

## Peerfeedback geven en gebruiken voor revisie;

### Een vergelijking tussen twee peerfeedbackmethoden

JANNEKE STUULEN, RENSKÉ BOUWER & HUB VAN DEN BERGH

In deze studie is onderzocht hoe leerlingen peerfeedback geven met verschillende beoordelingsmethoden, en hoe zij deze peerfeedback gebruiken voor het reviseren van hun eigen tekst. De ene groep leerlingen gaf feedback aan de hand van paarsgewijs vergelijken (PV; n = 22), de andere groep met een analytisch beoordelingsformulier (BF; n = 23). Uit de resultaten blijkt dat leerlingen uit de BF-conditie meer feedback geven op inhoudelijke aspecten dan leerlingen uit de PV-conditie, en dat zij deze feedback ook vaker verwerken. Er is echter geen effect van peerfeedback-conditie op tekstkwaliteit: in beide condities is de tekstkwaliteit na het herschrijven in dezelfde mate verbeterd.

We schrijven vaak en veel: e-mails, blogs, zakelijke brieven of berichten via sociale media. In de huidige kenniseconomie is het van groot belang complexe voorstellen en beslissingen duidelijk en coherent schriftelijk te verwoorden. Bovendien is een goede schrijfvaardigheid één van de belangrijkste voorspellers van het schoolsucces (Graham & Perin, 2007). Leraren zijn zich bewust van het belang van schrijven. Uit het voortgezet onderwijs komen echter signalen dat er wei-

nig tijd en aandacht is voor schrijfonderwijs (Elving, 2019). Zo wijst de Vereniging van Leraren in Levende Talen erop dat, door de nadruk op leesvaardigheid in het centraal examen, er naar verhouding veel onderwijstijd gaat naar lezen (Heemskerk, 2013). Dit gaat ten koste van het schrijfvaardigheidsonderwijs (Nederlandse Taalunie, 2015). Het voortgezet onderwijs moet daarom meer aandacht besteden aan de manier waarop leerlingen de kwaliteit van hun eigen teksten evalueren en verbeteren. In deze studie is onderzocht hoe leerlingen met behulp van verschillende peerfeedbackmethoden kritisch naar hun eigen en andermans teksten leren kijken en hoe leerlingen hiermee hun eigen schrijfproducten kunnen verbeteren.

#### Reviseren op basis van peerfeedback

Ondanks alle inspanningen slagen docenten Nederlands er niet voldoende in om hun leerlingen goede teksten te laten schrijven (Elving, 2019). Schrijven is complex en een lastig aan te leren vaardigheid (Flower & Hayes, 1981). Het is een proces dat het werkgeheugen enorm kan belasten.

Een manier om die cognitieve belasting

te beperken is het opdelen van het schrijfproces in afzonderlijke stappen zoals plannen, formuleren en reviseren (Kellogg, 2008). Uit diverse onderzoeken is gebleken dat een aparte revisiefase de tekstkwaliteit ten goede komt (Cho & MacArthur, 2010). Bovendien blijkt dat reviseren op basis van feedback nóg betere teksten oplevert (Braaksma et al., 2001). Feedback biedt schrijvers de mogelijkheid om vanuit het perspectief van de lezer inzicht te krijgen in de mate waarin het beoogde schrijfdoel is behaald en biedt concrete aanknopingspunten voor revisie (Koster et al., 2015; Rijlaarsdam, 1986). Daarom is feedback een cruciaal ingrediënt van goed schrijfonderwijs (Koster et al., 2015).

Feedback hoeft niet enkel van docenten te komen. Uit een recente meta-analyse (Double et al., 2019) blijkt dat ook peerfeedback effectief kan zijn voor schrijfprestaties. Door in de huid te kruipen van de lezer en zelf te ontdekken wat goede teksten kenmerkt (Rijlaarsdam, 1986), leren leerlingen schrijfdoelen beter te internaliseren (Van den Berg, 2003). Ook ontwikkelen leerlingen evaluatievaardigheden, die nodig zijn om de kwaliteit van hun eigen werk in te schatten en te reguleren (Nicol & Macfarlane-Dick, 2006; Tai et al., 2017). Door peerfeedback wordt de natuurlijke communicatielijntussen schrijver en lezer hersteld, een lijn die ontbreekt wanneer de leerling solitair teksten schrijft (Elving, 2019). Bovendien verwoorden klasgenoten hun feedback in taal die aansluit bij de taal van de leerlingen, zodat ook het herschrijven makkelijker zou moeten gaan (Rijlaarsdam, 1986). Het verwerken van feedback van verschillende medeleerlingen maakt een revisieopdracht levensechter en uitdagender (Bonset, 2007, p.6).

Het effect van peerfeedback lijkt echter wel afhankelijk van het soort feedback dat gegeven wordt. Leerlingen richten zich bij het geven van feedback voornamelijk op lagere-orde-aspecten in plaats van hogere-orde-aspecten van de tekst (Van Steendam et

al., 2010). Lagere-orde-feedback (LO) heeft betrekking op het microniveau van de tekst zoals spelling, interpunctie, grammatica en lay-out, terwijl hogere-orde-feedback (HO) op het macroniveau gericht is en gaat over inhoud, structuur of stijl (Faigley & Witte, 1981). Juist die HO-feedback is nodig om leerlingen in staat te stellen hun teksten te verbeteren (Bogaerds et al., 2017; Patchan et al., 2016). Het lijkt erop dat leerlingen voor het geven van effectieve peerfeedback op HO-aspecten van de tekst extra ondersteuning en instructie van hun docent moeten krijgen (Hoogeveen & Van Gelderen, 2013).

