

De beoordeling van schrijfvaardigheid

Een onderzoek naar de mogelijke implementatie van schaalmodellen in het
basisonderwijs

ERIK PHILIPPENS
JUNI 2013

Abstract

Het beoordelen van schrijfvaardigheid in het basisonderwijs blijkt een niet al te gemakkelijke taak: naast dat het begrip schrijfvaardigheid problemen oproept, is men het ook niet altijd eens over de wijze van beoordelen. Zo hebben de beoordelingsmethoden allen hun nadelen. Een van deze beoordelingsmethodes maakt gebruik van voorbeeldteksten op een schaal: de schaalbeoordelingsmethode. Uit onderzoek is gebleken dat de schaalbeoordelingsmethode voor een hoge beoordelaarsovereenstemming zorgt. Deze methode wordt echter niet implementeerbaar beschouwd omdat deze een erg hoge arbeidsintensiviteit met zich mee brengt voor de docent.

Dit onderzoek heeft getoetst of een schaalmodel adequaat kan fungeren bij het beoordelen van teksten, geschreven door kinderen in het basisonderwijs, die van onderwerp en/of teksthandeling verschillen. De resultaten van dit onderzoek suggereren dat een schaalmodel gebruikt kan worden ongeacht het onderwerp of teksthandeling.

Inleiding

Probleemstelling

Het leren van schrijfvaardigheid aan onze kinderen lijkt een vanzelfsprekend fenomeen. Er zitten echter meer haken en ogen aan dan gedacht. Voor het leren en doceren van schrijfvaardigheid is het uiteraard als eerste nodig voor de docenten om in staat te zijn de schrijfvaardigheid te toetsen. Dit doen we aan de hand van het beoordelen van geschreven teksten. Van deze teksten wordt de tekstkwaliteit beoordeeld om daarmee een schrijfvaardigheid aan een persoon toe te kennen. De kwestie die als eerst voor ons ligt is wat schrijfvaardigheid precies is. Van den Bergh en Meuffels laten zien dat schrijfvaardigheid een moeilijk punt is. Zij noemen vaardigheden ‘*hypothetische constructies*: wanneer wij iemand een grote schrijfvaardigheid toeschrijven, dan veronderstellen wij bij die persoon een dispositie (een attribuut) aanwezig die aan het geobserveerde gedrag ten grondslag ligt en die zich onder vergelijkbare omstandigheden ook op een andere tijd en plaats zou manifesteren’ (Van den Bergh en Meuffels, 2000, p. 122). Bij het hypothetische begrip schrijfvaardigheid wordt verwezen naar het voldoen aan bepaalde criteria die hieraan worden aangewezen. Wanneer iemand schrijfvaardig is, is deze persoon in staat om in iedere situatie aan de huidige schoolse maatstaven en criteria te voldoen.

Deze criteria zijn onder andere opgenomen in Referentiekader taal en rekenen, dat in opdracht van het ministerie van onderwijs, cultuur en wetenschap (OCW) is geschreven. Dit referentiekader beschrijft de basiskennis en –vaardigheden waaraan Nederlandse leerlingen moeten voldoen wat betreft taal en rekenen. Het zorgt daarnaast voor een onderscheid tussen een fundamenteel niveau en eens streefniveau (Meijerink, Letschert, Rijlaarsdam, van den Bergh & van Streun, 2009). Aan de hand van deze criteria kan het mogelijk zijn om de schrijfvaardigheid van scholieren te beoordelen.

Hoewel deze criteria om de schrijfvaardigheid te beoordelen goed te gebruiken zijn, blijven er problemen aan plakken. De criteria focussen zich namelijk op de resultaten van het schrijven en niet op bijvoorbeeld het proces. Een van de problemen noemen van den Bergh en Meuffels dan ook dat een vaardigheid als schrijven niet direct observeer- en dus meetbaar is, maar moet worden afgeleid uit iemands handelingen, gedragingen en de resultaten daarvan (Van den Bergh & Meuffels, 2000). De vraag is natuurlijk of op dat moment – wanneer alleen de resultaten getoetst worden en niet het proces - schrijfvaardigheid daadwerkelijk gemeten wordt.

Deze uitspraak legt het probleem rondom schrijfvaardigheid bloot: een tweeluik bestaande uit een meetprobleem en beoordelingsprobleem. Beoordelaars blijken het onderling niet eens te zijn over waar een tekst aan moet voldoen en wanneer schrijfvaardigheid gemeten wordt. Overigens komt hierdoor de betrouwbaarheid tussen beoordelaars in het gedrang: hoe kunnen we iemand schrijfvaardigheid toewijzen wanneer een andere beoordelaar een andere notie heeft van dit begrip? De beoordelaarsbetrouwbaarheid wordt gekarakteriseerd door de consistentie van de beoordelaar (Van den Bergh en Meuffels, 2000). De betrouwbaarheid tussen beoordelaars, of beoordelaarsovereenstemming kenmerkt zich door dezelfde of overeenkomstige scores tussen de beoordelaars per tekst.

Van den Bergh en Meuffels (2000) wijten een lage beoordelaarsbetrouwbaarheid en beoordelaarsovereenstemming aan de onduidelijkheid in kenmerken van tekstkwaliteit – zo is

het niet altijd even duidelijk waar een tekst van hoge kwaliteit aan moet voldoen.

De vraag is dan ook of bij een goede kenmerking van de tekstkwaliteit de stabiliteit voldoende zou zijn. Wanneer dit het geval zou zijn, is de beoordelaarsbetrouwbaarheid hoog wanneer er een goede methode wordt ontwikkeld. De huidige meestgebruikte methoden zullen verderop in dit onderzoek aan het bod komen.

Naast de problematiek die het begrip schrijfvaardigheid en diens beoordeling op zich af roept, blijft het probleem bestaan dat de schrijfvaardigheid van leerlingen erg laag is. De basis van de te lage schrijfvaardigheid ligt volgens McColly (1970) in de educatie. Het is volgens McColly echter niet te wijten aan de studenten. Het probleem lijkt voornamelijk te liggen in de onenigheid over wat goed schrijven is en de onmogelijkheid voor leraren om objectief te blijven bij het beoordelen van hun studenten (McColly, 1970).

Uit zowel het beoordelaarsprobleem als het meetprobleem komt naar voren dat de onenigheid over het construct schrijfvaardigheid het beoordelen van teksten en daarmee het determineren van de schrijfvaardigheid vermoeilijkt. Omdat schrijfvaardigheid uit de resultaten afgelezen moet worden is het van groot belang om tot een gelijke beoordeling te komen om discrepantie in het gegeven onderwijs te voorkomen. Om de schrijfvaardigheid te verhogen wordt de focus gelegd op de beoordelingsmethode.

Dit onderzoek zal zich als eerste bezig houden met een uiteenzetting van de meest voorkomende beoordelingsmethoden: het analytisch beoordelingsmodel; het holistisch beoordelingsmodel en het schaalbeoordelingsmodel.

Analytisch beoordelingsmodel

Om een uniforme beoordeling in de educatie te krijgen zijn er verschillende evaluatie methoden om teksten te beoordelen. Het Europees referentiekader talen stelt dat het analytisch beoordelingsmodel al lange tijd gebruikt wordt voor het beoordelen van schrijfvaardigheid (Liemberg en Meijer, 2004). Dit model werkt door fouten aan te strepen in een tekst en door middel van puntenaftrek een cijfer te bepalen. Aan de hand van categorieën wordt de beoordeling opgedeeld en kan er nadruk gegeven worden op, voor de onderzoeker, belangrijke elementen.

