



Universiteit
Leiden
The Netherlands

Van struikelblok naar bouwsteen: een vakdidactisch ontwerponderzoek naar het schrijven van begrijpelijke alinea's door leerlingen uit het voortgezet onderwijs

Winden, A.W.C. van

Citation

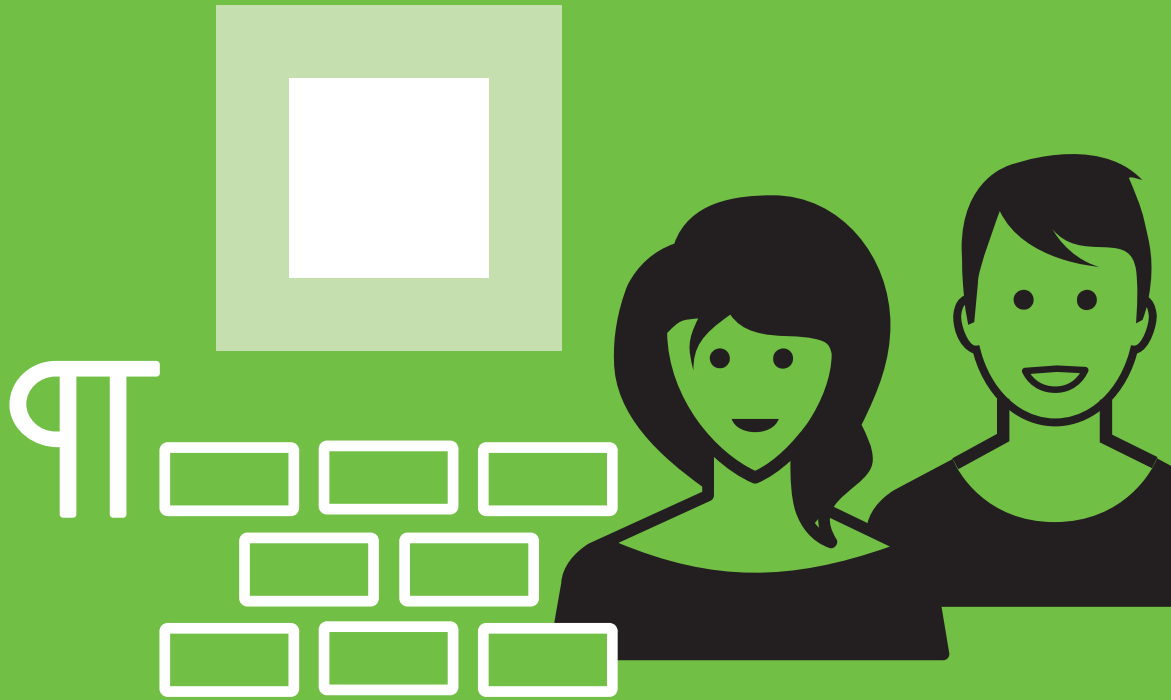
Winden, A. W. C. van. (2024, January 10). *Van struikelblok naar bouwsteen: een vakdidactisch ontwerponderzoek naar het schrijven van begrijpelijke alinea's door leerlingen uit het voortgezet onderwijs*. Retrieved from <https://hdl.handle.net/1887/3677397>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3677397>

Note: To cite this publication please use the final published version (if applicable).



“Ik merkte dat ik niet kritisch genoeg was op mijn teksten. Ik dacht gewoon dat ik het allemaal wel wist. Dat komt omdat de alineanormen heel herkenbaar waren. Dat zag ik ook in mijn klas: alle leerlingen zeiden steeds: ‘Ja mevrouw, dat weet ik wel.’”

Leerling van 5-havo, februari 20209

VI

Wat zijn de effecten van het werken met een toolbox op kennis, self-efficacy en vaardigheden van leerlingen die alinea's schrijven?

Over de kenmerken van alinea's in e-mails en betogende teksten van havo-leerlingen

Dit hoofdstuk zal verschijnen in Tijdschrift voor Taalbeheersing.

Winden, A. van, Haafden, T. van, Stukker, N., Schooten, E. van, Janssen, F. & Glopper, K. de (te verschijnen). *De effecten van het werken met een toolbox op de kennis, self-efficacy en vaardigheden van leerlingen die alinea's schrijven.*

Samenvatting

Dit hoofdstuk verkent de kennis over en de vaardigheden in het schrijven van alinea's en de *self-efficacy beliefs* van leerlingen in het voortgezet onderwijs. Omdat samenhang binnen een tekst een bijdrage levert aan de begrijpelijkheid ervan, is het van belang dat leerlingen in het voortgezet onderwijs leren hoe zij coherente teksten kunnen schrijven. Hiertoe zijn voor leerlingen van 2-havo en 5-havo alineatoolboxen ontwikkeld voor twee genres (informerende e-mails en betogende teksten). Vier ontwerpprincipes waren hierbij leidend: besteed aandacht aan de normen voor het schrijven van begrijpelijke alinea's, start met de genrespecifieke schrijfopdracht als hele taak, werk met een zelfevaluatie-instrument op basis van het ontwikkelde alineanormenkader en bied leerlingen vervolgens hulp op maat aan in de vorm van extra uitleg en oefeningen. Vervolgens is door middel van een interventiestudie onderzocht wat de effectiviteit is van de ontwikkelde alineatoolboxen. Hiervoor hebben we in een quasi-experimenteel design een vragenlijst uitgezet onder 300 leerlingen en daarnaast hebben we ruim 2100 van hun alinea's geanalyseerd op de toepassing van alineanormen. De resultaten laten zien dat er geen causaal verband is tussen het gebruik van de ontwikkelde alineatoolboxen in het onderwijs en de toename van kennis, vaardigheden en *self-efficacy beliefs* bij leerlingen.

