

A close-up photograph of a person's hand holding a black pen, writing on a white page. The background is a blurred grid of light blue circles. A large, semi-transparent blue square with rounded corners is overlaid on the bottom left, containing white text.

**Dankzij de speelse  
schrijfopdracht  
vonden mijn leerlingen  
het echt leuk om aan  
de tekst te werken!**

# Het toetsen van schrijfvaardigheid<sup>1</sup>

## ■ Uriël Schuurs

Dr. U.R.I. Schuurs was van 2012 tot zomer 2021 werkzaam voor Stichting Cito, afdeling Centrale Toetsen en Examens  
E-mail: [uriel.schuurs@deaanzet.nl](mailto:uriel.schuurs@deaanzet.nl)

Toen ik in de jaren '70 de middelbare school afsloot, kreeg ik bij het examen Nederlands een tiental opsteltitels voorgeschoteld. Daaruit moesten we er een kiezen en binnen twee uur een tekst schrijven die grammaticaal correct, goed opgebouwd én boeiend was voor de lezer. Ik schreef een maatschappijkritisch betoog – met de nodige humoristische voorbeelden erin om de lezer geboeid te houden. Mijn eigen docent vond die tekst een 9 waard, de tweede corrector gaf een 3: “Mooi betoog, maar humor hoort daar niet in thuis”, zo motiveerde hij zijn oordeel.

## Beoordeling van tekstkwaliteit

Tot 1998 werd schrijfvaardigheid getoetst in het centraal examen, in de vorm van een keuze uit een aantal opsteltitels of – mondjesmaat – in de vorm van gericht schrijven, met een door de leerling zelf aangelegde documentatiemap. Vanaf 1998 wordt schrijfvaardigheid op havo en vwo niet meer centraal getoetst, maar komt schrijven aan de orde in het schoolexamen. Dit geeft docenten meer vrijheid in de vormgeving van hun onderwijs, maar tegelijkertijd heeft deze verandering het zicht verkleind op wat er precies gebeurt in het schrijfvaardigheidsonderwijs (Bonset, 2010). Bovendien bestaat de indruk dat hierdoor het schrijven minder aandacht krijgt op school dan voorheen en dat de schrijfvaardigheid van leerlingen flink achteruit is gegaan. Zie bijvoorbeeld Rooijackers (2007) die een aantal vakdeskundigen over deze kwestie heeft geïnterviewd. Wat leerlingen mogen of moeten gebruiken aan documentatie bij het schoolexamen schrijfvaardigheid verschilt

nogal tussen scholen (Goosen, 2001). Ook de weging van het cijfer voor het schoolexamen schrijfvaardigheid varieert nogal, van 15% en 45% (Meestringa & Ravesloot, 2013), net als de lestijd die aan schrijfvaardigheid wordt besteed. Het aantal genres dat aan de orde komt, laat juist weinig variatie zien: als er wordt geoefend, is dat met de brief, het betoog en de beschouwing (Meestringa & Ravesloot, 2014). Blijkens een enquête onder docenten Nederlands is een meerderheid er voorstander van om schrijfvaardigheid weer in het centraal examen op te nemen (Advies examen Nederlands, 2018).

Uit de vakliteratuur blijkt intussen dat de beoordeling van schrijfvaardigheid op zijn minst problematisch is: de oordelen kennen doorgaans een lage betrouwbaarheid en de overeenstemming tussen beoordelaars is vaak laag. Dat maakt het minder aantrekkelijk om het centraal leesvaardigheidsexamen minder afnametijd te gunnen ten gunste van aandacht voor schrijfvaardigheid in het examen.

<sup>1</sup>Aan het onderzoek waarover hier wordt gerapporteerd, is een belangrijke bijdrage geleverd door Corien Stevens, Eva Cornuijt en Fanny Versteegh (allen werkzaam bij Cito). Onze dank gaat uit naar de docenten die hebben meegewerkt aan de ontwikkeling van de schrijfpoddrachten, beoordelingsmodellen en lesmaterialen: Esra Joustra (Veluws College Apeldoorn), Gerdien Keita (Stedelijk Lyceum Zutphen), Martijn de Bont (Stedelijk Gymnasium Nijmegen) en Martijn Koek (Keizer Karel College Amstelveen).

Immers, minder tijd voor het leesexamen betekent automatisch minder opgaven en daardoor een afname in betrouwbaarheid. Dat lijkt niet verkieslijk voor het centraal examen van een kernvak.