Deze methode lijkt zeer goed te werken doordat belangrijke elementen uit een tekst kunnen worden geselecteerd en beoordeeld op aanwezigheid en correctheid. Er kleven echter problemen aan. Zo is niet altijd even duidelijk aan welke criteria een beoordelaar zich moet houden bij het nakijken. Er worden categorieën gemaakt waaraan de beoordelaar zich houdt tijdens het beoordelen. Voorbeelden hiervan zijn bijvoorbeeld spelling en argumentering. Aan deze categorieën wordt echter niet altijd dezelfde waarde gehecht – soms is argumentering belangrijker dan spelling en andersom - en soms worden zelfs categorieën weggelaten (Wesdorp, 1981). Dit geeft ernstig problematische situaties waarin het niet duidelijk is of aan de hand van de gekozen categorieën dezelfde notie van schrijfvaardigheid wordt gemeten als bij een andere indeling. Wanneer de categorieën tussen beoordelingen namelijk van elkaar verschillen dan is het niet duidelijk of beide beoordelingen dezelfde notie van schrijfvaardigheid meten. Denk aan het weglaten of de toevoeging van een categorie als interpunctie. Wanneer deze categorie wordt meegenomen in de beoordeling wordt er een andere beoordeling mogelijk dan wanneer dit buiten beschouwing wordt gelaten.

Daarnaast is een analytische beoordeling erg tijdrovend. Door de aanwezigheid van

verschillende categorieën is het nodig om een tekst vaker met verschillende invalshoeken te lezen; een docent kan immers niet tegelijk de spelling beoordelen als de argumentatie. Hierdoor gaat de arbeidsintensiviteit sterk omhoog. Daarnaast geeft Wesdorp aan dat het mogelijk is dat de beoordelaar zich laat beïnvloeden door tekstkenmerken die eerder in de analyse van de tekst naar voren zijn gekomen (Wesdorp, 1981). Doordat de beoordelaar een tekst vaker leest kan dit niet uitgesloten worden. Tot slot vermeldt Wesdorp (1981) dat er onduidelijkheid bestaat over of de beoordelaarsovereenstemming bij deze methode hoog uitvalt. Het is dus niet vanzelfsprekend dat de analytische beoordelingsmethode zorgt voor een goede beoordelaarsovereenstemming.

Door de beperkingen die deze methode geeft, lijkt het niet vanzelfsprekend om deze veel te gebruiken in het basisonderwijs. Hierdoor verdient deze methode niet de voorkeur.

Holistisch beoordelingsmodel

Een andere manier om schrijfvaardigheid in teksten te beoordelen is het holistische beoordelingsmodel. Bij dit model wordt de tekst als geheel beoordeeld. Het model is gebaseerd op de theorie dat iedere tekst groter is dan de onderdelen bij elkaar (Mullis, 1984). De verschillen tussen deze methode en het analytische beoordelingsmodel zijn dat het analytische model meer tijd in beslag neemt en dieper op tekstuele kenmerken in gaat maar er niet in slaagt de tekst als geheel te beoordelen, waar het holistische model met het omgekeerde probleem kampt.

Door te beoordelen aan de hand van het holistische model kunnen beoordelaars weliswaar snel te werk gaan en hebben zij geen training nodig, maar ontwikkelen zij ook ieder hun eigen criteria voor goede en slechte schrijfvaardigheid (Mullis, 1984). Hierdoor komt de betrouwbaarheid tussen en binnen beoordelaars sterk in het gedrang.

Daarnaast geeft het holistische beoordelingsmodel weinig informatie over de schrijfvaardigheid. Een oordeel of cijfer voor een gehele tekst geeft niet aan waar de sterke en zwakke punten liggen. Zelfs wanneer de beoordelaar zijn eigen of vooropgelegde criteria hanteert zullen deze bij de beoordeling niet naar voren komen. In het algemeen hebben andere methoden de voorkeur waarin criteria op zijn minst enigszins gespecificeerd worden.

Schaalbeoordelingsmodel

Een derde methode maakt gebruik van ankerteksten die een plaats innemen in een schaalmodel. Zo zijn er in dit voorbeeld vijf ankerteksten waar de beoordelaar de te beoordelen tekst mee kan vergelijken. De beoordelaar kan deze tekst dan een score toewijzen die correspondeert met de ankertekst waar deze het meest op lijkt. De ankerteksten nemen een plaats in op een schaal en staan daarmee voor een x-aantal punten. De te beoordelen tekst krijgt dus de punten aan de hand van zijn positie op de schaal. Door de teksten die beoordeeld worden op de voor de beoordelaar juiste plaats in het model te plaatsen is het mogelijk de corresponderende punten hier aan te geven. Het schaalmodel wordt onderbouwd met uitleg over waarom de ankerteksten op hun plaats staan en dient als ondersteuning bij het beoordelen. Door gebruik te maken van voorbeelden die hun plaats innemen in een schaalmodel is het gemakkelijk om zonder al te veel training consistente beoordelingen te geven aan kleine teksten. De beoordeling via een schaalmodel werkt als volgt: bij iedere tekst die de beoordelaar voor zich krijgt, gaat deze na waar de tekst zich op de schaal bevindt. Hij

of zij zoekt naar overeenkomsten tussen de te lezen tekst en de ankertekst en laat zich daarbij sturen door de uitleg die bij de schaal gegeven wordt. Aan de hand van de plaats die de beoordelaar aan de tekst toewijst krijgt de tekst de overeenstemmende punten. Wat betreft de schaalmodellen en analytische aandachtspunten, geldt dat de nadruk gelegd dient te worden op argumentatiestructuur en tekstcohesie.

In een artikel van Kuiken & Vedder (2012) waarin beoordelaars van schrijfvaardigheid zelf factoren noemen die primair betrekking hebben op de schrijfvaardigheid, komt namenlijk naar voren dat ‘duidelijk geformuleerde en met elkaar verbonden argumenten en sprekende voorbeelden, het vermijden van onnodige herhalingen, het gebruik van verbindigswaarden en signaalwoorden, het belang van een goede en heldere inleiding en conclusie’ de belangrijkste factoren zijn voor communicatieve adequaatheid (Kuiken & Vedder, 2012 p. 97). Dit gaat hand in hand met de gedachte dat structuur van de tekst en argumentatiestructuur belangrijker zijn dan spelling en interpunctie. Het schaalmodel en diens uitleg dienen zich dus aan deze gedachte te houden.

Het gebruik van het schaalbeoordelingsmodel zit qua tijd en arbeidsintensiviteit tussen het analytische en het holistische beoordelingsmodel in. Het meeste werk zit in de voorbereiding om de benodigde ankerteksten uit te kiezen. Twee belangrijke voordelen van schaalmodellen worden door Pollmann et al. (2012) gegeven: het normaal – in zowel het analytische als het holistische beoordelingsmodel - voorkomende sequentie-effect, waarin de volgorde van invloed is op de beoordeling, komt bij het gebruik van schaalmodellen niet voor. Iedere tekst wordt namelijk opnieuw voorgelegd aan de ankerteksten. Daarnaast komt ook normverschuiving niet voor. Bij de normverschuiving gaat de beoordelaar mee in de kwaliteit van de tekst en past zich daar aan. Omdat de teksten niet onderling, maar aan de hand van ankerteksten beoordeeld worden, is de beoordelaar niet meer geneigd om zich aan het niveau aan te passen.

