

WIE SCHRIJFT DIE BLIJFT... AFSTEMMEN

METHODE VOOR EEN BETERE BEOORDELAARSAFSTEMMING IN ÉÉN LESUUR

Karin Echten, Ariane Linnert & Liselore van Ockenburg

Uit onderzoek blijkt dat het moeilijk is om schrijfvaardigheid te 'meten', dat het cijfer voor een schrijfpdracht slechts voor een klein deel wordt bepaald door de schrijfvaardigheid van leerlingen. Er spelen ook andere zaken mee. Veel docenten vinden het moeilijk om schrijfpdrachten van hun leerlingen eerlijk en objectief te beoordelen. Om de betrouwbaarheid van beoordelingen te vergroten, hebben we een stappenplan ontwikkeld.

Wat was het vertrekpunt?

Veel docenten vinden het moeilijk om schrijfpdrachten van hun leerlingen eerlijk en objectief te beoordelen. En uit onderzoek blijkt dat het inderdaad zó moeilijk is om schrijfvaardigheid te 'meten', (Bouwer & Van der Bergh, 2015) dat het cijfer voor een schrijfpdracht slechts voor een klein deel wordt bepaald door de schrijfvaardigheid van leerlingen. Er spelen namelijk een heleboel andere zaken mee. Ten eerste het onderwerp waarover de leerlingen schrijven. Weten ze veel van het onderwerp af? Vinden ze het een interessant onderwerp? Als het antwoord op deze vragen 'ja' is, zullen ze een betere tekst schrijven dan wanneer het antwoord op één of beide vragen 'nee' is. Ten tweede het tekstgenre. Schrijven de leerlingen een informatieve tekst, argumentatieve test, column of brief? Ook dit heeft invloed op hun schrijfpdracht. Ten slotte maakt het uit door wie de tekst wordt beoordeeld, want er bestaan grote verschillen tussen beoordelaars. Verschillende docenten kunnen dezelfde schrijfpdracht soms heel anders beoordelen en niet zelden zitten er hele punten verschil tussen die beoordelingen. Zelfs binnen één beoordelaar kunnen er grote verschillen ontstaan: dezelfde docent kan dezelfde tekst op verschillende tijdstippen heel anders beoordelen. Als hij honger heeft, geeft hij de tekst misschien een zesje, terwijl hij hem na een lekkere maaltijd of een goede nacht slaap een 7,5 zou geven. Dit is problematisch omdat schrijfvaardigheid binnen het schoolvak en dus ook in de schoolexamens Nederlands een belangrijke plek inneemt. Bij alle toetsen, maar zeker bij schoolexamens, moeten leerlingen ervan uit kunnen gaan dat ze op een eerlijke manier getoetst zullen worden en dat hun prestaties op dezelfde manier beoordeeld zullen worden.

Een deel van de hierboven beschreven problemen zou kunnen worden opgelost door iedere tekst te laten beoordelen door verschillende docenten en het gemiddelde van deze oordelen te nemen, maar dat kost erg veel tijd. Daarom zoeken secties de oplossing vaak in het opstellen van uitgebreide rubrics of lijsten met criteria waaraan de tekst moet voldoen. De uiteindelijke beoordeling van de tekst wordt dan bepaald door de som van de (eventueel gewogen) deelscores. Het voordeel van deze methode is dat de verschillen tussen beoordelaars een

stuk kleiner worden. Maar er zijn ook nadelen. Het kost bijvoorbeeld veel tijd om voor iedere schrijfpdracht een goede rubric of criterialijst te maken. Daarnaast moet de beoordelaar een tekst vaak meerdere keren lezen om alle aspecten te kunnen beoordelen. En ten slotte is het de vraag of de eindbeoordeling, uitgedrukt in een totaalaantal punten, een goed beeld geeft van de échte kwaliteit van de tekst. Want laat een complexe vaardigheid als schrijven zich wel opdelen in deelaspecten? En omvat de rubric of criterialijst echt alle relevante aspecten? Wanneer er te veel aspecten 'los' moeten worden beoordeeld, is de beoordeling van de tekst als geheel bijna onmogelijk.

Welke acties zijn er ondernomen?