De beoordeling van de kwaliteit van teksten wordt beïnvloed door de volgende vijf factoren:

### **1 Het aantal schrijfpoddrachten**

De prestaties van leerlingen blijken sterk te verschillen van schrijfpoddracht tot schrijfpoddracht. Om die reden is het noodzakelijk om de schrijfvaardigheid met meerdere schrijftaken te toetsen (zie bijvoorbeeld Bouwer & van den Bergh 2015, p. 11).

### **2 Het aantal verschillende tekstgenres dat de leerling schrijft**

Bepaalde tekstgenres blijken makkelijker te zijn voor leerlingen dan andere genres. Zo lijkt een mailtje aan de schoolbibliotheek eenvoudiger dan een doorwrocht essay of een onderzoeksverslag. Hoe beter de schrijver bekend is met een genre, des te minder moeite die zal hebben een tekst te schrijven die voldoet aan de conventies behorende bij het genre en die de intenties goed overbrengt bij het beoogde publiek (Schreurs et al., 2017).

### **3 De tekstkenmerken die worden beoordeeld**

De beoordeling van een schrijfproduct kan zich richten op een veelheid aan tekstkenmerken: spelling of zinsbouw, of – op hoger niveau – de mate waarin de auteur zich richt op het beoogde publiek of genrespecifieke kenmerken correct en op een voor de lezer herkenbare manier in de tekst verwerkt (zie Schuurs & De Groot, 2018). Als je je in het voortgezet onderwijs beperkt tot beoordeling van spelling – door leken het meest van belang geachte aspect – dan zul je weinig verschil tussen leerlingen zien. Ga je kijken naar stijl, tekstopbouw, alinea-indeling of argumentatie, dan gaan de prestaties van een leerling per tekst aanzienlijk verschillen.

### **4 Het aantal beoordelaars dat wordt ingeschakeld**

Er zijn meerdere beoordelaars nodig om een betrouwbare beoordeling te krijgen. Zo gebruiken bijvoorbeeld Kuhlemeier et al. (2017) maar liefst negen beoordelaars per tekst om tot robuuste resultaten te komen. Niet alleen speelt de vraag hoeveel beoordelaars nodig zijn, ook is het de vraag wie als beoordelaar optreedt. Zo is het verleidelijk om de zogenoemde *self-assessment* en de *peer-assessment* als didactisch hulpmiddel op te nemen in het schrijfonderwijs (zie bijv. Sluismans et al., 2013). Op die manier kunnen aan de leerling de standaarden worden overgebracht waaraan een schrijfproduct zou moeten voldoen. Elders wordt vooral de gebrekkige betrouwbaarheid van *peer-assessment* en *self-assessment* benadrukt (zie Schuurs en Verhoeven, 2011, voor een meta-analyse op dit punt).

### **5 De beoordelingsmethode die wordt gehanteerd**

Er is een veelheid aan soorten beoordelingen mogelijk. Veelgenoemd is het onderscheid tussen analytische beoordelingsmodellen die op gedetailleerde, tevoren gedefinieerde tekstkenmerken focussen, en de globale beoordeling die een algemene kwaliteitsindruk moet weergeven. Bij globale beoordeling spreekt de beoordelaar een oordeel uit over de kwaliteit van de tekst als geheel. De analytische beoordeling is tot op heden niet betrouwbaarder gebleken dan de globale; de ontwikkelde analytische beoordelingsschema's zijn volgens Bonset en Braaksma (2008, p. 129) vaak onvoldoende van kwaliteit.

Ook kan men gebruikmaken van rubrics waarbij per beoordelingsaspect de schaalpunten nauwkeurig omschreven zijn. Daarnaast is het mogelijk een beoordelingsschaal te hanteren waarbij de punten op de beoordelingsschaal op beredeneerde plekken staan – bijvoorbeeld elk een standaarddeviatie van elkaar geplaatst – en zogenoemde anker teksten als voorbeeld dienen voor de schaalpunten. De

laatste jaren krijgt de paarsgewijze vergelijking veel aandacht. Bij deze methode vergelijken men een groot aantal keren een aantal teksten paarsgewijs met elkaar. Deze methode lijkt de betrouwbaarheid van de beoordeling te vergroten (Bouwer & Koster, 2016); de claim dat deze beoordelingsmethode tijdbesparend zou werken, lijkt vooralsnog niet onderbouwd (Coertjes et al., 2017).

