

Oplosbare en niet-oplosbare (echte) dilemma's

bij het beoordelen van schrijf- en spreekproducten

14 sept '12
Les 178

Op de conferentie BVNt2 werden twee workshops gegeven over het beoordelen van spreek- en schrijfproducten, door Jacqueline de Maa, Tiba Bolle en Inge van Meelis van het ITTA. In de workshops beoordeelden zij met de deelnemers een aantal schrijfproducten en presentaties van studenten. Aan de hand daarvan kwam de discussie goed op gang. In de workshops kwam naar voren welke ingewikkelde beslissingen een assessor soms moet nemen. Ook is besproken wat je kan doen om een beoordeling zo betrouwbaar mogelijk te laten zijn. Een kort verslag.

Allereerst schetsen we de kaders. Een taalassessor in het mbo en vmbo beoordeelt taal aan de hand van het REFERENTIEKADER TAAL EN REKENEN¹. Het Referentiekader beschrijft de eindniveaus die in het onderwijs behaald moeten worden. In het (v)mbo werkt men toe naar niveau 2F, alleen voor de mbo-4 opleidingen wordt niveau 3F als eindniveau gevraagd. Deze niveaus zijn enigszins te vergelijken met respectievelijk niveau B1 en niveau B2 uit het RAAMWERK NT2. De beschrijvingen in het REFERENTIEKADER TAAL EN REKENEN zijn globaal, ze zijn niet geformuleerd in de vorm van beoordelingscriteria.

Hiermee hebben we gelijk het eerste dilemma te pakken. Hoe kunnen we bepalen wat er precies met deze globale niveaubeschrijvingen wordt bedoeld en wat de exacte niveaugrenzen zijn? Een beoordelingsformat, waarin de globale beschrijvingen herschreven zijn tot beter meetbare beoordelingscriteria is onmisbaar om een beslissing te nemen. Hoe kom je tot een cijfer (bijvoorbeeld 'een 6 op 2F' of een '5 op 3F'), zoals gevraagd wordt in de instellingsexamens spreken en schrijven? Met deelnemers zijn we dieper ingegaan op deze vragen door concrete schrijfproducten en opnames van studenten gezamenlijk te beoordelen.

Geen eenduidige beoordelingen

De deelnemers bestuderen een voorbeeldtekst van een student uit de HANDREIKING BIJ DE INSTELLINGSEXAMENS MBO². We vergelijken vervolgens hun ideeën over niveau 2F en 3F met de in de handreiking bijgevoegde normbeoordeling. Er zijn verschillende kenmerken waarop de tekst beoordeeld wordt en waarop de student punten kan krijgen. De kenmerken van de uitvoering (zie kader) blijken voor de ongeoefende beoordelaar op verschillende manieren te interpreteren. Neem een tekst gericht op een jong lezerspubliek, zoals een folder over SOA's. Hoe erg is het dan om 'kan je' te schrijven in plaats van 'kun je'? Reken je het fout, is dat dan een grammaticale fout of een minpunt op het kenmerk 'afstemming op de doelgroep'? Of betreft het een fout in woordkeuze? Hoeveel keer mag zo'n soort fout eigenlijk terugkomen in een tekst? En telt dezelfde fout maar één keer mee? Dit zijn vragen waarop niet makkelijk eenduidig antwoord te geven is. De kenmerken in het referentiekader zijn bedoeld om een beschrijving te geven van de niveaus. Ze zijn niet geschreven als toetscriteria, daarvoor zijn ze niet specifiek genoeg geformuleerd.

Dat betekent dat met het ene format de beoordeling veel milder kan uitpakken dan met het andere, omdat er bijvoorbeeld gecompenseerd mag worden



KENMERKEN VAN DE UITVOERING (uit het beoordelingsformat van de Handreiking Instellingsexamens)

OPDRACHTSPECIFIEKE INHOUDSKENMERKEN

Samenhang

- Gebruikt veelvoorkomende verbindingswoorden (als, hoewel) correct.
- De tekst bevat een volgorde: inleiding, kern en slot.
- Kan alinea's maken en inhoudelijke verbanden expliciet aangeven.
- Maakt soms nog onduidelijke verwijzingen en fouten in de structuur van de tekst.

Afstemming op doel

- Kan in teksten met een eenvoudige lineaire structuur trouw blijven aan het doel van het schrijfproduct.

Afstemming op publiek

- Past het woordgebruik en toon aan op het publiek.

Woordgebruik en woordenschat

- Varieert het woordgebruik, fouten met idiomatische uitdrukkingen komen nog voor.

Spelling, leestekens en grammatica

- Vertoont een redelijke grammaticale beheersing.

Leesbaarheid

- Gebruikt titel en tekstkopjes.
- Heeft bij langere teksten (meer dan twee x A4) ondersteuning nodig bij aanbrengen van lay-out.