#### Beoordelingsformulieren versus paarsgewijs vergelijken

Om leerlingen te ondersteunen bij het beoordelen en analyseren van de kwaliteit van hun eigen en/of andermans schrijfproducten kunnen docenten Nederlands rubrics of beoordelingsformulieren gebruiken. Op deze manier proberen docenten leerlingen bewust te maken van de verschillende aspecten van tekstkwaliteit. Bij deze analytische manier van beoordelen geven leerlingen geen feedback op de tekst als geheel, maar richten ze zich op in het beoordelingsmodel gespecificeerde tekstkenmerken. Dat kan zijn weerslag op leerlingen hebben: ze krijgen weliswaar feedback over criteria waar ze goed of minder goed op scoren, maar leren ze ook iets over de communicatieve effectiviteit van hun tekst als geheel?

Een alternatieve methode voor het geven van peerfeedback is paarsgewijs vergelijken (Bouwer, Goossens et al., 2018). Hierbij vergelijken leerlingen het werk van medeleerlingen met elkaar in willekeurig samengestelde paren en leren ze naar de kwaliteit van teksten als geheel te kijken. Paarsgewijs vergelijken gaat uit van het principe dat mensen beter zijn in het vergelijken van twee producten

dan in het beoordelen van één product los van de rest (Laming, 2004). Bij elk paar geeft de leerling aan welk werk van hogere kwaliteit is. Alle vergelijkingen samen leveren uiteindelijk betrouwbare kwaliteitsoordelen op (Verhavert et al., 2019). Na elke vergelijking kunnen leerlingen feedback geven op de teksten. Hoe meer beoordelaars dat doen, hoe rijker de feedback is en hoe beter de schrijver kan inschatten hoe de tekst overkomt op verschillende lezers. Leerlingen zien verschillende teksten langskomen en leren daardoor goede en minder goede teksten van elkaar te onderscheiden.

Paarsgewijs vergelijken heeft recentelijk zijn weg in het hoger onderwijs gevonden, waar het al veelvuldig gebruikt wordt om op een betrouwbare en valide manier peerfeedback te geven op tekstkwaliteit (Bouwer, Goossens et al., 2018). Ook blijkt uit onderzoek dat studenten bij paarsgewijs vergelijken van teksten zich voornamelijk richten op hogere-orde-aspecten (Bouwer, Lesterhuis et al., 2018). Maar we weten niet of dit ook voor leerlingen in het voortgezet onderwijs geldt. Hebben zij voldoende kennis van tekstkwaliteit om teksten goed met elkaar te vergelijken? Of hebben zij hier meer sturing bij nodig in de vorm van analytische beoordelingsformulieren? Verder zijn studies naar de effectiviteit van peerfeedback met name uitgevoerd in de bovenbouw van de basisschool (Hoogeveen & Van Gelderen, 2015; Bouwer & Koster, 2016) of de onderbouw van het voortgezet onderwijs (Rijlaarsdam 1986; Denneman, et al 2020). Het is tot op heden onbekend hoe leerlingen in de bovenbouw peerfeedback geven en gebruiken voor revisie.

### Onderzoeksvragen

In dit onderzoek kijken we naar de verschillen in peerfeedback tussen twee beoordelingsmethoden: paarsgewijs vergelijken en

analytisch beoordelen. Daarbij proberen we een antwoord te vinden op onderstaande onderzoeksvragen:

1. In hoeverre geven leerlingen feedback op hogere- of lagere-orde aspecten van elkaars teksten?
2. In hoeverre gebruiken leerlingen peerfeedback bij het reviseren van hun tekst?
3. In hoeverre is er een relatie tussen het verwerken van feedback en de kwaliteit van de herschreven teksten?

Op basis van eerdergenoemd onderzoek verwachtten we dat leerlingen aan de hand van paarsgewijs vergelijken meer HO-feedback geven en gebruiken bij het reviseren van hun teksten, dan leerlingen die peerfeedback geven aan de hand van analytische beoordelingsformulieren. Ook verwachten we dat een toename in het gebruik van peerfeedback op HO-aspecten leidt tot betere herschreven teksten.

### Methode

#### Participanten

Het onderzoek is uitgevoerd in twee klassen van één docent met in totaal 55 havo-5-leerlingen van één school. Op het moment van de dataverzameling waren er 10 leerlingen absent. Dit heeft geresulteerd in een klas die peerfeedback gaf door middel van paarsgewijs vergelijken (PV-conditie;  $n = 22$ ) en een klas die een beoordelingsformulier heeft gebruikt om elkaar van feedback te voorzien (BF-conditie;  $n = 23$ ). De klassen zijn random toegewezen aan een van de twee condities.

#### Materiaal en procedure

Het onderzoek werd uitgevoerd binnen zes lessen van 60 minuten. De opdracht voor beide groepen was om een motivatiebrief voor een vervolgopleiding te schrijven. Deze opdracht is onderdeel van het PTA (Programma van

Toetsing en Afsluiting) zakelijk schrijven. Tijdens de eerste les ontvingen leerlingen een syllabus met theorie over zakelijke brieven, een voorbeeldbrief en een voorbeeld van een curriculum vitae (cv). De docent instrueerde leerlingen over genrekenmerken van een motivatiebrief en criteria voor tekstkwaliteit, en lichtte per alinea de briefconventies toe. Een les later verzamelden de leerlingen individueel informatie over vervolgopleidingen. Leerlingen kozen zelf drie opleidingen uit waar ze de motivatiebrief voor wilden schrijven. Tijdens de derde les liet de docent een voorbeeld-cv zien en maakten leerlingen hun eigen cv.

Het schrijven van de eerste versie van de motivatiebrief gebeurde de volgende les digitaal, binnen een lesuur. Leerlingen maakten hierbij gebruik van de verschillende bronnen die ze verzameld hadden voor hun vervolgopleiding. Ook hun cv hielden ze erbij.