De resultaten van het onderzoek van Pollmann et. al. (2012) laten zien dat de overeenstemming en betrouwbaarheid van de oordelen van de beoordelaars met het gebruik van schaalmodellen hoog waren.

De vergelijking van deze drie beoordelingsmethoden geeft aan dat de schaalbeoordelingsmethode de voorkeur krijgt boven de holistische en analytische beoordelingsmethode. Een volgende stap hier is om naar de implementatie van de schaalbeoordelingsmethode te kijken.

Het gebruik van schaalmodellen zou wellicht op grote schaal kunnen worden toegepast in de educatie in Nederland. Een erg belangrijk punt van kritiek hiervan is echter of de schaalmodellen gebruikt kunnen worden buiten hun oorspronkelijke onderwerp en zelfs teksthandeling. Er is, zoals in het onderzoek van Pollmann et. al. (2012) aangetoond dat schaalmodellen werken en consistent te gebruiken zijn bij het beoordelen van teksten. Het draait hier slechts om schaalmodellen die specifiek voor een tekst of opdracht gemaakt zijn. De huidige notie is dat voor ieder tekstkenmerk of opdracht telkens een nieuwe schaalbeoordeling geconstrueerd dient te worden (van den Bergh en Meuffels, 2000). Het is echter ondoenlijk om voor iedere afzonderlijke schrijfopdracht een nieuw schaalmodel te creëren. Het zou daarom van groot belang zijn als de schaalmodellen niet onderwerp en/of teksthandeling afhankelijk zouden zijn. Om te onderzoeken of dit wel zo is, zal allereerst

gekeken worden naar de invloed van het onderwerp en de eigenschappen en verschillen tussen teksthandelingen.

Teksthandelingen

Het onderwerp van een tekst heeft bij het beoordelen een kleinere rol dan het tekst type. Bij het beoordelen van een tekst wordt namelijk de structuur bekeken op aanwezigheid en correctheid van belangrijke elementen zoals argumentatie. Pander Maat (2002) beschrijft dat voor een goede tekstbeoordeling grondige kennis van zowel het genre als communicatieve doel van de tekst nodig is. Bij een beoordeling moet de nadruk dus op deze twee constructen gelegd worden.

Volgens Pander Maat (2002) zijn genres en genrenamen – zoals de brief of het betoog - opgebouwd uit een of meerdere teksthandelingen. Dit construct verdient wat extra uitleg: volgens Pander Maat (2002) hebben teksten en passages van teksten een bepaalde strekking waardoor we van teksthandelingen spreken. Teksthandelingen hebben de vorm van een werkwoord en geven in feite aan wat het communicatieve doel van de tekst is – een tekst kan gemaakt worden met als doel een verzoek in te dienen, waardoor deze tekst de teksthandeling verzoeken volgt. Het construct teksthandelingen maakt daarom een perfecte aanvulling in de discourse. Het belangrijkste bij het beoordelen zit hem zoals eerder genoemd in het communicatieve doel en het genre, die beiden in de teksthandelingen te vinden zijn.

Het verschil tussen verschillende teksthandelingen zit in de structuur van een tekst. Om dit uit te leggen zal de structuur van een tekst naar aanleiding van verschillende teksthandelingen besproken worden.

Een tekst is in feite opgebouwd uit allerlei uitingen zoals bijvoorbeeld ‘het water is koud’. Deze uitingen vormen gezamenlijk tekstpassages die op hun beurt onderdeel zijn van de tekst. Al deze uitingen hebben een bepaald doel. Zo heeft het voorbeeld ‘het water is koud’ een informatief of wellicht een expressief doel. Dit hoeft echter niet te betekenen dat een tekst met een bepaald doel (verzoeken) alleen maar verzoekende uitingen in zich heeft. Een tekst heeft om deze reden een bepaald thema of een overkoepelende teksthandeling. Het construct thema wordt gebruikt om het onderwerp van een informerende tekst aan te geven en het construct hoofdhandeling wordt gebruikt om de overkoepelende teksthandeling weer te geven (Pander Maat, 2002).

Teksthandelingen zijn dus ‘op twee manieren in de tekststructuur aanwezig: als globale handeling, gevormd door een reeks samengevoegde uitingen, of als hoofdhandeling, ondersteund door de overige uitingen in de tekst’ (Pander Maat, 2002, p. 265). Dit houdt in dat teksthandelingen gebruikt kunnen worden als ondersteuning van een bepaald thema (bij bijvoorbeeld een informerende tekst) of als hoofdhandeling – als de belangrijkste uiting in bijvoorbeeld een verzoekende tekst. Hierdoor zien we dat wanneer een teksthandeling verschilt, de structuur van een tekst ook drastisch verandert. Een overtuigende (directieve) tekst met de teksthandeling verzoeken zal immers gebruik maken van uitingen (in dit geval argumenten) die leiden tot de hoofdhandeling van de tekst – het verzoek. Daarentegen zal een informerende tekst juist uitleg geven over een bepaald thema. Een teksthandeling die dit doel dient is bijvoorbeeld beschrijven. De wetenschap of we te maken hebben met een themagecentreerde tekst of een tekst met een hoofdhandeling, kan ons dus helpen de structuur van een tekst te ontleden. De aanwezigheid van een thema of hoofdhandeling in een tekst

geeft een andere structuur van de tekst aan (Pander Maat, 2002).

Pander Maat laat ons dus zien dat de structuren van teksten kunnen verschillen op basis van het communicatieve doel. Deze informatie is erg belangrijk wanneer gewerkt wordt met schaalmodellen. Wanneer het schaalmodel gebaseerd is op de ene structuur en gebruikt wordt bij het beoordelen van een andere structuur kan de beoordeling namelijk sterk belemmert worden. Als een schaalmodel bijvoorbeeld wordt gebaseerd op een directieve tekst die een verzoek vervult dan zal bij het analyseren van deze tekst ingegaan worden op de centrale uiting en de eventuele argumenten. Kijken we echter met dit schaalmodel dat gebaseerd is op een directieve tekst naar een beschrijvende tekst, beoordelen de beoordelaars een themagecentreerde tekst met een schaalmodel op basis van een centrale uiting. Dit houdt in dat er discrepantie ontstaat tussen de beoogde wijze van beoordelen en de structuur van de te lezen tekst. Zo kan er gezocht worden naar argumenten in een tekst waar geen argumenten in voorkomen. Deze discrepantie kan uiteraard zorgen voor uiteenlopende beoordelingen en daardoor een mindere beoordelaarsovereenstemming.

Hoewel er verschil gevonden kan worden tussen teksthandelingen, spreekt Pander Maat (2002) ook van overeenkomsten: zo kunnen bijvoorbeeld instruerende passages ook onder een directieve teksthandeling vallen. Hiermee wordt bedoeld dat de structuur van instruerende passages kan lijken op die van directieve passages. Dit komt doordat het, net als bij directieve teksten, bij instruerende teksten om de handeling van de lezer gaat. Bij zowel directieve als instruerende draait de tekst er om de lezer aan het werk te zetten. Wanneer de structuur dus op elkaar gaat lijken kan het mogelijk zijn dat een instruerende tekst gemakkelijker te beoordelen valt aan de hand van een schaalmodel van een directieve tekst – dan bijvoorbeeld een informatieve tekst - omdat de instruerende en directieve teksten qua teksthandeling dicht bij elkaar liggen.