VI Wat zijn de effecten van het werken met een toolbox op kennis, self-efficacy en vaardigheden van leerlingen die alinea's schrijven?

Over de kenmerken van alinea's in e-mails en betogende teksten van havoleerlingen

1 Inleiding

In de huidige lesmethodes Nederlands is er nauwelijks aandacht voor het schrijven van alinea's. Wel worden kenmerken van een alinea besproken als onderdeel van leerstof waar een ander thema centraal staat, bijvoorbeeld leesvaardigheid of formuleren, maar een samenhangende en enigszins complete behandeling van de normen waar een begrijpelijk geschreven alinea aan moet voldoen, zien we niet (zie hoofdstuk 2). Dat is opmerkelijk, want alinea's hebben binnen teksten de belangrijke functie bij te dragen aan het creëren van samenhang. En samenhang wordt gezien als het belangrijkste kenmerk van een begrijpelijke tekst (Pander Maat, 2002).

Omdat samenhang een bijdrage levert aan de begrijpelijkheid van een tekst, is het van belang dat leerlingen leren hoe zij coherente teksten kunnen schrijven. Het schrijven van alinea's vormt binnen de tekst juist dat onderdeel waarin de structuur van de gehele tekst en de formulering op zinsniveau samenkomen. Dat maakt de alinea tot een belangrijke bouwsteen van een begrijpelijke tekst. Als we leerlingen in het voortgezet onderwijs willen leren hoe zij begrijpelijke teksten kunnen schrijven, is aandacht voor de alinea daarom op zijn plaats. Om die reden hebben we in een eerdere studie een reconstructie gemaakt van de normen waaraan alinea's zouden moeten voldoen om bij te dragen aan een begrijpelijke tekst (zie hoofdstuk 2).

Eerder presenteerden we in een deelstudie een nulmeting waarbij we de kwaliteit van de alinea's in teksten voor twee genres onderzochten. Daaruit bleek dat de alinea-vaardigheden tekortschieten bij leerlingen in 2-havo en in 5-havo (zie hoofdstuk 3). Vervolgens beschreven we in een deelstudie een nulmeting van de kennis van het schrijven van alinea's en de *self-efficacy beliefs* van leerlingen in 2-havo en in 5-havo (zie hoofdstuk 4). De resultaten van deze studies laten zien dat leerlingen onderwijs in het schrijven van alinea's nodig hebben. Op basis van de resultaten uit deze eerste drie studies hebben we een didactisch ontwerponderzoek uitgevoerd waarin we toolboxes ontwikkeld hebben (zie hoofdstuk 5). In dit hoofdstuk beschrijven we een interventiestudie waarin we de effectiviteit van de ontwikkelde alineatoolboxes onderzoeken

2 Theoretisch kader

Voor het ontwikkelen van vier alineatoolboxes waarmee we leerlingen het schrijven van alinea's hebben willen leren zijn in hoofdstuk 5 vier ontwerpprincipes gepresenteerd: één inhoudelijk ontwerpprincipe dat betrekking heeft op alineanormen en drie algemene ontwerpprincipes voor effectief onderwijs. Het eerste, inhoudelijk, ontwerpprincipe betreft het presenteren van elf alineanormen waaraan begrijpelijk geschreven alinea's moeten voldoen. Deze normen zijn gereconstrueerd uit Nederlands-talige taaladviesboeken en samengebracht in een kader (zie tabel 1). Vanuit theoretische en empirische inzichten uit de taal- en tekstwetenschap zijn er aanwijzingen dat het volgen van deze normen bijdraagt aan de totstandkoming van begrijpelijk gestructureerde en geformuleerde alinea's. Voor een uitgebreide bespreking van de vier categorieën binnen het normenkader, de individuele alineanormen, hun theoretische onderbouwing en het empirisch bewijs verwijzen we naar hoofdstuk 2.

TABEL 1

Op basis van Nederlandstalige taaladviesboeken (N = 29) gereconstrueerde alineanormen verdeeld over vier categorieën met in de laatste kolom de vermelding hoe vaak de norm aangetroffen werd in taaladviesboeken.

Categorie	Nr.	Norm	Aantal keren aanwezig
Afbakening van de alinea	1	De schrijver gebruikt een van de volgende manieren om te markeren dat een nieuwe alinea begint	
		[a] inspringen	17
		[b] een witregel gebruiken	20
	2	Een alinea bestaat uit ten minste twee zinnen	12
Coherentie op globaal structuurniveau	3	De functie van de alinea is herkenbaar	7
	4	De alinea hangt samen met de centrale boodschap van de tekst	3
	5	De kernzin ondersteunt de centrale boodschap van de tekst	4
Coherentie op lokaal structuurniveau: thematische samenhang	6	De alinea beslaat één deelonderwerp van de tekst	26
	7	De belangrijkste uitspraak over het deelonderwerp staat in de kernzin	21
	8	De overige informatie sluit aan bij de kernzin	17
		Nieuwe informatie in de alinea wordt gekoppeld aan wat de lezer al weet	
		[a] door deze nieuwe informatie te koppelen aan kennis over woordbetekenissen	1
	9	[b] door deze nieuwe informatie te koppelen aan kennis over de wereld om ons heen	2
Coherentie op lokaal structuurniveau: relaties tussen zinnen		Binnen de alinea worden de zinnen verbonden tot een samenhangend geheel	
	10	[a] door het gebruik van verwijswaarden en/of	15
		[b] door het gebruik van verbindingswoorden en/of	22
		[c] door impliciete verbanden	3
		De alinea is opgebouwd op een van de volgende manieren:	
11	[a] als piramide: de plaats van de kernzin is aan het begin van de alinea	21	
	[b] als zandloper: de plaats van de kernzin is in het midden van de alinea	6	
	[c] als trechter: de plaats van de kernzin is aan het einde van de alinea	15	

Uit de eerder genoemde nulmetingen van de vaardigheden in het schrijven van alinea's (zie hoofdstuk 3) en de kennis over alinea's en *self-efficacy beliefs* van leerlingen (zie hoofdstuk 4) kunnen we op hoofdlijnen enkele conclusies trekken³¹.