In de vijfde les moesten leerlingen feedback geven op elkaars teksten. De leerlingen in de PV-conditie gaven peerfeedback op basis van paarsgewijs vergelijken. Dit gebeurde aan de hand van een online tool, *Comproved* (<https://comproved.com/>). Leerlingen kregen op hun scherm steeds twee willekeurige teksten van hun medeleerlingen te zien die zij vergeleken. Na elke vergelijking gaven leerlingen aan waarom ze de ene tekst beter vonden dan de andere, aan de hand van drie tips en drie tops per tekst. Elke leerling maakte binnen het lesuur tien vergelijkingen en gaf dus feedback op twintig teksten. De feedback ontvingen leerlingen overzichtelijk in een schema met alle tips en tops op hun schrijfproduct onder elkaar. Deze peerfeedback was anoniem. Voor een voorbeeld van peerfeedback met paarsgewijs vergelijken, zie Stuulen (2022).

De leerlingen uit de BF-conditie gaven feedback aan de hand van een beoordelingsformulier (zie bijlage 1). Dit gebeurde niet digitaal, maar op papier. Per alinea schreven

leerlingen op het beoordelingsformulier aan welke inhoudelijke eisen de brief voldeed. Op de achterkant van het formulier checkten de leerlingen de onderdelen briefconventies, layout, stijl en spelling. Ook moesten leerlingen na het invullen van het beoordelingsformulier drie algemene tips en tops noteren voor de feedbackontvanger. Met deze analytische beoordelingsmethode konden de leerlingen gedurende een lesuur maar één brief van feedback voorzien.

De volgende les herschreven leerlingen hun teksten op basis van de feedback van hun klasgenoten. In beide condities gaven leerlingen vervolgens aan welke feedbackpunten zij mee wilden nemen bij het herschrijven van de brief. Dit noteerden ze op een apart formulier en leverden ze na het herschrijven bij de docent-onderzoeker in.

#### Coderen peerfeedback

Voor de analyse zijn verschillende stappen gezet. Eerst is de inhoud van de gegeven tips en tops per leerling gecodeerd (tabel 1). Hierbij is een onderscheid gemaakt tussen de hogere-orde-kenmerken inhoud, structuur en stijl, en de lagere-orde-kenmerken grammatica, interpunctie, lay-out/conventies, spelling. Dit onderscheid geeft inzicht in de aspecten waarop leerlingen peerfeedback geven. Een aantal opmerkingen is gecodeerd als Algemene Opmerking (AO). Hierbij is het commentaar niet gericht op hogere- of lagere aspecten van de tekst, maar bestond het uit algemene opmerkingen.

Om antwoord te krijgen op de vraag hoe leerlingen de peerfeedback gebruiken bij het reviseren van hun tekst hebben we alleen naar de tips gekeken. De tops hebben we buiten beschouwing gelaten, want die vragen immers niet om revisie. Hierbij hebben we ons tot de relevante feedback beperkt, dat wil zeggen, feedback die leerlingen ook echt kunnen verwerken. Een deel van de feedback van de tips in beide condities bleek namelijk

Type feedback	Voorbeeld
<b>HO</b>	Inhoud <i>Je noemt heel duidelijk de argumenten waarvoor je geschikt bent voor deze studie</i>
	Structuur <i>Alinea's zijn niet heel duidelijk te onderscheiden</i>
	Stijl <i>Het lijkt alsof je er een woordenboek bijgehouden hebt, komt niet echt over</i>
<b>LO</b>	Grammatica <i>Let op je grammatica</i>
	Interpunctie <i>Havo moet met kleine letters</i>
	Lay-out/conventies <i>Ziet er niet netjes uit</i>
	Spelling <i>Het is SFEER in plaats van zweer</i>
<b>AO</b>	Algemene opmerkingen <i>Zeg je eerlijk het is gewoon niet goed</i>

Tabel 1. Codering en voorbeeld peerfeedback (HO: feedback op hogere-orde-aspecten; LO: feedback op lagere-orde-aspecten).

irrelevant (11%) en is dus niet meegenomen in de analyse. Voorbeelden hiervan zijn: 'Noem echt een datum', terwijl de datum al duidelijk vermeld stond. Of 'Vervolgstudie ontbreekt', terwijl de vervolgstudie duidelijk genoemd werd in de brief. Voor elke leerling is vervolgens gescoord hoeveel van elk type feedback wel of niet is verwerkt in de tweede versie van de tekst.

**Bepalen van tekstkwaliteit**

Om na te gaan wat het effect van peerfeedback is op tekstkwaliteit, is de kwaliteit van zowel de eerste als de tweede versie van de motivatiebrieven beoordeeld door drie beoordelaars met ten minste tien jaar ervaring in het beoordelen van teksten. Zij beoordeelden teksten aan de hand van een ankeropstelschaal: een reeks voorbeeldopstellen van oplopende kwaliteit. De anker teksten zijn in een eerder stadium geselecteerd uit alle te beoordelen teksten. Dit gebeurde door dezelfde beoordelaars op basis van een impressie-analyse, zie

Pollman et al., 2012. De middelste anker tekst is een gemiddelde tekst met een arbitraire score van 100 punten, de andere anker teksten liggen een standaarddeviatie boven (115 punten) en een standaarddeviatie onder het gemiddelde (85 punten). De beoordelaars vergeleken de te beoordelen teksten met deze drie anker teksten en kwamen zo tot een holistisch oordeel over de kwaliteit van de tekst. De schaal is continu, wat betekent dat er ook scores gegeven konden worden tussen de ankerpunten in. Deze procedure leidt vaak tot een hoge overeenstemming tussen beoordelaars (Bouwer et al., 2016). Ook in dit onderzoek bleek de overeenstemming tussen de drie onafhankelijke beoordelaars hoog ( $\alpha = 0,93$ ). De gemiddelde tekstkwaliteitscores van de drie beoordelaars zijn gebruikt voor verdere analyse. Met een anova voor herhaalde waarnemingen is nagegaan wat het effect is van het reviseren na peerfeedback in beide condities op de uiteindelijke tekstkwaliteit.