### Een genre-specifieke benadering

In een poging enkele van de genoemde problemen te ontlopen, hebben we een exploratief onderzoek uitgevoerd waarbij we genre-analyse als uitgangspunt hebben genomen. Genre-analyse is een relatief nieuwe tak van tekstwetenschap waarin specifieke kenmerken van tekstgenres worden bestudeerd. In algemene zin lijkt genre-georiënteerd lees- en schrijfonderwijs een goede gelegenheid te bieden om leerlingen de essentie van teksten in een bredere maatschappelijke context te leren begrijpen (Vis, 2019). De onderscheidende kenmerken van een genre blijken goed beschrijfbaar aan de hand van kenmerken van het vereiste Taalgebruik, van de Inhoud en van de Vorm en structuur.

Het idee achter ons onderzoek was, dat in overleg met docenten tekstgenres worden gekozen en genrespecifieke kenmerken (kenmerken van tekststructuur, specifieke woordkeuze, stijlkenmerken) worden benoemd die de basis vormen voor meer eenduidige analytische beoordelingsschema's.

Het onderzoek bestond uit drie fasen:

fase A constructie: genres kiezen en materialen maken

fase B uitvoering: leerlingen schrijven de teksten en beoordelen elkaars teksten

fase C analyse: klassiek beoordelaarsonderzoek

In de constructiefase lieten we elk van de vier deelnemende docenten drie genres kiezen waarvoor ze een schrijfpodracht en introdu-

cerend lesmateriaal zouden gaan maken (zie Tabel 1).

Tabel 1. De gekozen tekstgenres en de beoogde leerlingniveaus

Tekstgenre	Niveau/ leerjaar
Mail naar instantie	havo 4
Foodblog	havo 3
Nieuwsbericht	havo 2
Creatief essay	vwo 5
Verzoekmail	vwo 3
Motivatiebrief	havo 4
Column	vwo 3
Recensie	vwo 1
Verslag van een gebeurtenis	vwo 1
Ideële reclame	vwo 5
Instructie	vwo 1



**Met genrespecifieke eisen in gedachten wordt mijn onderwijs veel gericht**

We hebben erop gelet dat deze genres vanuit genre-analytisch perspectief wel verschillend van elkaar zijn. Zo kan men denken dat de "mail naar instantie" en de "verzoekmail" erg op elkaar lijken, maar ze verschillen genre-analytisch gezien flink van elkaar: de "mail naar instantie" behelst de vraag om het entreegeld voor een afgelast concert terug te betalen, terwijl de "verzoekmail" het verzoek betrof om een gastles op school te verzorgen. De twee genres verschillen dus aanzienlijk in de verhouding tussen afzender en geadresseerde, wat ook consequenties heeft voor eisen aan het taalgebruik en bijvoorbeeld de mate waarin de zender moeite moet doen om de ontvanger te overtuigen.

Per genre ontwikkelden de deelnemende docenten:

- een prikkelende schrijfo opdracht
- een bijbehorend genrespecifiek beoordelingsmodel
- een kort lessenreeksje per genre

In Figuur 1 ziet u een voorbeeld van een van de schrijfo opdrachten. Per opdracht hebben we geprobeerd een zoveel mogelijk genrespecifiek beoordelingsmodel op te stellen. Daarbij maakten we gebruik van analytische beoordeling aan de hand van rubrics: in de beoordelingsmodellen zijn per beoordelingsaspect de eisen beschreven die gelden voor de twee uitersten van de gebruikte vierpunts-



Kruip in de huid van een journalist en schrijf een nieuwsbericht. Gebruik hierbij de bovenstaande vier foto's. Bekijk de foto's goed en bedenk wat er volgens jou is gebeurd. Deze zelfbedachte gebeurtenis beschrijf je in een nieuwsbericht.

Je mag de foto's in willekeurige volgorde gebruiken, maar zorg er wel voor dat je ze allemaal gebruikt.

Figuur 1. Voorbeeld van een schrijfo opdracht

schaal. In de lessenreeks werd het betreffende genre kort geïntroduceerd; daarbij werden bijvoorbeeld enkele geslaagde en minder geslaagde voorbeelden van het genre aan de orde gesteld. Daarna moesten de leerlingen de schrijfpdracht maken en tot slot beoordelen ze anoniem een tekst van een klasgenoot.