Het blijkt dat leerlingen in 2-havo en in 5-havo met name moeite hebben om coherentie op globaal structuurniveau (norm 5) en op lokaal structuurniveau (norm 6, 7 en 8) in alinea's aan te brengen. Qua kennis blijkt dat de leerlingen het moeilijk vinden om adequaat onder woorden te brengen wat een alinea is en wat de functie van een alinea in een tekst is. We zien dat hun declaratieve kennis beperkt is, terwijl leerlingen er wel van overtuigd zijn dat ze in staat zijn begrijpelijke alinea's te schrijven. Leerlingen kennis aanbieden over alinea's is dus wenselijk, want een bewust taalvaardige leerling heeft kennis van de normen nodig (zie hoofdstuk 4). Docenten Nederlands kunnen hiervoor het overzicht van de alineanormen uit tabel 1 gebruiken. Leerlingen iets leren bestaat echter uit meer dan alleen kennis in de vorm van taalnormen aanbieden.

De ontwerpstudie (zie hoofdstuk 5) is naast de presentatie van het inhoudelijke ontwerp-principe dan ook gericht op het identificeren van didactische ontwerp-principes voor een effectieve en praktisch bruikbare alineadidactiek. Voor het formuleren van deze drie ontwerp-principes voor effectief onderwijs zijn we te rade gegaan bij algemene onderwijstheorieën over het onderwijzen en leren van complexe taken. Het schrijven van alinea's is zo'n complexe taak, omdat op het niveau van de alinea formulering en structuur samenkomen. Het tweede ontwerp-principe is het starten van een les(senreeks) met de presentatie van de genrespecifieke schrijfofdracht als complete taak. Door leerlingen een genrespecifieke schrijfofdracht als hele taak eerst te geven, daagt de docent zijn leerlingen uit hen zelf na te laten denken over de vragen die zij zich moeten stellen om de taak te kunnen volbrengen. Dit helpt leerlingen grip te krijgen op een complexe opdracht (Merrill, 2002; Janssen et al., 2016). Het derde ontwerp-principe is het werken met een zelfevaluatie op basis van de gereconstrueerde alineanormen. Na het schrijven van een eerste versie van de genrespecifieke opdracht controleren de leerlingen hun opdracht met behulp van een zelfevaluatie. Dit is een vorm van formatieve evaluatie waarbij leerlingen reflecteren op de kwaliteit van hun werk met als doel om hun schrijfproduct in een volgende versie te verbeteren (Andrade & Valcheva, 2009). De focus ligt daarbij op de talige aspecten van het genre in combinatie met de alineanormen. Het vierde en laatste ontwerp-principe bestaat uit het op maat aanbieden van hulp en oefeningen (Janssen et al., 2016). We geven leerlingen de verantwoordelijkheid om zelf te bepalen bij welke onderdelen van de alinea zij nog hulp nodig hebben om de gehele taak goed te volbrengen. Per onderwerp krijgen de leerlingen theorie, extra uitleg en oefeningen aangeboden.

³¹ Deze nulmetingen van vaardigheden (zie hoofdstuk 3), kennis en *self-efficacy beliefs* (zie hoofdstuk 4) betreffen een deel van het materiaal waar we in dit hoofdstuk verslag van doen; het zijn in de context van dit hoofdstuk de voormetingen van de vaardigheden, kennis en *self-efficacy beliefs*.

Hierdoor hoeven leerlingen zich niet met zaken bezig te houden die ze al beheersen. Dat zorgt voor motivatie en mede daardoor voor een effectieve, gedifferentieerde manier van leren.

Niet alleen de effectiviteit voor leerlingen telt; het is ook belangrijk dat de alineadidactiek praktisch bruikbaar is voor docenten. Een nieuw ontwikkelde didactiek zal immers alleen in de lespraktijk geïmplementeerd worden als docenten de onderwijs-vernieuwing als praktisch ervaren. Het is dus van belang dat een interventie focust op de drie eisen waaraan een praktisch bruikbare didactiek voor docenten moet voldoen: leerlingen moeten er zelfstandig mee kunnen werken, extra tijd en middelen zijn niet of nauwelijks benodigd en er is sprake van congruentie met wat docenten doen en belangrijk vinden (Janssen et al., 2015).

Op basis van de bovengenoemde ontwerp-principes hebben we genrespecifieke toolboxes ontwikkeld om leerlingen kennis te laten maken met de alineanormen én hen hiermee te laten oefenen. Een toolbox is hier een set van didactische middelen die docenten kunnen gebruiken tijdens de schrijfvaardigheidslessen. Concreet bestaat elke toolbox uit een schrijfofdracht, een les waarin de tekst- en taalkenmerken van het betreffende genre centraal staan, een zelfevaluatie-instrument, hulp-op-maat-opdrachten en een herschrijfofdracht. De algemene opzet van elke toolbox is dezelfde, maar per genre en per jaarlaag is gewerkt met verschillende opdrachten, teksten en invalshoeken. We ontwikkelden vier toolboxes; de leerlingen in 2-havo en in 5-havo werkten met twee verschillende genres: een e-mail met een verzoek en een ingezonden brief/opinieartikel. De keuze voor de genres is een pragmatische; in het voortgezet onderwijs worden deze genres veel geschreven. Leerlingen leren gewoonlijk in een betogende tekst hun mening te formuleren en te beargumenteren en ook de e-mail met een verzoek is voor leerlingen een bekend genre dat zij niet alleen in de toekomst maar ook tijdens hun schoolcarrière al schrijven.

