

4 Werken met beoordelingsmodellen voor productieve vaardigheden

Inleiding

Voor het vak Nederlands in het mbo is het Referentiekader Nederlandse taal de basis vormt voor de examinering. Hieronder lichten we toe hoe de beoordeling van de kandidaatprestaties zo betrouwbaar mogelijk kan verlopen. Deze voorbeelden zijn een handreiking, geen regelgeving.

4.1 3-staps beoordelingsmodel

De beoordeling van toetsprestaties voor de productieve vaardigheden Schrijven, Gesprekken voeren en Spreken gebeurt met behulp van een 3-staps beoordelingsmodel.

De eerste twee stappen betreffen precondities.

- Als eerste wordt de leesbaarheid of de verstaanbaarheid van de prestatie beoordeeld. Denk bij leesbaarheid aan een leesbaar handschrift, maar ook aan een worddocument dat zo is opgemaakt dat het communicatief bruikbaar is.
- Als tweede de adequaatheid. Met adequaatheid bedoelen we dat de prestatie bij de toetsopdracht moet passen. Vraagt de opdracht om een klachtenbrief en wordt een klant memo opgeleverd, dan is de opdracht niet adequaat uitgevoerd. Ook de lengte van de opgeleverde tekst kan een reden zijn om de prestatie af te keuren. De opdracht vereist bijvoorbeeld een verslag van minimaal 400 en maximaal 600 woorden. Als de leerling zich niet aan deze grenzen houdt, dan wordt deze preconditie negatief beoordeeld.

Als de kandidaat op één van deze precondities negatief scoort, wordt de beoordeling niet verder voortgezet. Voldoet de prestatie aan de gestelde precondities, dan volgt stap 3 van de beoordeling, waar wordt bepaald of de geleverde prestatie voldoende scoort op het beoogde niveau van het Referentiekader Nederlandse taal.

4.2 Relatie Referentiekader Nederlandse taal en de beoordelingsaspecten

In stap 3 zien we de directe relatie met het Referentiekader Nederlandse taal: de standaard bij de beoordeling. De kenmerken van de taakuitvoering uit het Referentiekader Nederlandse taal vormen de beoordelingsaspecten.

Stap 3 uit het beoordelingsmodel is als volgt opgebouwd:

1 Inhoudskenmerken uit de toetsopdracht.

In de beschrijvingen van de referentieniveaus is weinig aandacht voor de 'inhoud' van de kandidaatprestatie. Die inhoud krijgt in dit beoordelingsmodel expliciet aandacht door de te beoordelen in welke mate de prestatie volledig en correct is.

2 Dan volgen de talige beoordelingsaspecten die rechtstreeks zijn gelinkt aan de kenmerken van taakuitvoering uit het Referentiekader Nederlandse taal. Per taalvaardigheid zijn er vijf (Spreken en Gesprekken voeren) of zes (Schrijven) aspecten.

Let op: het kan zo zijn dat in een toetsopdracht niet de complete beschrijving van een beoordelingsaspect aan de orde komt. In dat geval geldt: wat niet gevraagd wordt, kan ook niet worden beoordeeld.

De inhoudskenmerken zijn cruciaal. Dat wil zeggen dat voor dit onderdeel minimaal 1 punt (voldoende) gescoord moet worden. Is dat niet het geval, dan wordt de opdracht als onvoldoende beoordeeld en wordt de beoordeling gestopt.

4.3 Globaal beoordelen met een 3-puntsschaal

Aan de hand van de beoordelingscriteria en scoreschalen wordt de taalprestatie beoordeeld. De beoordelingsmodellen geven een globale omschrijving van onvoldoende, voldoende en goed (voorheen excellent).

Voor een betrouwbaar en evenwichtig beoordelingsmodel is een goed evenwicht nodig tussen beoordelingsaspecten (verticale as) en schaalpunten (horizontale as) om een oordeel uit te drukken. Theoretisch geldt het principe dat kan worden volstaan met een korte scoreschaal, als er maar voldoende beoordelingsaspecten zijn.