Deze verkenning van de verschillen tussen teksthandelingen heeft geleid tot een beter begrip van tekststructuren en inzicht in de mogelijkheid van eventuele invloed bij tekstbeoordeling – vooral wanneer er bij de beoordeling discrepantie is tussen de te beoordelen tekst en het gebruikte schaalmodel. Het is echter nog niet duidelijk of, hoe en in hoeverre deze invloed zich gaat uiten.

Vraagstelling:

Het is duidelijk geworden dat de schaalbeoordelingsmethode de voorkeur geniet boven andere beoordelingsmethoden, maar dat nog niet duidelijk is of deze methode ook op grote schaal gebruikt kan worden. Kijkend naar het probleem dat hierdoor voor ons ligt rijst ons de vraag of het mogelijk is alle teksten van basisschoolleerlingen te beoordelen aan de hand van schalen die gebruik maken van ankerteksten en gemaakt zijn voor een bepaalde schrijfo opdracht. Hierbij is het belangrijk te kijken of deze beoordelingschalen ook voor andere onderwerpen en teksthandelingen te gebruiken zijn alvorens deze geïmplementeerd kunnen worden. Hieruit volgt de volgende vraag: Heeft het onderwerp en/of teksthandeling bij het beoordelen van teksten van basisschoolleerlingen door taalstudenten aan de hand van een beoordelingsschaal invloed op de beoordelaarsovereenstemming?

Er zal na worden gegaan wat de beoordelaarsovereenstemming is wanneer de opdrachten op verschillende wijze congrueren. De congruentie is hoger of lager wanneer het schaalmodel en diens ankerteksten respectievelijk dichterbij of verder af liggen van de te beoordelen

teksten. Wanneer de schaal wordt gebruikt om teksten met dezelfde teksthandeling en onderwerp na te kijken spreken we van een hoge overeenkomst, ofwel een hoge congruentie. De schaal en de teksten congrueren gemiddeld wanneer alleen het onderwerp verschilt en wanneer zowel onderwerp als teksthandeling verschillen dan spreken we van een lage congruentie. Hierdoor kan de volgende centrale vraagstelling geformuleerd worden:

In welke mate is de congruentie tussen de beoordelingschaal en de schrijfp opdrachten van invloed op de beoordelaarsovereenstemming?

In dit onderzoek wordt de focus vooral gelegd op de beoordelaarsovereenstemming bij de opdrachten met gemiddelde en lage congruentie¹. Er zal echter ook ter controle gekeken worden naar de beoordelaarsovereenstemming bij hoge congruentie.

Hypothesen

Verwacht wordt dat de beoordelaarsovereenstemming bij zowel een hoge als gemiddelde overeenkomst hoog zal zijn. De teksten zijn immers op het gebied van teksthandeling hetzelfde en verschillen alleen per onderwerp. Naar verwachting zal de schaal, ongeacht of deze van onderwerp verschilt van de tekst, voldoende houvast geven bij de beoordeling. De beoordelaarsovereenstemming bij een lage congruentie zal naar verwachting lager uitvallen dan zowel de hoge als de gemiddelde congruentie.

De beoordelaarsovereenstemming van de samengenomen hoge en gemiddelde congruentie zal zoals verwacht hoog zijn en hoger uitvallen dan de beoordelaarsovereenstemming bij een lage congruentie.

Methode

Materiaal

Voor dit onderzoek hebben wij de teksten van drie reeds gemaakte opdrachten gebruikt. De drie teksten verschillen van onderwerp en één heeft een andere teksthandeling. De opdrachten die vooraf zijn gegaan aan deze teksten zijn gegeven aan basisschoolleerlingen van groep 7 en 8, met een instructie. De drie opdrachten – vanaf nu te noemen Smurfen, Smikkel en Like – bevinden zich in respectievelijk bijlage 1A, 1B en 1C. De opdrachten zullen kort worden uitgelegd.

Smurfenopdracht

De leerlingen is een uitleg gegeven over de situatie waarin zij zich bevonden: de Smurfen die door een actie bij iedere 25 euro aan boodschappen te verkrijgen waren, waren bij de laatste keer boodschappen doen op. Er dient een brief gezonden te worden met als doel de smurfen die de kinderen nog te goed hebben, te ontvangen. Uit deze opdracht vloeiden directieve teksten voort met de teksthandeling verzoeken. De opdracht luidde als volgt:

Schrijf een briefje dat je meestuurt met de kassabon van 15 april. Leg uit dat bij SuperCoop

¹ Vanaf nu dient ‘congruentie’ als ‘congruentie tussen de beoordelingschaal en de schrijfp opdracht’ gelezen te worden.

om de hoek de Smurfen op waren.

Overtuig de firma SUPERCOOP ervan dat je vóór de sluitingsdatum van de actie de zakjes met Smurfen wilt ontvangen. Schrijf daarna de envelop.

Smikkelopdracht

De Smikkelopdracht gaf voor de teksten een zelfde teksthandeling, maar verschilde in onderwerp. Bij de Smikkelopdracht werd de leerlingen voorgelegd dat zij – net als bij de smurfenopdracht – met een spaaractie meededen waar punten bij elkaar moesten worden gespaard. De leerling staat voor het probleem dat het met slechts 8 punten niet aan de CD komt die voor 10 punten gehaald kan worden. De punten zijn echter niet meer in de repen te vinden terwijl de leerling wel nog repen heeft gekocht. De opdracht voor de leerling luidde als volgt:

Schrijf een briefje dat je meestuurt met je punten en wikkels. Vertel waarom je geen 10 punten kunt opsturen.

Overtuig de firma SMIKKEL ervan dat jij er niets aan kunt doen dat je geen 10 punten hebt, maar dat je wél die muziek-CD wilt ontvangen .

Zorg ervoor dat ze je de muziek-CD toch sturen!

Like-opdracht

Uit de Like-opdracht vloeiden instruerende teksten voort. De leerlingen moesten bij deze opdracht namelijk een nieuwe leerling, Like, die een jaar in het buitenland heeft gewoond welkom heten met een instructie waarin uitgelegd wordt hoe je in het Nederlands een goede tekst schrijft.

Deze teksten zijn allen bekeken en beoordeeld. Vijf teksten werden uit de beoordeelde teksten gekozen en als anker teksten geclassificeerd. Met deze anker teksten is een schaalverdeling gemaakt. Dit betekent dat de gemiddelde anker tekst lijkt op de gemiddeld beoordeelde teksten. De schaal werkt als interval schaal. Dit houdt in dat de beoordeling van een tekst van 0 tot oneindig kan lopen. De anker teksten krijgen echter de scores via de normaalcurve en variëren dus tussen 70 punten, de gemiddelde tekst 100 punten en 130 punten.

Van deze bulk teksten is een selectie teksten gemaakt: per opdracht zijn veertig teksten geselecteerd die een plaats innemen in de normaalcurve. Dit betekent dat er één tekst geselecteerd is die slechter scoorde dan 98% van de andere teksten, 4 teksten die beter scoorden dan 2% maar slechter dan 74% van de beoordeelde teksten, 15 teksten die beter dan 16% van de teksten scoorden maar slechter dan de helft van de teksten. En kant van de normaalcurve hetzelfde: één tekst die beter scoorde dan 98%, 4 teksten die slechter scoorden dan 2% maar beter dan 74% van de andere teksten en 15 teksten die slechter dan 16% maar beter dan de helft van de teksten werden beoordeeld. Deze selectie zorgt er voor dat de teksten die gekozen zijn een representatie zijn voor alle beoordeelde teksten.