Hoe functioneren de aldus ontworpen toolboxes in de praktijk? In dit artikel doen we verslag van een experimentele studie waarin we de effectiviteit van de hierboven beschreven toolboxes getest hebben bij leerlingen in 2-havo en in 5-havo. De onderzoeksvraag luidt: Wat is het effect van het werken met een alineatoolbox op de kennis van leerlingen over alinea's, op hun *self-efficacy beliefs* en op hun vaardigheid in het schrijven van alinea's?

3 Methode

3.1 Onderzoeksdesign

In deze studie hebben we gebruikgemaakt van een quasi-experimenteel onderzoeksdesign waarin we hele klassen aan de verschillende condities hebben toegewezen. Hoewel een dergelijk design een lagere interne validiteit heeft dan een experimenteel onderzoeksdesign, is de keuze voor dit design toch te verantwoorden, omdat de ecologische validiteit hoog is. Er is immers sprake van onderzoek in een natuurlijke setting.

De verzameling van de data vond in het schooljaar 2019-2020 plaats op de eigen school van de docent-onderzoeker en wel in twee jaarlagen: 2-havo en 5-havo. De keuze voor onderzoek doen op één school kent beperkingen. Generaliserende uitspraken doen naar aanleiding van de resultaten van dergelijk onderzoek is immers lastig. In dit onderzoek hebben we gemeend dat de vernieuwende aard van het alineaonderwijs deze keuze rechtvaardigt, omdat het ervoor zorgde dat we konden werken met een selectieve groep docenten die door de docent-onderzoeker persoonlijk geïnstrueerd kon worden en daarnaast was er zicht op de implementatie van het onderwijs.

De betreffende school achten we voldoende representatief voor de populatie. De school valt zowel qua resultaten als qua leerlingkenmerken te beschouwen als een gemiddelde school. De examenresultaten en de resultaten voor het vak Nederlands liggen voor de havo op het landelijk gemiddelde: havoleerlingen scoren op deze school voor hun eindexamen gemiddeld een 6.3 ten opzichte van een 6.3 landelijk en voor het vak Nederlands gemiddeld een 6.2 ten opzichte van een 6.2 landelijk. Ook de leerlingkenmerken laten zien dat deze school een gemiddelde is: 16% van de leerlingen heeft een niet-westerse migratieachtergrond ten opzichte van 16% landelijk; 31% komt uit een gezin met een relatief laag inkomen ten opzichte van 40% landelijk; 15% is afkomstig uit een eenoudergezin ten opzichte van 18% landelijk (Nationaal Regieorgaan Onderwijsonderzoek, 2020)³². In de lessen Nederlands wordt gewerkt met de methode *Nieuw Nederlands* (5^e editie). Landelijk is dit een van de meest gebruikte lesmethodes (Van der Leeuw et al., 2017).

In tabel 18 staat het quasi-experimenteel design voor 2-havo en 5-havo schematisch weergegeven. De vragenlijst over kennis en *self-efficacy beliefs* werd driemaal afgenomen: de voormeting voorafgaand aan de interventie, de tussenmeting halverwege en

³² Deze gegevens zijn afkomstig uit een vertrouwelijke bron van het Nationaal Regieorgaan Onderwijsonderzoek (2020). We vermelden de gegevens in dit artikel met toestemming van de directeur van de betreffende school.

de nameting aansluitend aan de interventie. Tussen de designs voor beide jaarlagen is één verschil zichtbaar. De leerlingen van 2-havo hebben de toolbox per genre tweemaal doorlopen³³, terwijl de leerlingen van 5-havo per genre slechts eenmaal werkten met de toolbox, omdat zij al meer ervaring hebben met schrijven in het algemeen en met deze twee genres in het bijzonder. Hierbij speelde ook een rol dat leerlingen in 5-havo een programma volgen dat vastligt in het Programma van Toetsing en Afsluiting (PTA), waardoor er minder tijd beschikbaar is. In de data-analyses zijn voor de voormeting de teksten van 2-havo en 5-havo uit ronde 1 (schrijftaken E1 en B1) gebruikt. Voor de nameting hebben we voor 2-havo de e-mails uit ronde 2 gebruikt (herschrijving E2) en konden we voor de betogende teksten alleen gebruikmaken van de teksten uit ronde 1 (herschrijving B1; zie noot 33). De nametingen van 5-havo bestonden uit de teksten van ronde 1 (herschrijving E1 en B1).

³³ Als gevolg van het sluiten van de scholen in het voortgezet onderwijs tijdens het begin van de COVID-19-epidemie in maart 2020 hebben we voor 2-havo de tweede ronde van de betogende tekst (B2-schrijftaak en B2-herschrijving) niet kunnen afnemen zoals gepland.