**Resultaten**

*Hoeveelheid en type feedback*

De leerlingen hebben in totaal 470 commentaren aan elkaar gegeven. Zes commentaren waren niet gericht op hogere- of lagere-orde aspecten, maar betroffen meer algemene opmerkingen, zoals: 'Ik denk dat je het echt heel goed hebt gedaan', 'Voldoet aan alle eisen' of 'Zeg je eerlijk is gewoon niet goed'. Deze vorm van feedback is niet relevant voor het onderzoek en is dan ook niet verder geanalyseerd. De rest van de commentaren was gericht op HO-aspecten (n = 272) en

LO-aspecten (n = 192). Gemiddeld kreeg elke leerling op tien punten commentaar. Er was een grote variatie in de hoeveelheid feedback die leerlingen elkaar gaven: het aantal varieerde van 0 tot 25 commentaren per tekst.

In tabel 2 is de HO- en LO-feedback opgesplitst in tips en tops per categorie. Het aantal tips en tops verschilde tussen condities ( $\chi^2(1) = 115, p < 0,05$ ). Opvallend is dat er in de PV-conditie relatief meer tops gegeven zijn dan in de BF-conditie, respectievelijk 168 en 13 tops ( $\chi^2(1) = 132,7, p < 0,01$ ). Het gaat hier om commentaar dat met name gericht is op inhoud zoals: 'Duidelijke informatie' of,

TYPE FEEDBACK	PV conditie				BF conditie				Toetsing	
	Tops		Tips		Tops		Tips		$\chi^2(1)$	p
	n	%	n	%	n	%	n	%		
<b>Hogere Orde (HO)</b>										
Inhoud	116 <sup>a</sup>	25,0	32 <sup>b</sup>	6,9	6	1,3	58	12,5	74,8	<0,01
Structuur	21 <sup>a</sup>	4,5	5	1,1	1	0,2	8	1,7	13,9	<0,01
Stijl	9 <sup>a</sup>	1,9	6	1,3	2	0,4	8	1,7	3,9	0,04
<b>Totaal</b>	<b>146<sup>a</sup></b>		<b>43<sup>b</sup></b>		<b>9</b>		<b>74</b>		<b>103,8</b>	<b>&lt;0,01</b>
<b>Lagere orde (LO)</b>										
Grammatica	2	0,4	3 <sup>b</sup>	0,7	0	0,0	17	3,7	7,5	<0,01
Interpunctie	0	0,0	3 <sup>b</sup>	0,7	0	0,0	11	2,4	4,1 <sup>c</sup>	0,03
Spelling	0	0,0	2 <sup>b</sup>	0,4	0	0,0	34	7,3	28,4 <sup>c</sup>	<0,01
Lay-out	20 <sup>a</sup>	4,3	72 <sup>a</sup>	15,5	4	0,9	24	5,2	0,7	0,39
<b>Totaal</b>	<b>22<sup>a</sup></b>		<b>80</b>		<b>4</b>		<b>86</b>		<b>12,0</b>	<b>&lt;0,01</b>
<b>Totaal</b>	<b>168<sup>a</sup></b>	<b>36,2</b>	<b>123<sup>b</sup></b>	<b>26,5</b>	<b>13</b>	<b>2,8</b>	<b>160</b>	<b>34,5</b>	<b>115,0</b>	<b>&lt;0,01</b>

a: PV > BF (p < .05); b: PV < BF (p < .05); c: alleen vergelijking aantal tips.

Tabel 2. Aantal tops en tips uitgesplitst per beoordelingsmethode en type feedback

'Jouw kwaliteiten worden goed uitgelegd'. In de PV-conditie werden daarentegen minder tips (n = 123) gegeven dan in de BF-conditie (n = 160;  $\chi^2(1) = 4,8, p = 0,02$ ). Dit verschil zit voornamelijk in het aantal HO-tips: 43 tips in de PV-conditie tegenover 74 tips in de BF-conditie. Er is nauwelijks verschil in de hoeveelheid LO-tips tussen beide condities: 80 tips in de PV-conditie tegenover 86 tips in de BF-conditie. In de BF-conditie was het leeuwendeel aan feedback op inhoud gericht (n = 58): 'Je moet specifieker zijn wat je met deze studie wilt bereiken'. Ook commentaar op grammatica ('Beter letten op je zinsbouw') en spelling werd veel gegeven. In de PV-conditie was het grootste deel van de tips (n = 72) gericht op lay-out: 'Briefhoofd ontbreekt'.

*Verwerken van feedback*  
Tabel 3 toont dat leerlingen meer feedback hebben verwerkt in de BF-conditie dan in de PV-conditie ( $\chi^2(1) = 5,5, p = 0,02$ ). We zien dat in de PV-conditie 67,5% van de feedback verwerkt is en in de BF-conditie 72,5%. Ook op HO-niveau verwerkten leerlingen meer feedback in de BF-conditie (73%) dan in de PV-conditie (53%). In de PV-conditie verwerkten leerlingen wel relatief vaker commentaar op lay-out-kwesties. Het gaat hierbij om commentaren als 'Je eigen gegevens moeten boven staan en die van de school daaronder', of 'Foute lay-out'. In de BF-conditie waren leerlingen het meest geneigd om feedback op grammatica en structuur te verwerken.