Om de betrouwbaarheid van beoordelingen te vergroten zónder de werkdruk te verhogen, zou een sectie een gestructureerd overleg kunnen voeren om meer onderlinge overeenstemming te creëren over de manier waarop schrijfpdrachten beoordeeld worden. Hiertoe hebben we een stappenplan (zie de schematische weergave verderop) ontwikkeld dat ruimte biedt om met eigen collega's in gesprek te gaan. Met het stappenplan willen we geen kant-en-klare oplossingen bieden, maar vooral handvatten om in korte tijd met elkaar te bepalen wat de sectie nu écht belangrijk vindt bij het beoordelen van schrijfpdrachten.

Tijdens het doorlopen van de stappen creëert een sectie een eigen beoordelingsschaal. Zo'n schaal bestaat uit verschillende voorbeeldteksten in oplopende kwaliteit en maakt duidelijk wanneer de opdracht voldoende, onvoldoende, goed of excellent is uitgevoerd. Deze aanpak heeft als uitgangspunt dat het beoordelen van schrijfpdrachten niet te reduceren is tot losse scores op afzonderlijke beoordelingsaspecten, (<https://score.hva.nl>) maar dat er juist vanuit het geheel beoordeeld wordt. Een ander voordeel is dat er niet voor iedere schrijftaak een nieuwe beoordelingsschaal ontwikkeld hoeft te worden. Onderzoek heeft namelijk laten zien dat één beoordelingsschaal per genre genoeg houvast biedt om ook teksten van andere schrijftaken net zo betrouwbaar te beoordelen (Koster, Besselink & Seinhorst, 2018). Als je als sectie dus gezamenlijk een beoordelingsschaal creëert waarover je het met elkaar eens bent, kan deze schaal ook helpen bij het beoordelen van andere schrijfpdrachten binnen hetzelfde genre.

Natuurlijk kost ook het creëren van een beoordelingsschaal tijd. Idealiter heeft een sectie Nederlands verschillende middagen per jaar de gelegenheid om los van de waan van de dag te overleggen over doorlopende leerlijnen, inspirerend lesmateriaal en mooie projecten. Helaas is de werkelijkheid vaak anders. Daarom is dit stappenplan bedoeld voor een overleg van 50 minuten. Het materiaal is vrij te downloaden, inclusief alle hieronder genoemde voorbeeldteksten, kwaliteitskaartjes, etc. Zo is het mogelijk om in vier rondes te bepalen wat belangrijk is om leerlingen te leren over schrijfvaardigheid en welke voorbeeldteksten deze kwaliteiten het best weerspiegelen. Eén sectielid (bijvoorbeeld de voorzitter) drukt vooraf het materiaal af, bereidt het gesprek voor, stuurt het gesprek bij als dat nodig is, en houdt de tijd in de gaten.

RONDE	TIJD	INHOUD
1	10 min.	Individueel drie informatieve voorbeeldteksten op volgorde van kwaliteit leggen en argumenten geven voor deze rangschikking.
2	15 min.	Uitkomsten vergelijken (is iedereen het met elkaar eens?), rangschikkingen bespreken en bepalen wat de belangrijkste aspecten zijn voor het beoordelen van tekstkwaliteit.
3	15 min.	Individueel twee nieuwe voorbeeldteksten toevoegen aan de rangschikking, vergelijken van de uitkomsten, eventuele verschillen bediscussiëren en bepalen of de kwaliteitsaspecten uit ronde 2 voldoen of moeten worden bijgesteld/aangevuld.
4	10 min.	Voor de vijf voorbeeldteksten bepalen waar de grens tussen een voldoende en een onvoldoende ligt. Daarna de voorbeeldteksten individueel een cijfer geven. Ten slotte deze cijfers vergelijken en eventuele verschillen bediscussiëren.

Korte inhoud stappenplan

Toelichting per ronde

Ronde 1: Wat vind ik een goede tekst?

Ieder krijgt drie informatieve voorbeeldteksten te lezen, die daarna moeten worden gerangschikt op kwaliteit. Dit gebeurt individueel, dus zonder daarover met elkaar te overleggen. Ieder schrijft kort voor zichzelf op waarom er is gekozen voor deze rangschikking: wat maakt een tekst erg sterk en een andere tekst juist veel minder sterk of ronduit zwak? Op deze manier verkent ieder voor zich welke zaken voor hem of haar impliciet een belangrijke rol spelen bij het bepalen van tekstkwaliteit.



Ronde 2: Kunnen we eens worden over kwaliteit?