TABEL 18

Overzicht van het quasi-experimenteel design voor 2-havo en 5-havo (VM = voormeting, TM = tussenmeting; NM = nameting, K+SE = kennis en self-efficacy, E = e-mail, B = betoog, 1 = ronde 1, 2 = ronde 2)

2-havo		Experimentele groep (N = 56)		Controlegroep (N = 59)	
Meetmoment	Instrument	Onderwijs	Instrument	Onderwijs	
VM K+SE	vragenlijst K+SE		vragenlijst K+SE		
VM E-mail	E1 schrijftaak	E1 schrijftaak	E1 schrijftaak	E1 schrijftaak	
		E1 toolbox		E1 reguliere lessen	
		E1herschrijving		E1herschrijving	
<i>Twee weken tussentijd</i>					
		E2 schrijftaak		E2 schrijftaak	
		E2 toolbox		E2 reguliere lessen	
NM E-mail	E2 herschrijving	E2 herschrijving	E2 herschrijving	E2 herschrijving	
TM K+SE	vragenlijst K+SE interviews		vragenlijst K+SE		
VM Betoog	B1 schrijftaak	B1 schrijftaak	B1 schrijftaak	B1 schrijftaak	
		B1 toolbox		B1 reguliere lessen	
		B1 herschrijving		B1 herschrijving	
<i>Twee weken tussentijd</i>					
		B2 schrijftaak		B2 schrijftaak	
		B2 toolbox		B2 reguliere lessen	
NM Betoog	B2 herschrijving	B2 herschrijving	B2 herschrijving	B2 herschrijving	
NM K+SE	vragenlijst K+SE interviews		vragenlijst K+SE		
5-havo		Experimentele groep (N = 111)		Controlegroep (N = 77)	
Meetmoment	Instrument	Onderwijs	Meetmoment	Instrument	
VM K+SE	vragenlijst K+SE		vragenlijst K+SE		
VM E-mail	E1 schrijftaak	E1 schrijftaak	E1 schrijftaak	E1 schrijftaak	
		E1 toolbox		E1 reguliere lessen	
NM E-mail	E1 herschrijving	E1 herschrijving	E1 herschrijving	E1 herschrijving	
TM K+SE	vragenlijst K+SE interviews		vragenlijst K+SE		
VM Betoog	B1 schrijftaak	B1 schrijftaak	B1 schrijftaak	B1 schrijftaak	
		B1 toolbox		B1 reguliere lessen	
NM Betoog	B1 herschrijving	B1 herschrijving	B1 herschrijving	B1 herschrijving	
NM K+SE	vragenlijst K+SE interviews		vragenlijst K+SE		

3.2 Het onderwijs met de toolbox en de lessen in de controleconditie

Voor beide condities gold dat de lessenseries voor beide genres een lengte hadden van drie à vier lessen. Elke les duurde 50 minuten. De kennis over de kenmerken van het genre presenteerde de docent via een powerpointpresentatie. Leerlingen maakten de schrijfpodrachten thuis af en leverden zowel de eerste als de herschreven versie in via de digitale leeromgeving.

De interventie werd geïmplementeerd in het bestaande jaarprogramma van het vak Nederlands. De lessen voor de experimentele groepen bestonden uit het werken met een toolbox waarin aandacht voor de alineanormen centraal stond volgens de ontwerpcriteria die hierboven zijn beschreven. De leerlingen kregen de schrijfpodracht als hele taak gepresenteerd. Zij maakten in de les kennis met het genre en de genrekenmerken. Daarna schreven zij een eerste versie van de tekst. In een volgende les kregen de leerlingen meer duidelijkheid over tekstkenmerken van het genre en de alineakenmerken, gericht op het formuleren en structureren van de inhoud. Daarna vulden de leerlingen voor een alinea een zelfevaluatie in en bepaalden zij met welke onderdelen van de alinea in het genre zij nog moesten oefenen. De leerlingen maakten één of meerdere hulp-op-maatopdrachten en keken deze zelf na aan de hand van antwoordboekjes. Vervolgens herschreven de leerlingen hun tekst met behulp van de opgedane kennis en leverden deze in. De schrijfpodracht, de vragen voor de zelfevaluatie, de hulp-op-maatopdrachten en de herschrijfpodracht ontvingen de leerlingen uit de experimentele groepen in een werkboekje per genre.

De leerlingen uit de controleklassen volgden de reguliere schrijfvaardigheidslessen waarin de fasen van het schrijfproces gevolgd werden: oriëntatie op de schrijftaak, bezigheden tijdens het schrijven en reflectie op en revisie van de schrijftaak. De lessen en de revisieformulieren waren gebaseerd op het materiaal uit de lesmethode, waarbij er geen leerstof en oefeningen met betrekking tot de alinea werden aangeboden. De lessen van deze controleklassen startten steeds met een klassikale instructie waarin de kenmerken van het genre centraal stonden, gevolgd door een oriëntatie op de schrijfpodracht die leerlingen in groepjes uitvoerden (Elving & Van den Bergh, 2015) waarna de leerlingen zelfstandig opdrachten maakten uit de lesmethode. Deze opdrachten pasten bij het genre en bij de fase van het schrijfproces.

3.3 Deelnemers

De deelnemers aan deze studie waren leerlingen van 2-havo en 5-havo. In 2-havo zaten 118 leerlingen verdeeld over vier klassen en in 5-havo zaten 195 leerlingen verdeeld over zeven klassen. De ouders van de deelnemers werden per brief geïnformeerd over het onderzoek en om toestemming voor deelname gevraagd. Daarnaast gaven de betreffende docenten toestemming voor het gebruik van de data. Ook de

directie van de school was akkoord met de uitvoering van het onderzoek³⁴. Van drie leerlingen uit 2-havo en van twee leerlingen uit 5-havo gaven de ouders geen toestemming voor deelname aan het onderzoek. Daarnaast waren er zeven leerlingen uit 5-havo chronisch ziek of gingen gedurende het jaar van school af. Hun gegevens werden uit de dataset verwijderd. Voor 2-havo ($N = 115$) was de gemiddelde leeftijd 13.9 jaar ($SD = .46$); 51.3% was man. Voor 5-havo ($N = 188$) was de gemiddelde leeftijd 17.3 jaar ($SD = .68$); 48.4% was man.