### Resultaten van empirisch onderzoek

De corona-uitbraak in 2020 maakte het onmogelijk om alle geplande lessen uit te voeren. In het najaar van 2020 heeft alsnog een beperkt aantal leerlingen de schrijfpdrachten gemaakt. Elk leerlingwerk is door vier beoordelaars beoordeeld volgens het schema uit Tabel 2.

In Tabel 3 staan de belangrijkste resultaten van het veldonderzoek weergegeven. De linkerkant van de tabel geeft de gemiddelde oordelen van de vier typen beoordelaars weer. Deze cijfers zijn een sommatie van drie oordelen over Taalgebruik, Inhoud en Vorm en structuur. Aan de rechterkant van de tabel staan de correlaties tussen de scores van de vier typen beoordelaars weergegeven.

Wat opvalt is dat de leerlingen aanzienlijk hogere scores toekennen aan het werk van hun klasgenoten dan de docenten. Ook laat Tabel 3 zien dat de scores van de leerlingen laag correleren met die van de docenten. Dat betekent dat de leerlingen niet alleen milder beoorde-

Tabel 2. Vier beoordelaars met verschillende beoordelingschema's

beoordelaar	Beoordelingsmodel
docent 1 (de eigen docent)	door Cito aangereikt, genre-specifiek beoordelingsmodel
docent 2	door Cito aangereikt, genre-specifiek beoordelingsmodel
docent 3	maakt zelf een beoordelingsmodel op basis van de drie categorieën Taalgebruik, Inhoud en Vorm en structuur
mede-leerling	maakt zelf een beoordelingsmodel op basis van de drie categorieën Taalgebruik, Inhoud en Vorm en structuur

Tabel 3. Gemiddelde score per type beoordelaar en correlaties tussen oordelen

Beoordelaar	n	Oordelen		Correlaties		
		Gemiddelde score	st.dev.	Docent 1	Docent 2	Docent 3
Leerling	233	9,41	1,67	0,26	0,31	0,20
docent 1	252	8,25	1,59		0,62	0,41
docent 2	255	8,36	2,02			0,47
docent 3	255	8,35	1,89			



## Leerlingen leren er veel van als ze zelf beoordelingsnormen moeten opstellen en die dan moeten toepassen op andermans tekst

len dan docenten, maar ze ordenen het werk ook anders dan de docenten. Hun beoordelingen lijken om die reden niet bruikbaar als onderdeel van toetsing van schrijfvaardigheid<sup>2</sup>.

Overigens gaven leerlingen te kennen dat ze het wel heel leerzaam vonden om het werk van medeleerlingen te beoordelen; didactisch gezien is *peer evaluation* dus een waardevolle exercitie.

Verder valt in Tabel 3 op dat de correlatie tussen docent 1 en docent 2 het hoogste is. Dit zijn juist de docenten die met het door Cito aangereikte, genre-specifieke beoordelings-schema hebben gewerkt. Een belangrijke vraag is, of genrespecifiek beoordelen nu leidt tot relatief homogener oordelen dan op een niet genrespecifieke manier beoordelen. Om deze vraag te beantwoorden, hebben we de overeenstemming tussen docent 1 en 2 vergeleken met die tussen docent 1 en 3 en tussen docent 2 en 3. Dit hebben we gedaan met gebruikmaking van Cohens' kappa, een statistische maat die beter rekening houdt met toevalsovereenstemming dan gewone correlaties. In Tabel 4 geven we het aantal keren lichte tot matige overeenstemming weer tussen de drie types docent-beoordelaars.

De tabel laat zien dat er tussen docent 1 en 2 zestien keer sprake is van lichte tot matige beoordelaarsovereenstemming; tussen docent 1 en 3 is daarvan maar acht keer sprake en tussen docent 2 en 3 slechts zeven keer. Met de kanttekening dat het hier verkennend onderzoek betreft en geen hypothesetoetsend onderzoek, kunnen we voorzichtig conclude-

Tabel 4. Aantal keren lichte tot matige overeenstemming tussen de drie types docent-beoordelaars

	Taalgebruik	Inhoud	Vorm en structuur
tussen docent 1 en 2	5	6	5
tussen docent 1 en 3	2	4	2
tussen docent 2 en 3	2	1	4

<sup>2</sup>We hebben niet onderzocht hoe de leerlingen elkaars teksten beoordelen met gebruikmaking van de door Cito ontwikkelde, genrespecifieke beoordelingsschema's. Eerder onderzoek naar de kwaliteit van *peer evaluation* geeft echter weinig reden tot optimisme.

ren dat de genrespecifieke benadering een veelbelovend perspectief biedt op andersoortige beoordeling van schrijfvaardigheid: het kan leiden tot grotere beoordelaarsovereenstemming.