TYPE FEEDBACK	PV conditie		BF conditie		Toetsing	
	Wel verwerkt N (%)	Niet verwerkt N (%)	Wel verwerkt N (%)	Niet verwerkt N (%)	$\chi^2(1)$	p
<b>Hogere Orde (HO)</b>						
Inhoud	17 (53,1)	15 (46,8)	41 (70,7)	17 (29,3)	9,9 <sup>a</sup>	<.01
Structuur	3 (60,0)	2 (40,0)	7 (87,5)	1 (12,5)	1,6	0,21
Stijl	3 (50,0)	3 (50,0)	6 (75,0)	2 (25,0)	1,0	0,32
<b>Lagere orde (LO)</b>						
Grammatica	0 (0,0)	3 (100,0)	16 (94,1)	1 (5,9)	16,0 <sup>a</sup>	<.01
Interpunctie	2 (66,7)	1 (33,3)	5 (45,5)	6 (54,5)	1,3	0,26
Spelling	0 (0,0)	2 (100,0)	25 (73,5)	9 (26,5)	25,0 <sup>a</sup>	<.01
Lay-out / conventies	58 (80,6)	14 (19,4)	16 (66,7)	8 (33,3)	23,8 <sup>b</sup>	<.01
<b>Totaal</b>	<b>83 (67,5)</b>	<b>40 (32,5)</b>	<b>116 (72,5)</b>	<b>44 (27,5)</b>	<b>5,5<sup>a</sup></b>	<b>0,02</b>

a: PV > BF (p < .05); b: PV < BF (p < .05)

Tabel 3. Frequenties en percentages van verwerkte – en niet verwerkte feedback-elementen per conditie

CONDITIE	Versie 1		Versie 2	
	Gemiddelde score	SD	Gemiddelde score	SD
PV	93,3	16,5	99,5	13,6
BF	89,2	20,2	96,5	18,3
Totaal	91,2	18,4	98,0	16,1

Tabel 4. Gemiddelde tekstkwaliteit (en standaarddeviatie) per meetmoment en per conditie (N= 45)

**Tekstkwaliteit**

Gemiddeld steeg de tekstkwaliteit tussen de eerste en tweede versie met bijna 7 punten, zie tabel 4. Dit verschil in kwaliteit tussen beide versies is significant (F (1, 44) = 18,43, p < 0,001, d = 0,41). Maar het interactie-effect tussen conditie en versie bleek niet significant (F (1, 44) = 0,07, p = 0,79); in beide condities is de vooruitgang tussen versie 1 en versie 2 dus even groot.

Het verschil in tekstkwaliteit tussen voor- en nameting is positief gerelateerd aan het aantal ontvangen tips (r = 0,29). Het aantal tips is noch gerelateerd aan de tekstkwaliteit op de voormeting, noch aan de tekstkwaliteit op de nameting. Het verschil tussen beide versies bleek wel gerelateerd aan de hoeveelheid feedback die de leerling verwerkt had (r = 0,39, p = 0,02). Hoe meer feedback verwerkt was, hoe beter de tekst was herschreven. Met name commentaar gericht op structuur bleek een significant effect te hebben op de uiteindelijke tekstkwaliteit (r = 0,31, p = 0,03). Het verwerken van commentaar als: 'Je moet een extra alinea toevoegen' of 'Alinea's zijn niet heel duidelijk te onderscheiden' leverde een verbeterde tweede versie op. Commentaar op LO-aspecten bleek niet gerelateerd te zijn aan de kwaliteit van de herschreven tekst (r < 0,27, p > 0,07).

**Conclusie en discussie**

In deze studie hebben we in twee havo-5-klassen onderzocht hoe leerlingen feedback geven op elkaars teksten in twee condities: paarsgewijze vergelijking (PV) en beoordelingsformulier (BF). Ook is nagegaan welke feedback leerlingen gebruiken bij het herschrijven en welke niet. Daarnaast is de relatie tussen het verwerken van de feedback en de uiteindelijke tekstkwaliteit nader geanalyseerd.

Voor het geven van peerfeedback blijkt er een verschil tussen beide condities te zijn in de mate waarin leerlingen feedback op sterke dan wel minder sterke aspecten geven. Leerlingen in de PV-conditie hebben meer tips gegeven dan leerlingen in de BF-conditie; deze tips waren met name gericht op HO-aspecten van elkaars teksten zoals inhoud en structuur. Leerlingen in de BF-conditie hebben daarentegen meer tips gegeven dan leerlingen in de PV-conditie, zowel voor HO- als LO-aspecten. De tips van leerlingen in de PV-conditie waren vooral gericht op LO-aspecten.

De resultaten laten ook verschillen zien tussen condities wat betreft het gebruik van peerfeedback voor het reviseren van de eigen tekst. Zo verwerkten leerlingen in de PV-conditie vaker lay-outkwesties,

terwijl leerlingen in de BF-conditie vaker grammatica- en spellingkwesties verwerkten. Verder zien we dat leerlingen uit de BF-conditie relatief meer feedback op inhoud hebben verwerkt dan leerlingen in de PV-conditie.

Ondanks de verschillen in het geven en gebruiken van peerfeedback, blijkt er geen effect te zijn van conditie op de kwaliteit van de herschreven teksten. In beide condities is de tekstkwaliteit na het herschrijven in dezelfde mate verbeterd. Wel laten de resultaten zien dat hoe meer feedback er verwerkt wordt, hoe beter de uiteindelijke tekst geschreven wordt.

Concluderend kunnen we stellen dat de beoordelingsmethode een effect heeft op de manier waarop leerlingen peerfeedback geven en gebruiken. Leerlingen lijken met een analytisch beoordelingsformulier meer inhoudelijke feedback te geven en te verwerken dan leerlingen die peerfeedback geven aan de hand van paarsgewijs vergelijken. Toch blijken beide peerfeedbackmethodes even effectief te zijn voor de kwaliteit van herschrijven.