De leerlingen in de vier klassen van 2-havo kregen les van twee docenten (gemiddelde leeftijd: 47.5 jaar ($SD = 9.2$), gemiddeld aantal jaren ervaring in het voortgezet onderwijs: 13 jaar ($SD = 4.2$)). De leerlingen in de zeven klassen van 5-havo kregen les van vijf docenten (gemiddelde leeftijd: 41.2 jaar ($SD = 13.6$), gemiddeld aantal jaren ervaring in het voortgezet onderwijs: 16.8 jaar ($SD = 11.9$)). De betrokken docenten gaven allen les in één van beide condities. Alleen de docent-onderzoeker gaf les aan 2-havo én 5-havo. Zij gaf les in de experimentele conditie.

Van de leerlingen van 2-havo zijn alle teksten uit de voor- en nameting in de analyse betrokken. Van de leerlingen van 5-havo werden per klas de eerste 13 teksten in de analyse betrokken.

3.4 Instrumenten en data-analyse

Voor de meting van de kennis, self-efficacy beliefs en vaardigheden hebben we verschillende instrumenten gebruikt. In deze paragraaf staan de vragenlijst, het analysemodel en de interviews beschreven. Daarnaast beschrijven we ook de wijze van data-analyse, de betrouwbaarheid van de instrumenten en de berekening van somscores.

3.4.1 Vragenlijst: kennis van leerlingen over alinea's

Met behulp van een vragenlijst onderzochten we de kennis en self-efficacy beliefs van de leerlingen. Deze vragenlijst is in bijlage 4 opgenomen. Het eerste onderdeel van de vragenlijst over de kennis van leerlingen over alinea's schrijven betrof twee opdrachten: 'Geef in je eigen woorden weer wat volgens jou een alinea is' en 'Leg uit waarom het belangrijk is dat een tekst in alinea's is opgedeeld' en vervolgens dertig gesloten vragen naar de kennis die een leerling heeft over alinea's. Het doel van de twee opdrachten was om te kunnen bepalen welke informatie leerlingen op eigen kracht kunnen noteren over alinea's en over het belang van het schrijven in alinea's. De dertig gesloten vragen bevatten achttien stellingen over alinea's, waarvan

³⁴ Het onderzoeksproject waar dit artikel de effectstudie van is werd goedgekeurd door de ethische commissie van het ICLON / Universiteit Leiden en geregistreerd onder projectnummer IREC_ICLON 2019-06.

de deelnemers moesten aangeven of ze deze 'waar' of 'niet waar' vonden. Dit gedeelte van de vragenlijst volgde het voorbeeld van een vragenlijst over metacognitieve kennis van lezen en schrijven uit het project Nelson (Schoonen et al., 2003) en werd gebaseerd op de gereconstrueerde alinea's (zie hoofdstuk 2). Tot slot van het kennisgedeelte van de vragenlijst moesten de deelnemers van twaalf signaalwoorden aangeven welk verband deze aanduiden: oorzaak-gevolg, tegenstellend, redengevend of samenvattend. Bij alle dertig gesloten items werd aan de leerlingen gevraagd hoe zeker ze waren van hun antwoord: op een schaal van 0 ('ik weet het antwoord echt niet, ik gok') tot 100 ('ik ben er helemaal zeker van dat mijn antwoord klopt') mochten ze elk getal geven dat ze wilden (Stankov & Crawford, 1997; Lundeborg et al., 1994).

TABEL 19

Constructie van de somscores Definitiekwaliteit en Onderkenning belang

Definitiekwaliteit		Onderkenning belang	
0	Misconceptie gegeven	0	Misconceptie gegeven
1	Geen van de concepten zijn benoemd	1	per concept <i>overduidelijk, lengte, voor het opzoeken van informatie, leesbaarheid of begrip</i>
2	<i>Onderdeel, markering of lengte</i> zijn benoemd		
3	<i>Onderdeel en markering of lengte</i> zijn benoemd		
4	<i>Onderdeel en markering en lengte</i> zijn benoemd / <i>Deelonderwerp</i> is benoemd		
5	<i>Deelonderwerp en onderdeel</i> zijn benoemd samen met <i>markering of lengte</i> / <i>Deelonderwerp en markering en lengte</i> zijn benoemd		
6	<i>Deelonderwerp en onderdeel</i> zijn benoemd samen met een van <i>markering of lengte</i>		
7	Alle concepten zijn benoemd		

De antwoorden op de twee opdrachten hebben we kwalitatief geanalyseerd door te tellen welke concepten leerlingen benoemen. Bij opdracht 1 hebben we langs inductieve weg de volgende categorieën onderscheiden: *een stukje tekst, een deel van een tekst, een deelonderwerp, een opmerking over de markering, een opmerking over de lengte van de alinea*. Bij opdracht 2 hebben we de volgende categorieën onderscheiden: *overzichtelijkheid, duidelijkheid, lengte, voor het opzoeken van informatie, leesbaarheid of begrip*³⁵. Voor beide opdrachten hebben we somscores berekend door de 0/1-scores voor de af- of aanwezigheid van de concepten in de antwoorden van de leerlingen gewogen bij elkaar op tellen. We beschouwen deze somscores voor 'definitiekwaliteit' en 'onderkenning belang' als indicator voor de kwaliteit van de antwoorden van de leerlingen. Immers, antwoorden die meer concepten bevatten zorgen voor een rijkere definitie van de alinea. Bij vraag 1 liep de somscore van 0 tot en met 7, bij vraag 2 van 0 tot en met 4. Zie voor de bepaling van de scores tabel 19. Een tweede beoordelaar

³⁵ Omdat de concepten *een stukje tekst en een deel van een tekst* bij opdracht 1 inhoudelijk gelijk zijn, hebben we deze samen genomen in het concept *onderdeel*. Bij opdracht 2 geldt ditzelfde voor de concepten *overzicht* en *duidelijkheid*. Deze hebben we samengevoegd tot het concept *overduidelijk*.

heeft een steekproef van antwoorden geanalyseerd om op deze manier de interbeoordelaarsbetrouwbaarheid te kunnen meten. De overeenstemmingspercentages bij het vaststellen van de aanwezigheid van de concepten zijn hoog; ze liggen tussen 89% en 100%.