Mild of streng

We praten door over de formats. Er circuleren verschillende beoordelingsformulieren, uiteenlopend van heel streng (onvoldoende voor één kenmerk betekent sowieso gezakt) tot

soepel (je mag onvoldoendes compenseren). Tekstkenmerken worden dus niet in ieder format even streng afgerekend. Sommige formats zijn heel algemeen, andere juist heel specifiek, met uitgewerkte lijsten van verschillende soorten fouten en criteria met absolute aantallen of percentages ('maximaal 5% van de woorden mag fout geschreven zijn'). Dat betekent dat met het ene format de beoordeling veel milder kan uitpakken dan met het andere, omdat er bijvoorbeeld gecompenseerd mag worden. Ook door het splitsen van kenmerken in twee aparte criteria kunnen cijfers heel anders uitvallen. Neem bijvoorbeeld een tekst waarin wel veel variatie in woordgebruik is maar veel fouten voorkomen in woordkeuze. In het ene format kun je voor beide aspecten één apart punt krijgen, in de andere moet je middelen om tot een score te komen. Wat weegt dan zwaarder? Het is aan de assessor.

Typische Nt2-fouten

Voor de deelnemers aan de BVNt2-conferentie is het natuurlijk vooral interessant om een schrijfproduct van een Nt2-leerder te beoordelen. Vragen die leven bij de deelnemers aan de conferentie waren: Wat betekenen de verhoogde eisen voor Nederlands voor een Nt2-leerder in het (v)mbo? Hebben deze studenten nog wel een kans om hun examen Nederlands en daarmee een diploma te halen? We bekeken gezamenlijk een brief van een mbo-student met een

Nt2-achtergrond. De brief die zij moest schrijven kwam uit een toets waarin gevraagd werd aan de gemeente toestemming te vragen om een muur te schilderen en te informeren naar de beschikbare subsidie. Hieronder een fragment uit haar brief:

Wij bewoners van Raadhuyslaan 4 willen graag aan gemeente toestemming vragen voor het muur schilderen. Daroem frag wij ook als er is een subsidie beschikbaar zodat kunnen de kosten ontdekken. en hoe komen wei aan? Ik hope dat zo snel mogelijk een reactie krijgt.

Natuurlijk springen bij deze brief de fouten in spelling, zinsbouw en woordkeus direct in het oog. Duidelijk is dat het om een schrijver met een andere moedertaal gaat. De vraag is welke fouten en hoeveel fouten zijn toegestaan op 2F. We beoordelen ook deze tekst met het beoordelingsformat uit de voorbeeldexamens (zie kader 1). De zwakke punten uit deze brief (spelling en zinsbouw) behoren bij het kenmerk grammatica. We besluiten gezamenlijk daar geen punten aan toe te kennen. Ondanks de vele fouten kan deze student wel degelijk slagen voor 2F als zij twee punten haalt op een ander kenmerk, bijvoorbeeld afstemming op doel. Dus: wie een samenhangende tekst, goed afgestemd op doel en publiek, weet te schrijven, maakt met veel spelfouten toch een kans om nog een voldoende te halen. Dat was in dit geval niet aan de orde,



maar het biedt Nt2-studenten de kans om ondanks hun moeite met spelling en grammatica toch 2F of zelfs 3F te behalen. Voor sommige docenten is dit een opluchting, voor anderen druist het in tegen hun rechtvaardigheidsgevoel.

Spreken en Gesprekken

Ook bij Spreken en Gesprekken is het voor Nt2-sprekers mogelijk om hardnekkige fouten in zinsbouw en spelling te compenseren door extra goed te letten op de andere beoordelingscriteria. Maar er zijn nog andere dilemma's. In de workshop bekijken we twee fragmenten. Een van de criteria bij Spreken luidt bijvoorbeeld 'kan luisteraars boeien door concrete voorbeelden en ervaringen' (kenmerk van de uitvoering 'afstemming op publiek' op niveau 2F). Dit kenmerk roept altijd discussie op. Moet een presentatie op 2F echt boeiend zijn en zo ja, hoe beoordelen we dat dan? Is het voldoende als over concrete voorbeelden en ervaringen verteld wordt? En wat doe je als een student door zenuwen of uit het hoofd leren weinig inspirerend spreekt? Spelen houding, toon en intonatie ook een rol bij het 'boeien' in dit beoordelingscriterium? We kijken naar een presentatie van een zenuwachtig meisje dat iets over haar eetpatroon vertelt. De beoordelingen in de groep blijken ver uiteen te lopen, van een 4 op 2F (ge-zakt) tot een 7 op 2F (geslaagd).

Oorzaak voor de verschillen ligt in de eisen die men aan grammaticale correctheid stelt, maar ook is er verschil in interesse in het onderwerp tussen de beoordelaars. Het gevoel van sympathie voor het meisje dat zo haar best doet en volgens sommige deelnemers veel beter kan, verhoogde soms de beoordeling.

Uitdagen tot een gesprek op niveau

Ook de opdracht die de spreker krijgt, kan de beoordelaar voor een aantal dilemma's plaatsen. Een opdracht moet duidelijkheid geven over het publiek en aanwijzingen bevatten over hoe lang iets moet zijn en welke hulpmiddelen wel of niet zijn toegestaan. In sommige examenopdrachten wordt de student onvoldoende uitgedaagd en krijgt hij niet de kans om te laten zien of hij voldoet aan de alle criteria. Of een opdracht is te eenvoudig en kan ook uitgevoerd worden op een lager taalniveau dan de toets vereist.