Hoe kan het dat leerlingen in beide groepen hun tekst even goed konden herschrijven, ondanks dat leerlingen uit de BF-conditie meer inhoudelijke feedback kregen van hun peers? Een verklaring hiervoor zou kunnen zijn dat leerlingen in de PV-conditie vanwege de vele positieve opmerkingen over de inhoud van hun teksten gemotiveerder waren om hun tekst te herschrijven, los van andere specifieke peerfeedbackcommentaren. Eerder onderzoek laat namelijk zien dat positieve feedback leidt tot meer motivatie (Hattie & Timperley, 2007; Underwood & Tregidgo, 2010). Vervolgonderzoek zou moeten uitwijzen in hoeverre tops tot meer revisie aansporen. Een andere verklaring zou kunnen zijn dat leerlingen in de PV-conditie veel verschillende teksten langs zien komen, waardoor ze een beter beeld hebben gekre-

gen van wat tekstkwaliteit is (Bouwer, Lesterhuis, et al., 2018). Deze kennis kunnen ze meenemen bij het herschrijven van hun eigen tekst, waardoor ze die op een hoger niveau kunnen herschrijven dan leerlingen uit de BF-conditie, die bij het herschrijven meer afhankelijk zijn van de gegeven peerfeedback. Verder onderzoek is daarom nodig naar wat leerlingen leren van het zien en vergelijken van meerdere teksten en in hoeverre dit het proces van herschrijven kan beïnvloeden.

Een ander opmerkelijk resultaat in dit onderzoek zijn de grote aantallen peerfeedbackcommentaren op lay-out in de PV-conditie. Leerlingen zien de twee brieven voor zich en reageren op duidelijk in het oog lopende verschillen. Dat uit zich in opmerkingen als 'korte brief'. Deze bevinding is niet in lijn met eerder onderzoek waaruit juist blijkt dat beoordelaars met de methode van paarsgewijs vergelijken vooral op hogere-orde-aspecten letten (Lesterhuis, 2018). Wellicht helpt het leerlingen om voor het geven van peerfeedback al te oefenen met paarsgewijs vergelijken zodat ze een reeks voorbeelden gezien hebben en weten waarop ze moeten letten bij elkaars werk (Bouwer, Lesterhuis, et al., 2018). Ook is gebleken dat met de methode van paarsgewijs vergelijken de beoordelingen van de leerlingen sterk correleerden met de beoordelingen van ervaren beoordelaars ( $r = 0,71$ ). Dit suggereert dat leerlingen de kwaliteit van elkaars werk goed kunnen beoordelen, maar dat ze wellicht nog moeite hebben om dit goed onder woorden te brengen.

Opgemerkt moet worden dat deze studie slechts op één school in twee klassen havo 5 is uitgevoerd. Bovendien is binnen deze studie gebruik gemaakt van slechts één genre, namelijk de zakelijke brief. Vervolgonderzoek met meer leerlingen en andere schrijftaken moet uitwijzen in hoeverre de gerapporteerde resultaten

generaliseerbaar zijn. Ook moet bij dergelijk onderzoek voldoende aandacht besteed worden aan de sociale relaties tussen leerlingen. Leerlingen zijn snel bang om te kritisch te zijn in hun beoordeling en een klasgenoot van streek te maken (Cartney, 2010). Onderzoekers wijzen daarom op het belang van training en ondersteuning bij het geven en ontvangen van peerfeedback (Van den Berg, et al 2003; Sluijsmans et al 2002). Kwalitatief goede peerfeedback moet daarnaast verder gaan dan het enkel aanstippen van sterke en minder sterke aspecten in de tekst, ze moet duidelijk maken hoe de schrijver deze aspecten kan versterken en verwerken (Hattie & Timperley, 2007). Zo zagen we dat de commentaren in de PV-conditie korter waren dan in de BF-conditie: "te weinig info", "brief is niet af", "veel herhaling" versus "Je moet specifieker zijn wat je met deze studie wil bereiken" of "Je persoonlijke kenmerken staan er niet duidelijk genoeg in". Wellicht is dat ook een reden dat leerlingen in de BF-conditie de inhoudelijke feedback meer verwerkten: er werd meer uitleg gegeven en de commentaren bevatten concretere tips voor verbetering dan de kortere PV-feedback.

Dit onderzoek laat zien dat herschrijven na peerfeedback werkt ongeacht de beoordelingsmethode. Met name het aantal HO-commentaren kan een positief verband aantonen. Daarom is het goed om het onderwijs vooral te richten op die HO-feedback, want daarmee kunnen leerlingen hun teksten goed herschrijven. Een instructie vooraf in de vorm van feedbacklessen waarbij leerlingen oefenen met het geven en verwerken van peerfeedback op HO-tekstkenmerken lijkt daarbij onontbeerlijk. Als leerlingen hun feedback met name richten op HO-feedback, is de kans groter dat klasgenoten iets veranderen aan de inhoud of structuur van hun tekst en op die manier de tekstkwaliteit verbeteren.