De sectie vergelijkt de rangschikkingen en de onderbouwingen. De sectie probeert daarbij vragen te beantwoorden als: Welke criteria komen terug in alle / de meeste onderbouwingen? en Wat vindt de sectie uiteindelijk de belangrijkste criteria voor het beoordelen van tekstkwaliteit? Als een sectie het moeilijk vindt om deze vragen te beantwoorden, kan een envelop worden opengemaakt met daarin kwaliteitskaartjes. In alle methodes worden criteria gebruikt om teksten te beoordelen. Deze criteria hebben we verzameld, gecategoriseerd en afgedrukt op kaartjes. Op de voorkant van een kaartje staat de overkoepelende categorie en op de achterkant staat toegelicht welke aspecten eronder zouden kunnen vallen. Deze kaartjes zijn echter niet bedoeld als aanwijzing wat allemaal moet worden meegenomen in de discussie, maar ze zijn bedoeld om een impuls te geven aan de discussie over wat voor de sectie het allerbelangrijkst is bij het beoordelen van tekstkwaliteit. Binnen 15 minuten bepaalt de sectie de definitieve kwaliteitsrangschikking van de drie voorbeeldteksten en de drie belangrijkste kwaliteitscriteria.



Ronde 3: Kunnen we de afspraken naar tevredenheid toepassen?

Ieder sectielid voegt individueel twee extra voorbeeldteksten toe aan de door de sectie vastgestelde rangschikking van de eerste drie teksten. Bij het beoordelen van de kwaliteit van de nieuwe teksten wordt uitgegaan van de drie belangrijkste kwaliteitscriteria zoals die zijn vastgesteld aan het eind van ronde 2. Als iedereen dit heeft gedaan, kan de plaatsing van de nieuwe teksten worden bediscussieerd. Denkt ieder sectielid hetzelfde over de plaats in de rangschikking of zijn er nog verschillen? Aan het eind van deze ronde stelt de sectie de definitieve rangschikking vast van de vijf teksten vast en maximaal vijf criteria voor het beoordelen van tekstkwaliteit.

Ronde 4: Kunnen we ons gevoel voor kwaliteit uitdrukken in een cijfer?

De sectieleden maken de stap van kwaliteitsrangschikking naar cijfers. Ieder bepaalt individueel tussen welke twee teksten de scheidslijn tussen een voldoende en een onvoldoende ligt. Dan deelt ieder de uitkomsten van deze oefening en bepalen de sectieleden in hoeverre ze het eens zijn. Vervolgens geeft ieder individueel een cijfer voor iedere tekst en ook deze uitkomsten worden vergeleken. Als de cijfers ver uiteenlopen, probeert de sectie uit te vinden wat er aan de cijfers ten grondslag ligt. Beoordeelt iedereen echt op basis van de vastgestelde beoordelingscriteria? Gebruikt iedereen de volledige breedte van de beoordelingsschaal?

Als er meer tijd beschikbaar is, biedt het stappenplan ook vervolgstappen om te kijken hoe goed het lesmateriaal voor schrijfvaardigheid aansluit bij wat de sectie belangrijk vindt en in hoeverre verschillende tekstsoorten (uiteenzettend, overtuigend) anders beoordeeld worden, maar het allerbelangrijkst is het natuurlijk dat sectiegenoten steeds met elkaar in gesprek blijven over hun beoordelingspraktijk, bij voorkeur aan de hand van schrijfopdrachten uit de eigen lespraktijk. Secties kunnen ook aan het begin van een schooljaar afspreken van welke opdracht er een paar uitwerkingen door een paar collega's nagekeken en bediscussieerd worden voordat iedereen het werk van de hele klas nakijkt. Zo kunnen nieuwe collega's steeds meegenomen worden in de gebruiken van één sectie. In kleine groepjes schrijfopdrachten ontwerpen helpt een sectie ook één lijn te ontwerpen.

Wat zijn de bevindingen?

Tijdens een bijeenkomst van de werkgroep Schrijven in de Schoolexamens vroegen we de deelnemers om ons stappenplan zelfstandig uit te voeren om zo te testen of de instructies en opdrachten helder waren. We deelden de deelnemers willekeurig in 'secties' in en daarna gingen ze aan de slag met de opdrachten. Wij liepen intussen rond om overal mee te luisteren en te kijken hoe het stappenplan in de praktijk functioneerde. En wat bleek? De eerste test werd positief ontvangen, maar verliep zeker niet vlekkeloos.