### Zicht op centrale toetsing van schrijfvaardigheid?

Hoe zou het op basis van het verrichte onderzoek nu verder moeten met de centrale toetsing van schrijfvaardigheid in plaats van de huidige toetsing in de vorm van schoolexamens? In de eerste plaats lijkt het verstandig om het experiment met genrespecifiek beoordelen nog eens te herhalen in de vorm van hypothesetoetsend onderzoek. Met name kan – achteraf bezien – een aantal beoordelingsschema's nog wel specifiek worden geformuleerd voor de betreffende genres. Verder verdient het aanbeveling om docenten meer kennis te laten maken met de genrespecifieke benadering. Dit kan bijdragen aan meer diversiteit in het schrijfonderwijs en het deelvak aantrekkelijker maken voor de leerlingen.

Wanneer deze benadering verder zou worden uitgewerkt en hout blijkt te snijden, is het denkbaar dat de centrale toetsing van schrijfvaardigheid in iets andere vorm kan terugkeren. Het is denkbaar dat er op de scholen gedurende de bovenbouw-periode drie sessies worden belegd waarin de leerlingen de taak krijgen om drie verschillende genres te schrijven – die dan met genrespecifieke beoordelingsschema's worden beoordeeld. Zo wordt tegemoet gekomen aan de klacht dat niet-genrespecifieke analytische beoordelingsschema's kwalitatief onvoldoende zijn én worden leerlingen op meer dan één prestatie in het schrijfdomein beoordeeld. Het zou goed zijn als hierbij de schrijfopdrachten en de beoordelingsschema's centraal worden aangestuurd, zodat onwenselijke verschillen tussen scholen op dit vlak worden tegengegaan. Daarvoor is het nodig om eerst consensus te zoeken bij de docenten over de vraag welke genres zich het beste

lenen voor toetsing op deze manier. Liefst wordt dan gekozen voor genres die ook in de maatschappelijke context (in de praktijk van onderwijs of beroep) van nut zijn; de keuze voor een recensie of een creatief essay ligt dan minder voor de hand. Deze aanpak kan leiden tot gedeeltelijke centralisering: per schooltype worden dezelfde schrijfopdrachten inclusief beoordelingsschema's centraal aangeleverd, maar de afname vindt tijdens de schoolexamens plaats. Op die manier hoeft er geen aparte zitting in de examenperiode te worden belegd naast die voor Leesvaardigheid.

Tot slot is het denkbaar dat de hier bepleite gedeeltelijke centralisering een statusverhogend effect heeft, zodat er op scholen weer meer aandacht wordt gegeven aan het schrijven. Met name de in ons experiment gebruikte *peer review*, maar ook het bespreken van leerlingwerk in de klas is leerzaam en didactisch gezien van grote waarde, zoals door Schneiders in 1929 (!) al anekdotisch is beschreven (zie Figuur 2); daarbij beperkt de docent zich liefst niet tot grammaticale kwesties zoals door Schneiders benoemd. Door klassikale bespreking van opvallende tekstkenmerken hoeft meer oefenen met schrijven in de klas niet te leiden tot een substantiële taakverzwaring bij de docent. Bij de schrijfopdrachten die meetel-

En niet alleen het leesboek bewijst me goede dienst voor eigen-taalonderwijs en voor dat van mijn verwante collega's, ook het z.g. opstel levert daartoe uitstekend materiaal. Elke twee- weken weer (zolang m'n eerste klas 16 leerlingen telt; krijg ik er volgend jaar een van 30 dan maandeliks!) heb ik gelegenheid op fouten (en juist gebruik) te wijzen, op misslagen begaan tegen de woordorde, tegen interpunctie, tegen de vervoeging van het werkwoord (hebben of zijn, kwestie van aspekt), tegen modaliteit en fleksie. Ik laat ze voelen door vergelijking met analoge gevallen, dat het "getal" van dit werkwoord afhangt van het zo ver daar-van-af staande meervoudige substantief, dat ze dus "wilde*n*" moeten spellen.