In dit onderzoek worden de teksten alleen aan de hand van de Smurfenschaal (de schaal die gemaakt werd met de anker teksten van de Smurfopdracht) beoordeeld. Bij het schaalmodel werd een uitleg meegegeven zodat duidelijk was waarom elke anker tekst de score heeft gekregen die het had. Ook werd hierin een algemene uitleg gegeven op welke

aspecten gelet moest worden. Bijvoorbeeld dat het belangrijk was dat de beoordelaars meer urgentie gaven aan inhoud en argumentatie dan aan spelling en interpunctie. De uitleg en schaal zijn terug te vinden in respectievelijk bijlage 2A en bijlage 2B.

Proefpersonen

Voor dit onderzoek zijn vijf proefpersonen gezocht die op het gebied van studie bezig zijn met taal of communicatie. Dit zijn personen die tweede of derde jaars zijn en woonachtig in de buurt van Utrecht. De leeftijd van de proefpersonen varieerde tussen 20 en 26 jaar. De reden voor de keuze van deze proefpersonen is dat er wordt geacht dat zij een affiniteit met taal hebben en hierdoor ingeschat worden als goede beoordelaars van teksten.

Procedure

De proefpersonen werden in een ruimte geplaatst onder begeleiding van een van de onderzoekers om de teksten volgens de beoordelingschaal te beoordelen.

Voorafgaand aan de beoordeling kregen de proefpersonen een instructie van 15 minuten waarin hen werd gevraagd een instructie door te lezen (zie bijlage 3). In deze instructie werd uitgelegd dat zij met behulp van een schaalmodel de teksten gingen beoordelen. Er werd hun gevraagd niet terug te bladeren – om scores aan te passen – en nauwkeurig te werk te gaan. Na het lezen van de instructie kregen zij een drietal voorbeelden om mee te werken onder begeleiding van een van de onderzoekers. Vervolgens konden zij aan de slag.

De proefpersonen behandelenden als eerste de Smurfenteksten, dan de Smikkelteksten en dan de Like-teksten. Hierdoor liep de congruentie per 40 teksten af. De volgorde van de teksten waren voor iedere beoordelaar hetzelfde. Voor de beoordeling zelf, om de 120 teksten na te kijken, kregen zij maximaal twee uur.

Data-analyse

Uit dit onderzoek kwamen voor vijf beoordelaars 120 tekstbeoordelingen per persoon voort. Om de betrouwbaarheid tussen de vijf beoordelaars van de Smurfenschaal te bepalen is de Cronbachs alpha berekend. Deze is per opdracht uitgerekend en een maal werden de opdracht met hoge congruentie en de opdracht met middelmatige congruentie samengenomen. Dit leverde vier waarden van Cronbachs alpha op die daarna onderling vergeleken werden: de hoge, gemiddelde en lage congruentie werden met elkaar vergeleken en de samengenomen hoge en gemiddelde werd met de lage congruentie vergeleken. Deze vergelijkingen zijn uitgevoerd door een K-sample-significantietest, die meet of de verschillen tussen de waarden van betrouwbaarheid significant zijn.

Resultaten

De data die dit onderzoek naar voren heeft gebracht, is geanalyseerd in SPSS.

Om de beoordelaarsovereenkomst per opdracht te berekenen wordt de cronbachs alfa uitgerekend. Aan de hand van dit getal wordt er gekeken of de beoordelaars een hoge overeenstemming hebben. Een hoge overeenstemming wordt bereikt wanneer deze waarde significant is. De beoordelaarsovereenkomst bij de Smurfopdracht bedraagt .91. Dit betekent

dat de overeenkomst zeer hoog is. Bij de Smikkelopdracht bedraagt de beoordelaarsovereenkomst .92 Tot slot bedraagt de beoordelaarsovereenkomst van de Like-opdracht .84 Dit betekent dat bij het gebruik van de Smurfschaal, de beoordelaars bij alle drie de opdrachten hoog overeenstemmen.

Vervolgens wordt gekeken in hoeverre de waarden van de beoordelaarsovereenkomst van elke opdracht onderling verschillen. Het verschil van de waarden van beoordelaarsovereenkomst tussen de Smurf- en Smikkelopdracht is niet significant ($F(1, 39) = 1.23, p = .26$). Tussen de Smurf- en Like-opdracht is er ook geen significant, maar wel een groter verschil te vinden ($F(1, 39) = 0.64, p = .06$). Tussen de Smikkel en Like is wel een significant verschil gevonden ($F(1, 39) = 2.03, p = .02$).

Tot slot is er gekeken naar het verschil tussen teksthandelingen en daarvoor zijn de beoordelaarscores van de Smikkel- en Smurfenteksten samen genomen. Bij deze samenvoeging bedraagt de beoordelaarsovereenkomst .91 ($\alpha = .91$). Dit geeft een significant verschil in vergelijking met de Liketeksten ($F(1, 79) = 1.79, p = .01$).

Discussie

Dit onderzoek werd ingeleid met het probleem dat een te lage schrijfvaardigheid hebben. Hiervoor is het noodzakelijk om tot een goede beoordeling van de schrijfvaardigheid te komen. Het is echter gebleken dat dit een moeilijk opgave is. Om het probleem van de beoordeling van schrijfvaardigheid tegen te gaan wordt gebruik gemaakt van verscheidene beoordelingsmethoden, zoals de holistische, analytische methode en schaalbeoordelingsmethode. In dit onderzoek is de schaalbeoordelingsmethode onderzocht.

De vraag die centraal stond in dit onderzoek was of het onderwerp en/of teksthandeling bij het beoordelen van teksten van basisschoolleerlingen door taalstudenten aan de hand van de beoordelingsschaal, invloed heeft op de beoordelaarsovereenstemming. Aan de hand van de resultaten kan gesteld worden dat de teksthandeling van invloed is op de beoordelaarsovereenstemming. Er is gebleken dat teksten die de minste overeenkomst hadden met de anker teksten (op grond van het verschil in teksthandeling en onderwerp ten opzichte van de anker teksten) significant lager beoordeeld werden dan de teksten die meer overeenkomst vertoonden. Aan de hand van de resultaten kan gesteld worden dat wanneer de teksthandeling van de te beoordelen tekst niet overeenkomt met de anker teksten deze minder goed beoordeeld worden dan wanneer de teksthandeling wel overeenkomt. Dit betekent dat een schaalmodel dat gebaseerd is op een bepaalde teksthandeling minder bruikbaar is voor teksten die deze teksthandeling niet volgen. Ook is duidelijk geworden dat teksten met dezelfde teksthandeling en een verschillend onderwerp op dezelfde wijze beoordeeld kunnen worden aan de hand van één schaal. Het is dus mogelijk om in de praktijk per teksthandeling één schaal te maken.

