor de beoordeelde. En gebondenheid van de beoordeelde maakt beperking van beoordelaarsvrijheid beter mogelijk. Beperking van de beoordelaarsvrijheid kan in de literatuur op ruime instemming rekenen. Toch zijn er ook wel tegengeluiden (Meuffels, 1985). De meeste vragen worden echter opgeroepen door onvrijheid van de beoordeelde. De discussie over objectieve toetsen voor het meten van schrijfvaardigheid getuigt daarvan (Van Schooten & De Glopper 1990).

De discussie over de vrijheid van de beoordeelde wordt vooral in termen van meer en minder gevoerd. Dat verhindert het zicht op het eigenlijke vraagstuk dat ik de 'locatie' van de vrijheid zou willen noemen.

Bij open, bepaalde en gestructureerde schrijfopdrachten (Schoonen 1991) wordt aan leerlingen een schrijfopdracht verstrekt met aanwijzingen omtrent de gewenste inhoud, het te schrijven type tekst, het doel en het publiek. Naar aanleiding van de opdracht vervaardigen de leerlingen schrijfprodukten die vervolgens beoordeeld worden. De schrijfopdracht laat de schrijver meestal een zekere vrijheid in de wijze waarop de taak uitgevoerd moet worden. Bij correctie-opdrachten en bij gesloten opdrachten zijn de schrijfsituatie en de te produceren tekst in veel verdergaande mate vastgelegd: er moeten tekortkomingen gecorrigeerd of alternatieven beoordeeld worden in teksten die speciaal voor toetsdoeleinden zijn samengesteld. De vrijheid van de

schrijver is bij deze indirecte methoden veel verder ingeperkt. Een voorbeeld met betrekking tot de spelling kan duidelijker maken op welke plaats directe en indirecte methoden in vrijheid verschillen.

In teksten die naar aanleiding van opdrachten geschreven worden, kunnen spelfouten geteld worden. Zulke foutentellingen kunnen als indicatie voor de spelvaardigheid van schrijvers opgevat worden, net als prestaties op een dictee of een andersoortige spellingtoets. Bij een dictee kunnen schrijvers hun woorden niet zelf kiezen, bij een zelf te schrijven tekst tot op zekere hoogte wel. In een dictee kunnen bepaalde probleemttypen opzettelijk vertegenwoordigd worden; ook kan de moeilijkheidsgraad van de items (de te spellen woorden) door de toets-constructeur gecontroleerd worden. Bij een zelf te schrijven tekst kunnen bepaalde probleemttypen door de schrijver al dan niet opzettelijk vermeden worden; bij een zelf te schrijven tekst heeft de schrijver de moeilijkheidsgraad van de te schrijven woorden (de items) tot op zekere hoogte in de hand. Op deze plaats heeft de schrijver derhalve veel vrijheid. De kans is daarom groot dat de spelling van het merendeel van de te schrijven woorden geen probleem vormt. Ondanks de schijnbare lengte is de test daarom (te) kort.

Gaat het niet om de spelling, maar om de inhoud of compositie van teksten dan dienen schrijvers opnieuw op bepaalde plaatsen juist wel en op andere weer niet vrij gelaten te worden. Zo lijkt variatie in kennis over het onderwerp waarover geschreven moet worden, mij eigenlijk ongewenst. Die variatie kan teruggedrongen worden door leerlingen gedocumenteerde opdrachten aan te bieden: niet alleen maar een specificatie van het onderwerp, het doel, het publiek en de aard van de te schrijven tekst, maar ook een kennisbestand (b.v. een verzameling feiten, citaten) met daarbij een gerichte opdracht tot selectie en ordenen van relevante informatie. Moet de stilistische kwaliteit van het taalgebruik beoordeeld worden, dan zijn juist weer andere vormen van vrijheid en gebondenheid nodig. Afhankelijk van het te meten aspect van schrijfvaardigheid moeten steeds weer andere bronnen van variatie geneutraliseerd worden. Het gaat niet in de eerste plaats om de mate, maar om de locatie van de vrijheid.

Noten

1. De inhoud en de structuur van het KV-domein voor het vak Nederlands laten zich niet eenvoudig omschrijven. Welke leerstof maakt er deel van uit? Welke niet? Hoe laat het domein zich beschrijven? De dissertatie van Sijstra (1990) behandelt dergelijke problemen.
2. Deze variatie hangt o.a. sterk samen met verschillen in leerresultaten.

3. Het naschoolse leven is hier samengevat onder de noemers studie, werk en burgerschap (cf. De Glopper & Van Schooten 1990).
4. Vgl. resultaten van onderzoek van Blok & De Glopper (1983) en De Glopper & Van Schooten (1990).

Bibliografie

- Baltzer, J., K. de Glopper & E. van Schooten. *Taalvaardigheid in het hoger onderwijs*. Amsterdam: SCO, 1988.
- Blok, H. & K. de Glopper. *Taal voor het leven*. Harlingen: Flevodruk, 1983.
- Glopper, K. de. Het kan nog beter... In: Hulshoff, H. & G. Rijlaarsdam (red.) *Vernieuwing van het eindexamen Nederlands havo en vwo. Discussiebijdragen*. Amsterdam, Instituut voor de Lerarenopleiding van de Universiteit van Amsterdam, 1990, 85-88.
- Glopper, K. de & E. van Schooten. *De inhoud van de examens Nederlands voor HAVO en VWO*. Amsterdam: SCO, 1990.
- Koning, P. de. Functies van afsluitingen: een opsomming. *Pedagogische Studiën*, 1978, 433-449.
- Kreeft, H. Een nieuw examenprogramma Nederlands: Een opgave voor Neerlandici! *Spiegel*, 8, 1990, 31-45.
- Meuffels, B. Globaal versus analytisch beoordelen. *Moer*, 1985, 5, 9-14.
- Schooten, E. van & K. de Glopper. De validiteit van meerkeuze-instrumenten voor het meten van schrijfvaardigheid. *Tijdschrift voor Taalbeheersing*, 12, 1990, 93-110.
- Schoonen, R. *De evaluatie van schrijfvaardigheidsmetingen*. Amsterdam: SCO, 1991.
- Sijtsma, J. *Doel en inhoud van taalonderwijs*. Arnhem, Cito, 1990.
- Wesdorp, H. *Evaluatietechnieken voor het moedertaalonderwijs*. Den Haag, Staatsuitgeverij, 1981.
- Wesdorp, H. *Research vs. Tradition: The Unequal Fight. The Evaluation of the Traditional Grammar Curriculum in the Netherlands*. Amsterdam: SCO, 1984.