Figuur 2. Overgenomen uit A.J. Schneiders: *Ons Taal- en Talenonderwijs*, p. 113-127. In *Paedagogische Studiën*, Groningen, Den Haag, 1929.



len voor het examencijfer moeten terwille van de objectiviteit van de beoordeling wel twee docenten bij de beoordeling zijn betrokken, maar dat is niet anders dan nu gebruikelijk is bij examenwerk. ■

## Meer weten

Op de website van Cito zijn de voor het onderzoek ontwikkelde schrijfpoddrachten en beoordelingsmodellen terug te vinden. <https://www.cito.nl/onderwijs/voortgezet-onderwijs/centrale-examens-voortgezet-onderwijs/expertise-van-cito/onderzoek>

## Literatuur

- Advies examen Nederlands (juli 2018, versie 2, met draagvlakpeiling). Nederlands Nu! Sectiebestuur Nederlands Levende Talen. Utrecht (CvTE).
- Bonset, H. (2010). Nederlands in voortgezet en hoger onderwijs: hoe sluit dat aan? Deel 2. *Levende Talen Magazine*, 97(4), 4-8.
- Bonset, H., Braaksma, M. (2008). Het schoolvak Nederlands opnieuw onderzocht. Een inventarisatie van onderzoek van 1997 tot en met 2007. Enschede: SLO.
- Bouwer, R. & Van den Bergh, H. (2015). Toetsen van schrijfvaardigheid: hoeveel beoordelaars, hoeveel taken? *Levende Talen Tijdschrift*, 16(3), 3-12.
- Bouwer, R., & Koster, M. (2016). Bringing writing research into the classroom. The effectiveness of Tekster, a newly developed writing program for elementary students. University of Utrecht, Utrecht.
- Coertjens, L., Lesterhuis, M., Verhavert, S., Van Gasse, R., & De Maeyer, S. (2017). Teksten beoordelen met criterialijsten of via paarsgewijze vergelijking: een afweging van betrouwbaarheid en tijdsinvestering. *Pedagogische Studiën*, 94, 283-303.
- Goosen, H. (2001). De inrichting van het nieuwe examen schrijfvaardigheid. *Levende Talen Tijdschrift*, 2(2), 24-33.
- Kuhlemeier, H., Boom, S., Gitsels, H., Schuurs, U. (2017). Maken de toevoegingen van de vakverenigingen op de correctievoorschriften het CSE VO betrouwbaarder? *Examens*, 4, 7-15.
- Meestringa, T., & Ravesloot, C. (2013). Het schoolexamen Nederlands havo/vwo in kaart. *Levende Talen Magazine*, 14(2), 11-19.
- Meestringa, T., & Ravesloot, C. (2014). Schrijven in het voortgezet onderwijs / Een stand van zaken. Het schoolexamen Nederlands havo/vwo in kaart. *Levende Talen Magazine*, 15(8), 22-27.
- Rooijackers, P. (2007). "We zetten in op schrijfvaardigheid" / Schrijven waarschijnlijk terug in het eindexamen Nederlands. *Levende Talen Magazine*, 94(5), 5-7.
- Schneiders, A. J. (1929). Ons Taal- en Talenonderwijs. *Paedagogische Studiën*, 113-127.
- Schreurs, Z., Kuhlemeier, H., Schuurs, U. (2018). Het meten van tekststructurele vaardigheden. *Levende Talen Tijdschrift*, 29, 15-25.
- Schuurs, U., de Groot, E. Genre-analyse. In J. Karreman en R. van Enschot (red.), *Tekstanalyse* (p. 239 – 291). Assen, Van Gorkum, 2018.
- Schuurs, U., Verhoeven, L. (2011). Taaltoetsing in vmbo en mbo. *Levende Talen Tijdschrift*, 12,4, 38-48.
- Sluijsmans, D.M.A., Joosten-ten Brinke, D., & Van der Vleuten, C.P.M. (2013). Toetsen met leerwaarde. Een reviewstudie naar de effectieve kenmerken van formatief toetsen. Den Haag: NWO.
- Vis, M. 2019. Genre-based pedagogy – interconnection between reading and writing. Posterpresentatie JURE/EARLI, augustus 2019.