Voor de afzonderlijke beoordeling van de prestaties op de verschillende aspecten hebben we voor een 3-puntsschaal gekozen. Een omschrijving van de schaalpunten:

0 - onvoldoende

De kandidaat voldoet niet of nauwelijks aan de beschrijving die hoort bij de beheersing van het betreffende aspect op het getoetste niveau in het Referentiekader Taal. De kandidaat beheerst het niveau op dat aspect dus in meer of mindere mate niet.

1 - voldoende

De kandidaat voldoet aan de beschrijving die hoort bij de beheersing van het betreffende aspect op het getoetste niveau in het Referentiekader Taal. Hij doet wat op dat niveau bij het aspect past, maar niet meer dan dat. De kandidaat beheerst dus het niveau op dat aspect voldoende.

2 - goed

De kandidaat voldoet volledig aan de beschrijving die hoort bij de beheersing van het betreffende aspect op het getoetste niveau in het Referentiekader Taal en presteert daarbij optimaal binnen de kaders van het niveau. Hij doet wat op dat niveau bij het aspect past, maar laat een prestatie zien die op dat aspect meer, complexer en/of uitgebreider is dan dat wat er in de beschrijving staat (MITS adequaat uitgevoerd). De kandidaat beheerst dus het niveau op dat aspect goed.

LET OP: Een aspect hoeft binnen een prestatie op een bepaald niveau **NIET** op het volgende niveau te worden beheerst om 'goed' te scoren. De voorgelegde opdracht is immers op het getoetste niveau geconstrueerd en kan daarmee dus ook niet valide en niet betrouwbaar meten of de kandidaat het volgende niveau beheerst.

De keuze voor een 3-puntsschaal

In de praktijk wordt er soms makkelijk gedacht over het wijzigen van de 3-punts-beoordelingsschaal in een andere beoordelingsschaal, zoals een 4-puntsschaal. Er wordt dan bijvoorbeeld van uitgegaan dat een 4-puntsschaal per definitie eerlijker is voor een kandidaat, omdat er meer punten te vergeven zijn. Dat is niet het geval.

Hieronder vindt u een aantal ondersteunende argumenten vóór de 3-puntsschaal ten opzichte van de 4-puntsschaal:

- Het Referentiekader Taal is een functioneel kader, dat past bij de functionele en competentiegerichte benadering/beoordeling die veelal in het mbo wordt toegepast. Er wordt beoordeeld of de kandidaat kan functioneren op een bepaald taalniveau binnen contexten die er toe doen. Hij/zij kan het, of kan het niet. Het is een logische keus om een puntsschaal te hanteren die hierbij aansluit.
- De bestaande 3-puntsschaal heeft een ordening die aansluit bij een competentiegerichte benadering. De drie scores op deze schaal, 0-1-2, geven aan of de kandidaat in het licht van het getoetste niveau onvoldoende (0), voldoende (1) of goed (2) heeft gepresteerd. Scoort de kandidaat op alle beoordelingsaspecten een voldoende, dan krijgt hij voor zijn prestatie een 6: hij beheerst (net) het niveau. Werkt de kandidaat één of twee aspecten beter of uitgebreider uit dan verwacht, dan krijgt hij een hoger cijfer: de kandidaat doet immers (iets) meer dan mag worden verwacht op het niveau dat wordt getoetst. Hij presteert (ten minste op die aspecten) niet voldoende, maar goed. Werkt de kandidaat één of twee aspecten onvoldoende uit, dan krijgt hij een onvoldoende: de kandidaat presteert immers onder het getoetste niveau.
- Hoe meer punten er te vergeven zijn op een schaal, hoe meer variatie in interpretatie van de verschillende scores. Houd je de beoordelingsschaal klein, dan is de betekenis van de punten op de schaal eenduidiger, zal deze meer uniform geïnterpreteerd worden en is het voor beoordelaars makkelijker om op één lijn blijven in hun beoordelingen. Zeker wanneer het kader waarop de schaal gebaseerd is, in dit geval het Referentiekader taal, op sommige punten onduidelijk of onspecifiek is omschreven.
- Onder elke schaal hangt een verklaring/duiding van de schaalpunten. Een wijziging in schaalpunten brengt altijd een herziening van die duiding met zich mee. Zo'n herziening vraagt om een weloverwogen afweging.
Als je een 4-puntsschaal gaat gebruiken, moet je de schaalpunten herdefiniëren. Je zou bijvoorbeeld kunnen kiezen voor de vier punten 0-1-2-3. Maar: wat is de betekenis/waarde van een 1 dan? Wanneer verdient de kandidaat op een bepaald aspect een 2? Of een 3? Krijgt een kandidaat 0 of 1 punt bij een onvoldoende prestatie en 2 en 3 bij een voldoende prestatie? De vraag moet dan beantwoord worden wanneer een prestatie net onvoldoende (1) en wanneer zwaar onvoldoende (0) is. Hiervoor geeft het Referentiekader onvoldoende aanknopingspunten.
Of staat 0 voor een onvoldoende prestatie en 1-2-3 een voldoende prestatie? De vraag moet dan beantwoord worden wanneer een prestatie net voldoende (1), ruim voldoende (2) of goed (3) is. Ook hiervoor biedt het Referentiekader te weinig houvast. Bovendien geldt voor beide voorbeelden: waar leg je de cesuur (zak/slaaggrens) en wat houdt die in?
- Mocht een school/opleiding toch de keuze maken voor een 4-puntsschaal, dan zijn er nog twee andere punten om rekening mee te houden :
 - De 3-puntsschaal is in de ontwikkelingsfase van de voorbeeldtoetsen instellingsexamens gepretest onder tientallen leerlingen. De resultaten zijn door experts besproken, die gezamenlijk een cesuur hebben vastgesteld. Van een 4-puntsschaal ontbreken deze gegevens.