#### LITERATUUR

- Berg, B. van den (2003). *Peer assessment in universitair onderwijs. Een onderzoek naar bruikbare ontwerpen*. Utrecht: Proefschrift Universiteit Utrecht.
- Bogaerds, S., Bouwer, R., Bergh, H. van den, & Evers-Vermeul, J. (2017). Daar maak ik geen punt van! *Levende Talen Tijdschrift*, 18(2), 21–29.
- Bonset, H. (2007). *Handreiking schoolexamen Nederlands havo/vwo, tweede fase*. Enschede: SLO.
- Bouwer, R., Goossens, M., Mortier, A. V., Lesterhuis, M., & De Maeyer, S. (2018). Een comparatieve aanpak voor peer assessment: Leren door te vergelijken. In D. Sluijsmans, & M. Segers (Eds.), *Toetsrevolutie: Naar een feedbackcultuur in het hoger onderwijs* (pp. 92–106). Culemborg: Phronese.
- Bouwer, R., & Koster, M. (2016). Bringing writing research into the classroom. *The Effectiveness of Tekster, a newly developed writing program for elementary students*. Utrecht: Proefschrift Universiteit Utrecht.
- Bouwer, R., Lesterhuis, M., Bonne, P., & De Maeyer, S. (2018). Applying criteria to examples or learning by comparison: Effects on students' evaluative judgment and performance in writing. *Frontiers in Education*, 3(86), 1–12.
- Braaksma, M., Bergh, H. van den, Rijlaarsdam, G., & Couzijn, M. (2001). Effective learning activities in observation tasks when learning to write and read argumentative texts. *European Journal of Psychology of Education*, 16(1), 33–48.
- Cartney, P. (2010). Exploring the use of peer assessment as a vehicle for closing the gap between feedback given and feedback used. *Assessment & Evaluation in Higher Education*, 35(5), 551–564.
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20(4), 328–338.

- Denneman, E., Bergh, H. van den, & Evers-Vermeul, J. (2020). Peerfeedback in het schrijfvaardigheidsonderwijs. *Levende Talen Tijdschrift*, 21(3), 25–36.
- Double, K., McGrane, J., & Hopfenbeck, T. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, 32, 481–509.
- Elving, K. (2019). *Effectieve leeractiviteiten voor het schrijfonderwijs in havo 4*. Utrecht: Proefschrift Universiteit Utrecht.
- Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, 32(4), 400–414.
- Flower, L., & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365–387.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445–476.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Heemskerk, K. (2012). Effecten van sturing op examenresultaten voor de kwaliteit van het taalonderwijs. In A. Mottart, & S. Vanhooren (Eds.), *Zevenentwintigste conferentie Onderwijs Nederlands* (pp.313–315). Academia Press.
- Hooegeven, M., & van Gelderen, A. (2013). What works in writing with peer response? A review of intervention studies with children and adolescents. *Educational Psychology Review*, 25(4), 473–502.
- Hooegeven, M., & van Gelderen, A. (2015). Effects of peer response using genre knowledge on writing quality: A randomized control trial. *The Elementary School Journal*, 116(2), 265–290.
- Kellogg, R. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1(1), 1–26.
- Koster, M., Tribushinina, E., De Jong, P., & Bergh, H. van den. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research*, 7(2), 249–274.
- Laming, D. (2004). *Human Judgement: The eye of the beholder*. Thomson Learning.
- Lee, I. (2008). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing*, 17(3), 144–164.
- Lesterhuis, M. (2018). *The validity of comparative judgement for assessing text quality*. Proefschrift. Universiteit Antwerpen.
- Nederlandse Taalunie. (2015). *Schrijfonderwijs in de schijnwerpers. Naar een betere schrijfvaardigheid van Nederlandse en Vlaamse leerlingen*. Nederlandse Taalunie.
- Nicol, D., & D. Macfarlane-Dick. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education* 31(2), 199–218.
- Patchan, M.M., Schunn, C.D., & Correnti, R.J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology*, 108(8), 1098–1120.
- Pollman, E., Prenger, J., Gloppe, K. de. (2012). Het beoordelen van leerlingteksten met behulp van een schaalmodel. *Levende Talen Tijdschrift*, 13(3), 15–24.
- Rijlaarsdam, G. (1986). *Effecten van leerlingrespons op aspecten van stelvaardigheid*. Proefschrift. Universiteit van Amsterdam.
- Sadler, D.R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159–179.
- Sluijsmans, D.M.A., Brand-Gruwel, S., & Van Merriënboer, J. (2002). Peer assessment training in teacher education. *Assessment and Evaluation in Higher Education*, 27(5), 443–454.
- Steendam, E. van, Rijlaarsdam, G., Sercu, L., & Bergh, H. van den. (2010). The effect of instruction type and dyadic or individual emulation on the quality of higher-order peer feedback in EFL. *Learning and Instruction*, 20(4), 316–327.
- Stuulen, J. (2022). Peerfeedback door middel van paarsgewijs vergelijken. *Didactiek Nederlands – Zo kan het ook*. Geraadpleegd via <https://didactieknederlands.nl/zokanhetook/2022/01/peerfeedback-door-middel-van-paarsgewijs-vergelijken/>.
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., Panadero, E. (2018). Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher Education*, 76(3), 467–481.
- Topping, K. (2009). Peer assessment. *Theory Into Practice*, 48(1), 20–27.
- Underwood, J.S., & Tregidgo, A.P. (2010). Improving student writing through effective feedback: Best practices and recommendations. *Journal of Teaching Writing*, 22(2), 73–98.
- Verhavert, S., Bouwer, R., Donche, V., & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy, & Practice*, 26(5), 541–562.

JANNEKE STUULEN is docent Nederlands aan het Amstelveen College. Dankzij een docentenpromotiebeurs van het NWO doet zij onderzoek naar peerfeedback bij schrijfvaardigheidsonderwijs in de bovenbouw havo en vwo. Ze richt zich hierbij op de effecten van paarsgewijs vergelijken en het herschrijven van teksten.  
E-mail: j.a.stuulen@uu.nl

RENKE BOUWER werkt als universitair docent Taal en Educatie aan de Universiteit Utrecht. Zij doet onderzoek naar onder andere kwaliteitsvolle manieren van toetsen en beoordelen van schrijfvaardigheid.  
E-mail: r.bouwer@uu.nl

HUUB VAN DEN BERGH is als hoogleraar Didactiek en Toetsing van het Taalvaardigheidsonderwijs werkzaam bij het Departement Talen, Literatuur en Communicatie aan de Universiteit Utrecht. Zijn specialiteit ligt op het gebied van onderzoek naar de effectiviteit van taalonderwijs.  
E-mail: h.vandenbergh@uu.nl