We bekijken gezamenlijk een reflectiegesprek tussen een docent en een student. De student neemt weinig initiatief en heeft een beperkte inbreng, iets wat op 3F wel verwacht wordt. De discussie ontstaat of het gebrek aan initiatief van de student veroorzaakt wordt door een taalprobleem van de leerling of door de dominante houding van de docent die alleen maar gesloten vragen stelt. We zijn het er over eens dat je bij het beoor-

delen van gesprekken afhankelijk bent van de gesprekspartner. In hoeverre kun je dit de student aanrekenen? Duidelijke instructies voor de beide sprekers is dus van belang.

Tips voor assessoren

Gaandeweg verzamelen we met de deelnemers adviezen voor de beoordelaars. Hieronder volgt een greep uit de verzamelde tips:

- Beoordeel wat je ziet/hoort, niet wat je denkt dat de student zou kunnen (in een andere situatie of met een andere gesprekspartner).
- Blijf dicht bij het referentiekader en laat je onderwijservaring (het beeld dat je hebt van studenten) erbuiten.
- Zorg dat je consistent blijft beoordelen, elke keer weer (wees voor iedereen even streng of soepel).
- Straf niet dubbel af, laat een zwakke score op één kenmerk niet je oordeel over alle andere kenmerken negatief beïnvloeden.
- Zorg dat je een prestatie uitlokt op het juiste niveau, geef een goede opdracht en let op de rol, status en het niveau van de gesprekspartner.
- Beoordeel voldoende taalmateriaal: geef meerdere opdrachten en bepaal de minimale tijdsduur of het aantal woorden.
- Onderbouw je beoordeling en houd je aan het beoordelingsformulier (noteer voorbeelden van uitingen die je boven of beneden het niveau vindt).

Spelen houding, toon en intonatie ook een rol bij het 'boeien' in dit beoordelingscriterium?

- Bouw routine op door verschillende prestaties naar aanleiding van dezelfde opdracht op een zelfde aspect te beoordelen.

Niet oplosbare problemen (echte dilemma's)

De deelnemers zijn het er over eens: Een goede kennis van het Nederlands, zowel op het gebied van functioneel Nederlands (tekstopbouw, toon) en taalverzorging zijn de noodzakelijke basis voor een goede beoordeling. Uit de trainingen die het ITTA verzorgt voor taalassessoren, blijkt dat veel gericht oefening en overleg tussen beoordelaars essentieel is om te komen tot een gezamenlijke norm. Toch blijven er zich altijd dilemma's voordoen. De omschrijvingen van het Referentiekader zijn immers niet absoluut (niet kwantificeerbaar) en niet bedoeld om te toetsen. Een paar voorbeelden:

- De weging van de aspecten genoemd binnen een kenmerk van de uitvoering is subjectief en lastig. De tijd ontbreekt vaak om alles door een tweede assessor te laten beoordelen.

- Het komt regelmatig voor dat je iets niet kunt beoordelen omdat het niet zichtbaar is.
- Opnames van goede kwaliteit maken is tijdrovend en niet altijd mogelijk (maar wel nodig om rustig te beoordelen).
- Kennis van het onderwerp kan de ene student voorsprong geven op de andere. Wat ben je dan eigenlijk aan het beoordelen?
- En dan is er nog de kwestie van normering. Je kunt het niveau behalen met een zesje. Moet je om een niveau te beheersen niet gewoon aan alle aspecten voldoen?

De deelnemers zijn erg enthousiast over het samen beoordelen van studenten, ondanks de dilemma's waar niet een twee drie een antwoord op is te geven. Het gezamenlijk beoordelen levert interessante en levendige discussies op die leiden tot beter inzicht in het referentiekader en de niveaus. Een exercitie die onmisbaar is voor taaldocenten die hun studenten adequaat willen beoordelen en de juiste feedback willen geven.

*Inge van Meelis
Tiba Bolle*

*De auteurs zijn beiden senior projectleider bij ITTA Amsterdam
Correspondentie: e.c.bolle@uva.nl*

Noten

1. REFERENTIEKADER TAAL EN REKENEN, IN OPDRACHT VAN OCW, oktober 2009.

Zie: www.taalenrekenen.nl. Het referentiekader is ook wel bekend onder de naam van de voorzitter van de adviesgroep, Meijerink. De niveau-eisen zijn 1F in het basisonderwijs, 2F in vmbo en mbo-opleidingen op niveau 1, 2 en 3, 3F voor havo en mbo-4 opleidingen, 4F voor vwo.

2. INSTELLINGSEXAMENS NEDERLANDS IN HET MBO, EEN HANDREIKING.

Cito/ Cinop, november 2011.

Zie: www.steunpunttaalenrekenen.nl, onder: Examinering; schoolexamens; complete set voorbeeldexamens.