De deelnemers concludeerden na afloop van de testronde dat het stappenplan een voldoende structuur aan de discussie over tekstkwaliteit beoordelen geeft en kan leiden tot een bruikbare reeks ankerteksten. Doordat deelnemers steeds eerst hun eigen gedachten moeten formuleren en pas daarna met de sectie delen om de gemeenschappelijke elementen te vinden, ontstaat een echt gesprek. Dat werd als zeer positief ervaren.

Maar natuurlijk ontdekten we ook wat problemen. Bijvoorbeeld met de drie voorbeeldteksten die we in ronde 1 gebruikten voor de eerste rangschikcoefening. Deze teksten op basis van bronnen waren geschreven door leerlingen uit vwo 4, 5 en 6, in het kader van het LIFT-onderzoek: een landelijk onderzoek uit 2016-2017 naar het schrijfvaardigheidsniveau van leerlingen uit klas 4, 5 en 6-vwo (<https://liftwritingresearch.wpcomstaging.com/>).

De teksten zijn door meerdere onafhankelijke beoordelaars beoordeeld en van iedere tekst is bekend wat de kwaliteit is ten opzichte van de gemiddelde kwaliteit van de teksten uit dit onderzoek. De kwaliteit van de teksten lag ver uiteen en was geclassificeerd als *zwak*, *gemiddeld* en *sterk*. Toch ontstond er in sommige 'secties' een discussie die veel meer tijd in beslag nam dan we ervoor hadden begroot. Dat kwam onder andere doordat in de *sterke* tekst wat argumentatieve elementen leken te zitten. De tekst was, mede daardoor, prettig leesbaar vanwege de persoonlijke stijl en scoorde dus hoog op dit kwaliteitsaspect. Tegelijkertijd werd dezelfde tekst door andere sectieleden afgekeurd omdat hij niet volledig aan de opdracht voldeed. Daarin stond immers dat er een informatieve tekst geschreven moest worden, terwijl deze tekst argumentatieve elementen bevatte. Andere discussies over tekstkwaliteit werden veroorzaakt doordat twee leden van één sectie, ook na discussie, zowel de *zwakke* als de *sterke* tekst beter vonden dan de *gemiddelde* tekst. Sommige secties vonden het door deze dilemma's zo moeilijk om tot een consensus te komen, dat stap 2 van het plan niet kon worden afgerond en ze niet meer aan stap 3 en 4 toe kwamen. Ook vonden verschillende secties de kwaliteit van de *sterke* tekst niet goed hoog genoeg.

Ook leverden de kwaliteitskaartjes uit ronde 2 problemen op. Sommige secties ontdekten pas na afloop van de test dat ze kaartjes tot hun beschikking hadden. Andere secties lieten weten dat ze de kaartjes graag al aan het begin van ronde 1 zouden hebben ontvangen, om zo een compleet overzicht te creëren van aspecten die mogelijk van belang zijn bij het beoordelen van tekstkwaliteit. Toch hebben we er bewust voor gekozen dit niet te doen. Wij willen niet dat kaartjes leidend zijn, maar dat secties vooral bepalen wat zij, als groep, zelf het belangrijkste vinden. Andere problemen met de kaartjes waren dat er overlap leek te bestaan tussen verschillende categorieën zoals 'samenhang' en 'structuur' en dat niet alle secties hadden gezien dat de categorieën op de achterzijde van de kaartjes waren uitgewerkt in verschillende concrete criteria.

Ten slotte vonden sommige deelnemers de vragen in ronde 1 en 3 erg sturend, hoewel de vragen achteraf ook werden beoordeeld als eenvoudig en helpend. Het bleek moeilijk een balans te vinden tussen heldere, concrete vragen die in een zeer kort tijdsbestek kunnen worden beantwoord en het waarborgen van voldoende vrijheid om een echt gesprek over beoordelen en tekstkwaliteit op gang te laten komen. Gelukkig ontvingen we bruikbare tips om de vraagstelling aan te passen.

Conclusies

Hoewel de test van het stappenplan dus enkele problemen aan het licht bracht, was de algemene indruk dat het stappenplan secties kan helpen om meer op één lijn te komen bij het beoordelen van teksten. Het plan gaf voldoende ruimte voor een open gesprek en tegelijkertijd voldoende houvast voor een gestructureerde aanpak. Het stappenplan was in dat opzicht dus geslaagd. Daarom besloten we het plan op hoofdlijnen te behouden en de geconstateerde problemen aan te pakken.