De resultaten hebben echter ook aangetoond dat de beoordelaarsovereenstemming bij iedere opdracht hoog is. Dit geeft aan dat hoewel er een significant verschil is gevonden in beoordelaarsovereenstemming tussen de teksten verschillen qua taalhandeling met de anker teksten en de teksten die niet verschillen, dit niet betekent dat de beoordelaarsovereenstemming van de teksten die de minste overeenkomst vertonen, laag is. Dit betekent dat het mogelijk is te beoordelen aan de hand van schaalmodellen, ongeacht het

ondewerp of de teksthandeling – met in het achterhoofd dat de beoordelaarsovereenstemming bij een andere teksthandeling lager kan uitvallen. Aan de hand van deze resultaten is het mogelijk om schaalmodellen te gebruiken om de taalvaardigheid van basisschoolleerlingen aan te pakken. Dit onderzoek heeft aangetoond dat de beoordelaars op goede wijze de teksten konden beoordelen. Het is dus niet ondenkbaar dat in de toekomst docenten aan de hand van goede schaalmodellen teksten van leerlingen kunnen beoordelen. Verder biedt het nieuw inzicht in de reeds bestaande discussie waarin gedacht en gezegd werd dat schaalmodellen slechts voor één tekst gebruikt zouden kunnen worden. Het druist dus in tegen de notie van van den Bergh en Meuffels (2000) dat voor iedere nieuwe opdracht een nieuw schaalmodel gecreëerd dient te worden.

De verklaring voor dat deze resultaten naar voren komen (ondanks de verwachte resultaten) is dat er geen rekening mee is gehouden dat alle drie de opdrachten enigszins aan de hand van dezelfde wijze beoordeeld konden worden, zoals Pander Maat (2002) ons liet zien. Hiermee wordt bedoeld dat alle drie de opdrachten aan de hand van structuur op basis van argumentatie en aanwezigheid van belangrijke elementen konden worden beoordeeld. Omdat de Smurfen- en Smikkelteksten zodanig op elkaar lijken, gaf de beoordeling van de twee totaal geen problemen. Het onderwerp heeft er waarschijnlijk alleen voor gezorgd dat de proefpersonen zich opnieuw even hebben moeten inlezen. Daarnaast lijken ook de Like-teksten veel op de andere twee opdrachten. Zowel de instruerende als directieve teksten worden namelijk sterker op basis van goede argumentatie en aanwezigheid van wezenlijke elementen (Pander Maat, 2002). Dit verklaart waarom de Like-teksten aan de hand van de Smurfenschaal toch een hoge beoordelaarsovereenstemming hadden. De beoordelaars moesten zich richten op structuur en op aanwezigheid van belangrijke elementen. Hierdoor werden zij direct de goede kant uit gewezen.

Hoewel voorgaande conclusies getrokken mogen worden, dienen wel enkele zaken in het achterhoofd gelaten te worden. Uit de informatie van Pander Maat (2002) is naar voren gekomen dat instruerende teksten, zoals de Like-teksten, op directieve teksten kunnen lijken. Het is dus mogelijk dat de verwachte lage beoordelaarsovereenstemming van de Like-teksten uit is gebleven omdat de schaal die ontsloten is uit een directieve tekst, goed gebruikt kan worden bij het beoordelen van een instruerende tekst. Bij het beoordelen van een bijvoorbeeld informerende tekst aan de hand van een schaal van een directieve tekst zouden zich echter meer problemen voor kunnen doen. Deze lijken immers veel minder op elkaar. Dit onderzoek kan het helaas niet uitsluiten dat verschillende teksthandelingen niet met hetzelfde schaalmodel kunnen worden beoordeeld. Het is daarom van belang dat in verder onderzoek wordt ingegaan op de mogelijkheid van implementatie van schaalmodellen op teksten met teksthandelingen die verder van de teksthandeling van het schaalmodel liggen. Daarnaast is het ook interessant om te onderzoeken of – in een omgekeerd onderzoek – schaalmodellen, ontsproten uit instruerende teksten, gebruikt kunnen worden bij het beoordelen van directieve teksten.

Een ander punt van dat in het achterhoofd moet worden gehouden, is dat hoewel de proefpersonen die gekozen zijn allen affiniteit hebben met taal en een taalgerelateerde studie genieten, dit hen tot aanvaardbare beoordelaars, maar nog geen getrainde docenten maakt. De docentenopleiding geeft toch verschillen met de opleiding die de studenten volgen. De vraag blijft dan ook of docenten deze schaalmodellen op dezelfde wijze kunnen toepassen als de

studenten. Dit lijkt zich echter gemakkelijk op te lossen door een training te geven. De proefpersonen in dit onderzoek hebben niet meer dan drie oefenteksten en anderhalf a-4 instructie voor hen beschikbaar gehad. Dit kan voor docenten uiteraard uitgebreid worden.

Een andere mogelijkheid die zich voor zou hebben kunnen gedaan is het effect van volgorde. De Like-teksten, die het minst op de anker teksten leken, zijn pas als laatste beoordeeld – na tachtig andere teksten. Hierdoor is het mogelijk dat de beoordelaars beter zijn geworden in het beoordelen met een sterkere beoordelaarsovereenstemming ten gevolge. Dit is echter een speculatie en zal onderzocht kunnen worden door andere proefpersonen de Like-teksten het eerst te laten beoordelen.

Algemene conclusie

De beperkingen van dit onderzoek in acht genomen kan gesteld worden dat dit onderzoek heeft geleid tot nieuwe inzichten bij het beoordelen van schrijfvaardigheid. Het is duidelijk geworden dat het niet nodig is om beoordelingsschalen dezelfde teksthandeling en onderwerp hebben als de te beoordelen teksten. Dit biedt een oplossing voor het tot op heden bestaande praktische probleem dat de beoordelingsschalen zo arbeidsintensief zijn. Of deze oplossing echter voldoende is om alle problemen op te lossen is nog niet duidelijk, zoals het probleem dat er onenigheid is in wat de criteria, waaraan aan tekst dient te voldoen, zijn. Daarnaast (en daarom) is het een vraag of het vooronderzoek, waarin de anker teksten gekozen worden betrouwbaar genoeg is. Deze beperkingen dienen verder onderzocht te worden alvorens de beoordelingsschalen in de praktijk op grote schaal geïmplementeerd kunnen worden. Wel fungeert dit onderzoek als een opstap naar een nieuwe denkwijze, waarin het wellicht mogelijk is om schrijfvaardigheid betrouwbaarder te beoordelen.

Literatuur

- Bergh van den, H. & Meuffels, B. (2000). Schrijfvaardigheden en Schrijfprocessen. *Taalbeheersing als communicatiewetenschap*. 122-153.
- Kuiken, F. & Vedder, I. (2012). Kwantitatieve en Kwalitatieve Beoordeling van Schrijfvaardigheid in T2 en T1. In Jong de, N., Juffermans, K., Keijzer, M. & Rasier, L. *Papers of the Anéla 2012 Applied Linguistics Conference*. (pp. 96-104). Eburon Uitgeverij B.V.
- Liemberg, D. & Meijer, D. (2004). Taalprofielen. *Nationaal Bureau Moderne Vreemde Talen*.
- McColly, W. (1970). What Does Educational Research Say About the Judging of Writing Ability? *The Journal of Educational Research*, 64 (4), 147-156.
- Meijerink, H. P., Letschert, J. F., Rijlaarsdam, G. C. W., Bergh van den, H. H. & Streun van, A. (2009). Referentiekader Taal en Rekenen: de Referentieniveaus. *Doorlopende Leerlijnen Taal en Rekenen*.
- Mullis, I. V. S. (1984). Scoring Direct Writing Assessments: What Are the Alternatives? *Educational Measurement: Issues and Practice*. 3 (1), 16-18.
- Pander Maat, H. (2002). *Tekstanalyse: Wat Teksten tot Teksten Maakt*. Bussum. Coutinho.
- Pollmann, E., Prenger, J. & Glopper de, K. (2012). Het beoordelen van leerlingteksten met behulp van een schaalmodel. *Levende Talen Tijdschrift*. 13 (3), 15-24.
- Wesdorp, H. (1981). De evaluatie van de schrijfvaardigheid. In *Evaluatietechnieken voor het moedertaalonderwijs*. 's-Gravenhage: Stichting voor Onderzoek van het Onderwijs.