Voor de gesloten vragen hebben we scores berekend door eerst de foute antwoorden op kennisvragen te scoren als -1 en de goede antwoorden als +1. Vervolgens zijn deze vermenigvuldigd met het antwoord op de vraag naar hoe zeker leerlingen waren van hun antwoord. Dit leverde scores op met een mogelijke range van min 100 tot plus 100. We hebben gekozen voor deze combinatievariabele van kennis en zekerheid, omdat de raadkans van 50% bij de kennisvragen een negatief effect op de betrouwbaarheid heeft. We verwachtten dat deze toegevoegde informatie de betrouwbaarheid van de schaal ten goede zou komen en dat bleek ook zo te zijn: voor de voor-, tussen- en nameting was de betrouwbaarheid (Cronbach α) voor 'kennis sec' (zonder vermenigvuldiging met zekerheid) met respectievelijk .61, .61 en .68, beduidend lager dan de corresponderende waarden voor 'kennis*zekerheid' in tabel 20 hieronder (.67, .68 en .72). Deze laatste waarden geven aan dat de betrouwbaarheid van 'kennis*zekerheid' voor de drie meetmomenten voldoende was.

Ter toelichting: de betrouwbaarheidsanalyses voor 'kennis sec' en 'kennis*zekerheid' vertrokken vanuit de gegevens van de voormeting. In de eerste analyses waren voor beide typen scores alle 30 items betrokken. Stapsgewijs is steeds het item met de hoogste negatieve gecorrigeerde item-restcorrelatie verwijderd, totdat er geen items met een negatieve item-restcorrelatie over waren. In totaal zijn er zo voor de twee typen scores dezelfde zeven items uit de vragenlijst verwijderd, items die niet geassocieerd waren met een andere factor of dimensie en die waarschijnlijk door hun formulering niet bijdroegen aan de betrouwbaarheid.

De betekenis van de schaal 'kennis*zekerheid' valt overigens vrijwel samen met die van de schaal 'kennis sec'. De correlaties tussen de schalen beliepen .94, .92 en .90 voor respectievelijk de voor-, tussen- en nameting.

TABEL 20

Betrouwbaarheid (Cronbach Alpha) voor kennisitems gecombineerd met zekerheid en voor *self-efficacy beliefs*

Meetmoment	Cronbach α voor 23 items kennis*zekerheid	Cronbach α voor 14 items <i>self-efficacy beliefs</i>
Voormeting	.67	.91
Tussenmeting	.68	.91
Nameting	.72	.90

3.4.2 Vragenlijst: self-efficacy beliefs van leerlingen bij het schrijven van alinea's

Het tweede deel van de vragenlijst bevatte veertien vragen over *self-efficacy beliefs*. Deze vragen waren gebaseerd op een vragenlijst van Shell et al. (1989) die *self-efficacy beliefs* gemeten heeft ten aanzien van schrijftaken en schrijfvaardigheid. Het gedeelte over de schrijfvaardigheid was voor deze vragenlijst gespecificeerd voor het schrijven van alinea's. Daarbij is gebruikgemaakt van de alineanormen zoals die in hoofdstuk 2 gepresenteerd zijn. De schaalverdeling liep van 0 tot 100, waarbij 0 betekende dat de leerling er geen vertrouwen in heeft dat de in het item gepresenteerde taak hem zal lukken en 100 betekende dat de respondent er veel vertrouwen in heeft dat hij zal slagen in de in het item gepresenteerde taak. Voor deze schaalverdeling is gekozen, omdat deze een sterkere voorspeller van schrijfvaardigheid blijkt te zijn dan een Likertschaal (Pajares et al., 2001).

Alle items uit de vragenlijst zijn voor dit onderdeel in de analyse opgenomen. De betrouwbaarheid is voor de drie metingen hoog te noemen (zie tabel 20).

3.4.3 Analysemodel alinea's

De alineanormen die we presenteerden in tabel 1, hebben we geoperationaliseerd in een analysemodel. Zie bijlage 2. Daarbij hebben we gekozen voor coderen in ja/nee-vorm. We willen immers weten of de betreffende alineakenmerken voorkomen in de alinea's van leerlingen. De reconstructie en verantwoording van het normenkader staan in hoofdstuk 2 en een uitgebreide beschrijving van het analysemodel staat in hoofdstuk 3.

Om de interbeoordelaarsbetrouwbaarheid te meten hebben de docent-onderzoeker en een onderzoeksassistent in drie fasen alinea's beoordeeld. De eerste fase was een verkennende waarin het analysemodel getest werd op bruikbaarheid en betrouwbaarheid. De docent-onderzoeker en de onderzoeksassistent werkten in nauw overleg, testten de interbeoordelaarsbetrouwbaarheid en legden twijfelgevallen voor aan een van de supervisors. Nadat het analysemodel definitief vastgesteld was, zijn in de tweede fase de alinea's uit de teksten van de voormeting geanalyseerd. Onderdeel van deze fase was een steekproef (van 96 teksten met 327 alinea's) die de docent-onderzoeker en de onderzoeksassistent onafhankelijk van elkaar beoordeelden. In de laatste fase werden de alinea's uit de teksten van de nameting beoordeeld. Hierin werd een steekproef van 97 teksten met 384 alinea's door de onderzoekers onafhankelijk van elkaar beoordeeld. Zie voor een voorbeeld van een analyse bijlage 3.