We voegden een inhoudsopgave toe, voorafgaand aan het stappenplan, waarin duidelijk stond beschreven welke materialen de secties in welke rondes tot hun beschikking hebben. Ook de kwaliteitskaartjes worden genoemd in de inhoudsopgave, zodat secties deze niet over het hoofd zien. Van de kaartjes die te veel overlap vertoonden, hebben we de categorie en criteria gespecificeerd. Ten slotte hebben we de formulering van een aantal vragen aangepast, waardoor ze minder sturend waren. Waar er in de eerste versie van het plan in ronde 1 bijvoor-

beeld werd gesuggereerd dat er teksten van drie verschillende kwaliteiten moesten worden gerangschikt en er in ronde 3 twee tekst ertussen moesten worden geplaatst, hebben we deze opdrachten in de tweede versie neutraler geformuleerd.

De problemen die zich voordeden met betrekking tot het materiaal, konden we dankzij de tips van de deelnemers oplossen. We vervingen de *sterke* voorbeeldtekst met betogende elementen door een andere *sterke*, meer neutraal geschreven voorbeeldtekst uit het LIFT-onderzoek. Een tip van de deelnemers van om na ronde 2 de sectie hun rangschikking te laten vergelijken met de rangschikking die uit het LIFT-onderzoek was gekomen en daarbij aan te geven dat deze rangschikking is voortgekomen uit een betrouwbare peiling van het landelijke schrijfvaardigheidsniveau. Bij deze rangschikking moet bij iedere voorbeeldtekst een korte onderbouwing gegeven worden van de kwaliteit. Hetzelfde geldt voor de extra voorbeeldteksten die moeten worden toegevoegd in ronde drie. Tegelijkertijd staat deze toevoeging op gespannen voet met het doel van het stappenplan, namelijk afstemming vinden binnen je sectie, waarbij de volgorde van de teksten niet het belangrijkste is.

Om ervoor te zorgen dat het stappenplan in de beschikbare tijd kan worden afgerond, helpt het als één sectielid optreedt als gespreksleider en tijdbewaker. Om het stappenplan daadwerkelijk in 50 minuten af te kunnen ronden met een maximaal rendement is het belangrijk dat alle sectieleden een evenredige inbreng hebben en dat de leden niet te lang in discussies blijven hangen. Vooraf wordt dus bij voorkeur bepaald wie gespreksleider en tijdwaarnemer is.

Verdere plannen

Schrijven is één van de belangrijkste vaardigheden binnen het schoolvak Nederlands, maar aan het beoordelen van schrijfvaardigheid zitten flink wat haken en ogen. Dit project heeft laten zien dat een stappenplan een succesvolle manier kan zijn om als sectie met elkaar in gesprek te gaan over het beoordelen van tekstkwaliteit. Om secties te faciliteren die hierin geïnteresseerd zijn, maken we het aangepaste stappenplan op korte termijn digitaal beschikbaar. De bedoeling is dat secties er volledig zelfstandig mee aan de slag kunnen gaan om zo in 50 minuten meer overeenstemming te bereiken over het beoordelen van tekstkwaliteit.

Bij dit artikel hoort een bijlage, te vinden via <https://ltn.publicatieserver.nl/publication/3/150>

LITERATUUR

- Bouwer, R., & Van den Bergh, H. (2015). Toetsen van schrijfvaardigheid: Hoeveel beoordelaars, hoeveel taken? *Levende Talen Tijdschrift*, 16(3), 3-12.
- Koster, M., Besselink, E., & Seinhorst, E. (2018). Goed gedaan, maar kan nog beter...; Het ontwikkelingsgericht en betrouwbaar beoordelen van leerlingteksten. *Tijdschrift voor Lerarenopleiders*, 39(3), 83-94.
- Onderzoeksproject LIFT. *Feedback op het schrijven van syntheseseteksten*. <https://liftwritingresearch.wpcomstaging.com/>.

Het project Schoolexamen Schrijfvaardigheid (Ses) is een initiatief van LT sectie Nederlands en wordt uitgevoerd in samenwerking met de VU en Fontys. Het project SES wordt afgerond met een studiedag en een handleiding Schoolexamen Schrijfvaardigheid voor docenten, die voor iedereen beschikbaar gesteld wordt via SLO.