Bijlage 1A

Schrijftaak: SMURFEN

Stel je voor ...

Je vader doet wekelijks inkopen bij de supermarkt (SuperCoop) om de hoek. Voor elke vijftientig euro ontvangt hij een mini-Smurf (verpakt in een zakje). Je hebt inmiddels al een hele verzameling maar je mist nog twee smurfen: de Brilsmurf en Grote Smurf. Natuurlijk hoop je ook op de Gouden Smurf. Als je de Gouden Smurf hebt, maak je kans om een digitale fotocamera te winnen. Elke vrijdag kijk je nieuwsgierig in de boodschappentas welke Smurfen je bij je verzameling kunt zetten.

In de folder van Supermarkt “SUPERCOOP” staat:

SPAAR SMURFEN EN MAAK KANS OP EEN PRACHTIGE DIGITALE CAMERA!

Zo werkt het:

Bij elke 25 euro aan boodschappen, ontvangt u een zakje met een Smurf. Zo kunt u uw kinderen verrassen met een prachtige verzameling van deze populaire poppetjes.

Heeft u een Gouden Smurf, dan maakt u kans op een digitale camera van 4.0 megapixels!

Stuur uw Gouden Smurf in een gefrankeerde envelop naar:

SMURF SPAARACTIE
postbus 3333
1273 AD Etten-Leur

Stuur ook EUR 0,80 aan postzegels mee voor de portokosten. Vermeld duidelijk uw naam, adres, postcode en woonplaats en u krijgt vóór 1 mei 2008

bericht of u een camera hebt gewonnen

Deze actie loopt tot 20 april 2008.

Je hoort op 15 april van je vader dat de zakjes met Smurfen op waren. Hij heeft een kassabon meegekregen met een stempel, zodat je later de Smurfen alsnog kunt krijgen. Hij heeft in totaal EUR 110,34 aan boodschappen uitgegeven. Je besluit de kassabon op te sturen met de vraag om vóór 20 april nog zakjes met Smurfen te kunnen ontvangen. Je hoopt natuurlijk op een Gouden Smurf!

Opdracht

Schrijf een briefje dat je meestuurt met de kassabon van 15 april. Leg uit dat bij SuperCoop om de hoek de Smurfen op waren.

Overtuig de firma SUPERCOOP ervan dat je vóór de sluitingsdatum van de actie de zakjes met Smurfen wilt ontvangen. Schrijf daarna de envelop.

Bijlage 1B

Schrijftaak: SMIKKEL

Stel je voor ...

Op de SMIKKEL-repen, die je weleens eet, heb je zien staan dat je gratis muziek CD's kunt krijgen.

Er staat:

SPAAR VOOR EEN GRATIS muziek CD!!

Dit moet u doen:

Op de wikkels van de repen zit 1 spaarpunt.

Spaar 10 punten en stuur deze in een voldoende gefrankeerde envelop naar:

SMIKKEL SPAARACTIE
postbus 3333
1273 AD Etten-Leur

Stuur ook EUR 0,80 aan postzegels mee voor de portokosten. Vermeld duidelijk uw naam, adres, postcode en woonplaats en de gratis (GRATIS!) muziek CD wordt dan zo spoedig mogelijk thuisgestuurd.

Deze actie loopt tot 15 juli 2009.

Je hebt 8 punten bij elkaar gespaard, maar nu kun je nergens meer repen met punten vinden. Op de repen in de winkels zit geen spaarpunt meer, hoewel het nog geen 15 juli is. Je kunt je 10 punten dus niet bij elkaar krijgen. Toch wil je zo'n CD hebben.

Je stuurt daarom je 8 punten op en doet er 2 hele wikkels zonder punt bij.

Opdracht

Schrijf een briefje dat je meestuurt met je punten en wikkels. Vertel waarom je geen 10 punten kunt opsturen.

Overtuig de firma SMIKKEL ervan dat jij er niets aan kunt doen dat je geen 10 punten hebt, maar dat je wél die muziekCD wilt ontvangen .

Zorg ervoor dat ze je de muziek-CD toch sturen!

Schrijf daarna de envelop.

Bijlage 2A



Beste deelnemer,

De Universiteit Utrecht is bezig met een onderzoek naar het beoordelen van schrijfp opdrachten in het basisonderwijs. Uit voorgaand onderzoek is gebleken dat veel docenten hun oordeel baseren op een globale indruk van de geschreven tekst. Hierdoor lopen de beoordelingen van verschillende docenten vaak ver uiteen.

Een mogelijke oplossing voor dit probleem is beoordeling door middel van een beoordelingsschaal. Op de schaal staan vijf ‘ankerteksten’ die staan voor verschillende schrijfniveaus. Door andere schrijfp opdrachten met de anker teksten te vergelijken, kan er gemakkelijker een score gegeven worden.

Voor dit onderzoek vragen wij jou om een reeks teksten te beoordelen aan de hand van een beoordelingsschaal. Dit gaat als volgt in zijn werk:

- Voor je ligt een beoordelingsschaal met anker teksten. De bedoeling is dat je alle anker teksten zo dadelijk goed doorleest om jezelf vertrouwd te maken met de schaal.
- Je krijgt drie stapeltjes schrijfp opdrachten, die je gaat beoordelen door ze te vergelijken met de anker teksten. Je geeft de opdrachten op basis hiervan een puntenscore (de schaal loopt van 0 punten tot oneindig). Let op: je hoeft hierbij niet een puntenaantal te kiezen dat overeenkomt met een van de anker teksten! Een score van bijvoorbeeld 92 punten is dus ook mogelijk.
- De score schrijf je bij de juiste tekst op het scoreformulier.

De teksten die je gaat beoordelen zijn niet allemaal onderdeel van dezelfde opdracht. Je krijgt drie stapeltjes met teksten, die geschreven zijn aan de hand van drie verschillende opdrachten: opdracht ‘Smurfen’, opdracht ‘Smikkel’ en opdracht ‘Like’. Bovenop ieder stapeltje vind je een beschrijving van de opdracht die de leerlingen kregen. Hieruit kun je afleiden wat de eisen aan de opdracht waren.

Je zult alle teksten gaan beoordelen met dezelfde schaal. De anker teksten op de schaal komen uit opdracht ‘Smurfen’. (De teksten uit opdracht ‘Smikkel’ en opdracht ‘Like’ lijken inhoudelijk dus minder op de anker teksten.)