**BIJLAGE 1**

Beoordelingsformulier motivatiebrief

5H 2019-2020 Feedbackformulier ED-4	<b>Beoordelingsmodel:</b> (oefen)motiva- tiebrief		
	<b>Naam schrijver brief:</b>		
	<b>Degene die beoordeelt:</b> docent / klas- genoot	Max pnt.	Behaald aantal pnt.
<b>INHOUD</b>	<b>Naam klasgenoot:</b>		
Uitgangspunt	Documentatiemap in orde = 10 pnt	10	
<b>1e alinea: inleiding</b>			
• Wens tot toelating	1 pnt	<b>max 4 pnt</b>	
• met ingang van,	1 pnt		
• waar,	1 pnt		
• expliciet studierichting noemen.	1 pnt		
<b>2e alinea: kern 1</b>			
• Aankondiging, kort, waarom deze opleiding/studie.	2 pnt	<b>max 5 pnt</b>	
• Welke opleiding nu, profiel, keuzevakken, bijna klaar met school, graag aan vervolgstudie.	1 pnt		
• Geef groeimogelijkheden aan die je via studie zult krijgen en licht dat inhoudelijk toe.	2 pnt		
<b>3e alinea: kern 2</b>			
• Leg uitgebreid uit waarom je juist deze studie hebt gekozen (inhoudelijke argumenten).	3 pnt	<b>max 6 pnt</b>	
• Leg uit waarom je kiest voor deze instelling/universiteit (inhoudelijke argumenten).	3 pnt		

**4e alinea: kern 3**

Aan welke minimaal 2 belangrijke punten uit je informatiemap die geëist worden (noemen dus) denk jij te kunnen voldoen? Licht deze punten toe.

8 pnt

**max  
12 pnt.**

Wat maakt jou naar aanleiding van die punten dus geschikt voor deze studie? Dat leg je uitvoerig uit.

4 pnt

**5e alinea: kern 4**

Duidelijk maken dat je met jouw kwaliteiten **als persoon** uitermate geschikt bent voor deze studie. Verwerk hierbij uitgewerkte informatie uit je CV minimaal 2 punten (ervaring en vaardigheden via baantjes, cursussen, hobby's opgedaan enz.)

8 pnt

**max  
12  
pnt.**

Welke extra's denk jij te kunnen bieden t.o.v. andere kandidaten?

**Noemen + toelichten!**

Persoonlijke kwaliteiten zoals: vermogen tot samenwerken, inzet, doorzettingsvermogen, stressbestendigheid.

4 pnt

**6e alinea: slot**

Doel van deze brief:  
• voldoende informatie verschaft,  
• bereid/hoop uitgenodigd te worden voor nadere kennismaking,  
• om geplaatst te kunnen worden.

1 pnt

1 pnt

1 pnt

**max  
3 pnt.**

**Aantal punten**

**LET OP: op de achterkant van dit vel >>> brief conventies + eventuele lay-out aftrek**

**52**

**Z.O.Z!**

<b>5H</b> <b>2019-2020</b> <b>BRIEFCONVENTIES / LAY-OUT:</b>		<b>Vervolg:</b>  <b>Beoordelingsmodel: (oefen) motivatiebrief</b>
Aftrek fouten in vormconventies: 1 pnt per fout, maximaal 13 punten. (Inhoud aftrek (al.1 en al.6) reeds op vorige pagina verrekend.)	Let op: Zeer strak gecorrigeerd d.w.z. zelfs de kleinste fout, zoals vergeten van een komma, levert voor dat onderdeel -1 pnt op!	<b>Aftrek punten</b>
<ul style="list-style-type: none"> <li>• briefhoofd</li> <li>• dagtekening</li> <li>• geadresseerde</li> <li>• betreft</li> <li>• aanhef</li> <li>• afsluiting</li> <li>• ondertekening verkeerde plek</li> <li>• bijlage</li> </ul>		<ul style="list-style-type: none"> <li>-1</li> <li>-1</li> <li>-1</li> <li>-1</li> <li>-1</li> <li>-1</li> <li>-1</li> <li>-1</li> </ul>
<b>Lay out:</b>		<b>-5</b>
<ul style="list-style-type: none"> <li>• geen witregel tussen alinea's,</li> <li>• te veel tekst zonder nieuwe regel,</li> <li>• te ver met tekst doorgedaan op een bladzijde,</li> <li>• tekst niet mooi gespreid over 2 bladzijdes.</li> </ul>		
<b>EXTRA AFTREK:</b>		
Foutieve toonzetting: onbescheiden, te bescheiden, niet realistisch, plechtig, niet enthousiast. 1 punt per verkeerde toon.		<b>max. - 5</b>

Spellingfouten en vergeten leestekens, 1 pnt per spellingfout, 1 pnt per leestekenfout. Ook bij herhaalde fouten wordt doorge-rekend.		<b>max. - 10</b>
Slechtlopende zinnen 1 pnt per fout.		max. -10
Documentatiemap niet in orde.	Veel voorkomende fouten: • geen titelblad (-1) • geen inhoudsopgave (-1) • geen complete bronvermeldingen (-1) • te weinig bronnen (-1) • oefenbrieven + correctiemodellen niet voorin los opgenomen etc. (-5 per onderdeel dat mist) • CV niet aanwezig (-5)	Verrekend op voorblad (10 punten te behalen indien in orde)
<b>Totaal behaalde punten van vorige pagina: INHOUD, maximaal 52</b>		<b>+</b>
<b>Totaal aantal punten AFTREK zowel voor conventies als extra aftrek, maximaal 38</b>		<b>-</b>
<b>Eind totaal aantal punten:</b>		<b>Eindcijfer: behaalde aantal punten : te behalen aantal punten (=52) x 9 + 1</b>