- o Bij die cesuur is een passende omzettingstabel naar cijfers ontwikkeld. Die omzettingstabel kan niet zonder meer gebruikt worden voor een 4-puntsschaal. Daarvoor is een rekenmodel nodig dat de ruwe scores (het puntenaantal dat een kandidaat behaalt) kan omzetten naar schaalpunten.

Consensus

Tot slot willen we wijzen op de noodzaak van discussie en training. De enige manier waarop echt betekenis kan worden gegeven aan de schaalpunten op een beoordelingsschaal, is door voldoende af te stemmen met collega-beoordelaars. Alleen zo is het mogelijk om een gezamenlijk beeld te ontwikkelen van de scores 0-1-2, oftewel onvoldoende, voldoende en goed. Hogere beoordelaarsovereenstemming zorgt namelijk voor een hogere betrouwbaarheid van een examen. Dit kan natuurlijk door alle kandidaten gezamenlijk te beoordelen, maar bijvoorbeeld ook door regelmatig beoordelingen met elkaar na te bespreken, door organisatiebreed een terugkerend assessorenoverleg in te plannen of door intervisie. De website www.gripoptaal.nl biedt hierbij ondersteunend materiaal. Tot slot is investeren in een training rondom beoordeling van taalvaardigheid een prima start voor die permanente discussie.

4.4 Taalverzorging bij schrijven beoordelen

Voor de categorie: spelling, leestekens en grammatica bij de subvaardigheid Schrijven is voor een iets andere aanpak gekozen. Het Referentiekader Nederlandse taal is vrij expliciet over Taalverzorging en werkt deze onderdelen zeer gedetailleerd uit (zie Hoofdstuk 4). Omdat met name in schrijfproducten fouten in deze categorie als zeer storend worden ervaren, zijn in de [beoordelingsmodellen voor schrijven](#) de grenzen van de scoreschalen aangegeven. Voor de subcategorieën spelling, interpunctie en grammatica is het aantal toegestane fouten genoemd en niet een percentage. Door punten te tellen objectiever je de meting in die zin, dat alle fouten geteld worden die je (in die categorie) tegenkomt. Natuurlijk heeft ook deze methode het nadeel, dat de kandidaat bij langere teksten meer kans heeft om fouten te maken.

Uit het beoordelingsmodel Schrijven:

1 Spelling, leestekens en grammatica

Vertoont een betrekkelijk grote beheersing van de grammatica. Incidentele vergissingen, niet-stelselmatige fouten en kleine onvolkomenheden in de zinsstructuur kunnen nog voorkomen.