Tot slot gelden bij de beoordeling nog enkele spelregels:

- Let er bij het beoordelen vooral op of het doel van de tekst wordt bereikt (bij ‘Smurfen’ en ‘Smikkel’ is dat overtuigen, bij ‘Like’ instrueren).
- Beoordeel de teksten één voor één, in de volgorde van het beoordelingsformulier.
- Je mag niet overleggen met de andere beoordelaars.
- Het is niet de bedoeling dat een eenmaal gegeven score na het lezen van andere teksten nog wordt aangepast. Doe dit dus alleen bij hoge uitzondering.
- Twijfel niet te lang bij het geven van een score: maximaal een minuut per tekst blijkt in de praktijk meestal voldoende.

Na het beoordelen van de teksten willen wij je vragen om het bijgevoegde vragenformulier in te vullen.

Alvast onze hartelijke dank voor je medewerking.

Joëlle Bemelman
Sanne Leendertse
Hanneke Liemburg
Erik Philippens
Carla van Rooijen

M.Sc. Renske Bouwer

Universiteit Utrecht
Onderzoeksinstituut UiL/OTS
Trans 10
3512 JK Utrecht

Bijlage 2B

Schaalbeoordeling Smurfen

<p>ik zou graag de smurfen actie ontvangen voor 20 april na sluitingstijd. Ik wil graag daar want ik heb 110,34 uitgegeven voor boodschappen en de smurfen waren op. 15/4/2008</p> <p>van: <naam leerling> <adres leerling></p>	<p>Beste geachte 15 april</p> <p>Mijn vader heeft aan boodschappen 110,34 euro uitgegeven voor boodschappen bij de Albert Heijn. Maar de smurfen waren op. Dus we kregen een stempel op onze bon en dat wil we dit moesten inleveren voor 20 april.</p> <p>vreundelijke groet <naam></p>	<p>Beste Supercoop</p> <p>Op 15 april hoorde ik van mijn vader dat de smurfen op waren terwijl mijn vader recht had op 4 smurfen. Ik vraag u bij deze of wij die vier smurfen Alsnog mogen want we moeten nog maar 2 smurfen. Ik vraag u nogmaals mogen wij die smurfen. Als u antwoord hebt Alvast bedankt.</p> <p><namen></p> <p>bijge bijlage: het bonnetje</p>	<p>Hallo meneer/mevrouw,</p> <p>Mijn vader had laast boodschappen gedaan. bij de supermarkt supercoop. Maar daar was waren de smurfen op dus toen kreeg hij een bon, om de smurfen te krijgen. maar het is bijna 20 april dus ik zou het leuk vinden als ik ze voor 20 april krijg.</p> <p>het gaat om een bedrag van €110,34 ik spaar de smurfen allang dus ik zou ze allemaal wel willen hebben.</p> <p>En de goude smurf misschien zou ik ook heel heel graag willen.</p> <p>Dus ik zou het leuk vinden als ik ze voor 20 april zou kunnen krijgen</p> <p>Groetjes <naam leerling> <adres leerling></p>	<p><plaatsnaam> 15-04-08</p> <p>Geachte firma Supercoop.</p> <p>Mijn vader doet wekelijks boodschappen bij supermarkt Supercoop. Daardoor heb ik al een hele verzameling smurfen. Alleen mis ik er nog 2: Erislurf en Grote Smurf. Mijn vader ging vandaag weer boodschappen doen en kwam erachter dat de smurfen op waren. Ik zou het jammer vinden als ik daardoor mijn collectie niet compleet kan maken. Daarom zou ik graag aan u willen vragen of ik nog smurfen zou kunnen krijgen met de stempel op het bonnetje van de boodschappen. Het bonnetje zit hierbij in de envelop. Als het idee door zou kunnen gaan is hier mijn adres: <adres leerling></p>
70 punten	85 punten	100 punten (gemiddelde score)	115 punten	130 punten

Toelichting bij de ankerpunten:

De gemiddelde ankertekst (100 punten) bevat de volgende plus- en minpunten:

- Het probleem is beschreven, maar dat had nog helderder gekund: waarom had de vader recht op de smurfen?
- Er wordt een duidelijke vraag gesteld, maar deze is wel twee keer genoemd – waarom? Dit maakt de brief niet overtuigender.
- De brief bevat een conventionele aanhef en afsluiting.
- De brief is gericht aan de lezer: ‘ik vraag u ...’
- De zinnen lopen grammaticaal goed, het taalgebruik is formeel en er zitten geen spellingsfouten in.

Minpunten:

- Inhoudelijk bevat de brief nog te weinig informatie om echt overtuigend te zijn.
- Het adres is niet genoemd (alleen een e-mailadres), waardoor de smurfen niet daadwerkelijk toegestuurd kunnen worden.
- Het hoofdlettergebruik is niet in orde.

De tekst met 115 punten is beter dan deze gemiddelde tekst omdat:

- er meer informatie is gegeven: zo is het bedrag expliciet beschreven, wordt de einddatum van de actie genoemd en betreft de briefschrijver emoties in de brief om de vraag kracht bij te zetten (doet allang mee met de actie, zou heel heel graag de gouden smurf willen hebben etc.). Het probleem wordt hierdoor duidelijker en de brief in zijn geheel overtuigender.
- de briefconventies voldoende zijn: er is een aanhef en een afsluiting. Dit had wel formeler gekund.
- de organisatie van de inhoud ok is: eerst wordt het probleem beschreven, dan de vraag gesteld met de onderbouwing waarom dit zo belangrijk is.

Een minpunt van de brief is dat het taalgebruik niet zo gevarieerd is, er worden zelfs hele zinnen herhaald (“dus ik zou het leuk vinden als ik voor ...”). Ook zijn de zinnen erg kort en niet altijd logisch afgebroken.

De tekst met 130 punten is nog beter dan de tekst met 115 punten omdat:

- de brief overtuigend is: het probleem is heel helder beschreven, de lezer geeft veel extra relevante informatie, de vraag is heel duidelijk gesteld.
- de structuur prima is: de zinnen lopen goed en zijn logisch opgebouwd, er wordt goed gebruik gemaakt van voegwoorden. Er missen wel alinea's.
- de briefconventies in orde zijn: er is een formele aanhef, de datum en het adres is genoemd. Er mist wel een afsluiting, waardoor de brief niet helemaal af lijkt.
- het taalgebruik in de brief is gevarieerd en de stijl is formeel. Er zijn geen fouten in de grammatica, spelling en/of interpunctie.

De tekst met 85 punten is slechter dan de gemiddelde tekst omdat:

- het probleem is wel genoemd, maar er is geen duidelijke vraag gesteld over het opsturen van de smurven. Ook ontbreekt overtuigende informatie. Het is dus maar de vraag of de smurven worden toegestuurd en al helemaal of dit voor 20 april zal zijn.
- de aanhef niet goed is: “beste geachte”.
- spelling en hoofdlettergebruik niet helemaal goed gaat.

De tekst met 70 punten is nog slechter dan de tekst met 85 punten omdat:

- het inhoudelijk onder de maat is: er wordt te weinig en foutieve/onduidelijke informatie gegeven: bv. “smurfenactie”, “na sluitingstijd”. De brief is dus niet duidelijk.
- de brief structuur mist. Het is maar 1 zin en de informatie is daarbinnen ook niet logisch georganiseerd.
- het niet voldoet aan de briefconventies: er is geen aanhef en geen ondertekening. Ook staat de datum op de verkeerde plek.
- er geen gebruik is gemaakt van hoofdletters aan het begin van de zin. Er zijn geen fouten in de grammatica of spelling gemaakt.