Zie verder: Referentiekader Nederlandse taal Hoofdstuk 4: Begrippenlijst en Taalverzorging

Beoordeel eerst de 3 categorieën Spelling, Leestekens en grammatica en noteer dan in de rechterkolom met behulp van de omreken tabel de score van dit onderdeel.

5a Spelling

Subscores

0 = 5 of meer spellingsfouten

1 = 3 of 4 spellingsfouten

2 = 1 of 2 spellingsfouten

0	1	2
---	---	---

5b Leestekens

0 = 5 of meer leestekensfouten

1 = 3 of 4 leestekensfouten

2 = 1 of 2 leestekensfouten

0	1	2
---	---	---

5c Grammatica

0 = 5 of meer grammaticale fouten

1 = 3 of 4 grammaticale fouten

2 = 1 of 2 grammaticale fouten

0	1	2
---	---	---

Totaal spelling, leestekens en grammatica:

0	1	2	3	4	5	6
---	---	---	---	---	---	---

Omrekentabel van totalen:

0 – 1 = 0 punten

2-3-4 = 1 punt

5-6 = 2 punten

Er is discussie geweest over een grotere foutentolerantie bij 2F. We hebben daar van af gezien, omdat het aantal fouten kan niet los worden gezien van de tekst. Taalgebruikers met lagere niveaus zullen teksten met eenvoudige zinsconstructies en simpelere woordgebruik opleveren. Het aantal fouten is dus relatief!

Fouten in deze categorie die meerdere keren voorkomen elke keer apart fout te tellen.

Bijvoorbeeld:

- Er worden consequent aaneenschrijffouten gemaakt: *verpleeg huis; ochtend medicatie; inbus sleutel*.
- De werkwoordsvervoeging gaat consequent fout: *hij bepaald; zij snoeid, zij heeft gesnoeit, hij heeft bepaalt; evenals fouten tegen samengestelde werkwoorden: gesamenwerkt, geherinnerd*.

Een fout wordt maar 1x aangerekend als het exact dezelfde fout is. Bijvoorbeeld het woord *fabricage* wordt consequent als *fabrikage* geschreven.

4.5 Scoren

De kandidaat scoort **voldoende** als hij gemiddeld voor alle aspecten een voldoende heeft behaald. Dus minimaal 7 x 1 en 6 x 1.

Er zijn twee cruciale voorschriften:

- De inhoudskenmerken zijn cruciaal. Dat wil zeggen dat de kandidaat daarop ten minste 1 punt (voldoende) moet scoren om de opdracht met een voldoende cijfer af te sluiten.
- Een kandidaat mag op maximaal één aspect 0 punten scoren. De 0 punten moeten om tot een voldoende te komen wel gecompenseerd worden met een 2 puntenscore/goed.

Maximale score

De kandidaat kan een maximale score halen als hij op alle beoordelingsaspecten 2 punten scoort.

Voor Schrijven: $7 \times 2 = 14$

Voor Gesprekken voeren: $6 \times 2 = 12$

Voor Spreken: $6 \times 2 = 12$

Het totale aantal behaalde punten wordt vervolgens met behulp van een omreken tabel vertaald naar een cijfer.

4.6 Cijfer bepalen

We werken in de omreken tabel met decimalen, omdat een uitspraak over het niveau van deze taalvaardigheid nooit gebaseerd zal zijn op één toetsopdracht. De cijfers van de verschillende taken moeten gemiddeld worden. Een cijfer zonder decimalen zou de kandidaat te veel afrondingswinst opleveren.

Scoreschaal bij Schrijven 2F en 3F

Aantal punten	Cijfer
14	10
13	9,5
12	9,0
11	8,4
10	7,9
9	7,2
8	6,6
7	6,0
6	5,3
5	4,4
4	3,6
3	2,8
2	2,0
1	1,2
0	1,0

Scoreschaal bij Spreken en Gesprekken voeren 2F en 3F

Aantal punten	Cijfer
12	10
11	9,4
10	8,7
9	8,0
8	7,4
7	6,7
6	6,0
5	5,0
4	4,0
3	3,0
2	2,0
1	1,0
0	0