Op basis van beide steekproeven samen is het percentage overeenstemming tussen beide beoordelaars berekend. In tabel 21 staat dit percentage overeenstemming per alineanorm weergegeven. Voor alle normen ligt het percentage boven de 84%. Ook

Cohens kappa is als maat voor de interbeoordelaarsbetrouwbaarheid met waarden tussen de .69 en .95 voldoende tot goed te noemen. Er is echter een uitzondering. Dat is norm 1b ('De tekst bevat alinea's'). Hoewel het percentage overeenstemming voor deze alinea-norm hoog is, is de kappa zeer laag. Net als in de metingen van de teksten van de eerste versie van deze data (zie hoofdstuk 3) komt dit, doordat de beoordelaars zeer scheve verdelingen aanbrengen, maar het wel volledig eens zijn over de teksten waarin alinea's gemarkeerd zijn.

TABEL 21

Overeenstemmingsscores van de interbeoordelaarsbetrouwbaarheid voor de alinea-normen uit tabel 1

Norm	1a	1b	2	3	4	5	6	7	8	9a	9b	10a	10b	11
% overeenstemming	95	95.1	97.9	88.9	91.3	89.2	90.9	89.1	88.7	96.7	95.3	84.6	84.4	88.1
Cohens kappa	.84	.13	.95	.78	.82	.76	.82	.77	.76	.93	.90	.69	.69	.72

3.4.4 Interviews

Direct na de afronding van de lessen van een genre hield de docent-onderzoeker individuele interviews met de docenten van de experimentele groepen en focusgroep-interviews met enkele leerlingen per experimentele klas. Beide soorten interviews waren semi-gestructureerd en de docent-onderzoeker gebruikte een interview-leidraad. De interviews met de docenten hadden als doel te verhelderen hoe de lessen verlopen waren en de ervaringen van de docenten zelf en hun visie op de werkwijze van de leerlingen te achterhalen. De focusgroepinterviews waren gesprekken met groepjes leerlingen waarbij het doel was te achterhalen hoe leerlingen het werken met toolbox ervaren hadden. De vragen waren gericht op het proces van werken en op het uiteindelijke schrijfproduct. De interviews werden opgenomen, getranscribeerd en kwalitatief geanalyseerd.

3.5 Procedures

De interventie werd uitgevoerd in het schooljaar 2019-2020. Het schrijven van de e-mail vond zowel voor 2-havo als voor 5-havo aan het begin van het schooljaar plaats (september-oktober 2019). Het schrijven van de betogende teksten schrijven vond voor 5-havo plaats in de periode november 2019-januari 2020; 2-havo schreef de betogende tekst in de periode januari-maart 2020.

De schrijfpoddrachten voor de e-mail waren identiek voor 2-havo en 5-havo met als enig verschil dat de leerlingen van 2-havo de informatie die zij in hun tekst moesten opnemen puntsgewijs gepresenteerd kregen, terwijl 5-havo geacht werd deze informatie te destilleren uit de beschrijving van de opdracht. De schrijfpoddrachten voor

de betogende teksten waren verschillend voor de twee doelgroepen: 2-havo schreef een ingezonden stuk en 5-havo schreef een opinieartikel. De reden hiervoor is dat de opbouw van het opinieartikel complexer is dan die van het ingezonden stuk. Omdat leerlingen uit 5-havo in eerdere leerjaren ook betogende teksten geschreven hebben, verwachtten we dat zij deze complexiteit aankunnen. Zie voor de schrijfpoddrachten bijlage 1.

3.6 Statistische analyse

Zowel de data uit de vragenlijst als de data die beschikbaar kwamen na het analyseren van de alinea's uit de teksten variëren op verschillende niveaus³⁶. De scores voor kennis en *self-efficacy beliefs* variëren op twee niveaus: tussen leerlingen en tussen klassen. Er waren namelijk data beschikbaar van in totaal 303 leerlingen uit 11 klassen (5 klassen vormden de controlegroep en 6 klassen de experimentele groep). De scores voor de alinea-kenmerken variëren op drie niveaus: binnen teksten (en dus tussen alinea's), tussen teksten van leerlingen³⁷ en tussen klassen. De 2109 alinea's (1260 uit de betogende teksten en 849 uit de e-mails) komen uit de voor- en de nameting van de teksten van in totaal 311 leerlingen uit dezelfde experimentele en controlegroep. De data zijn dus hiërarchisch gestructureerd. Om deze reden hebben we voor de statistische analyse van de data gebruikgemaakt van multilevelanalyse in MLwiN (Rasbash et al., 2000).

3.6.1 Kennis en self-efficacy beliefs

Voor de analyse van de vragenlijsten hebben we vier afhankelijke variabelen gevormd: 'Kennis', een gemiddelde per vragenlijst per leerling van de 23 items kennis*zekerheid; 'Definitiekwaliteit', een somscore van de conceptuele rijkdom van de definitie van een alinea; 'Onderkenning belang', een somscore van de kenmerken van het belang van alinea's; 'Self-efficacy', een gemiddelde van de 14 items over *self-efficacy beliefs* met betrekking tot het schrijven van alinea's. Deze afhankelijke variabelen zijn continu; binnen een bepaald interval kan de variabele elke waarde aannemen.

Voor de vier variabelen zijn steeds zes modellen gespecificeerd. In model 0 schatten we de gemiddelde score (de intercept) en de variantie tussen de scores. Daarna voegen we in model 1 het leerlinglevel toe om te zien wat de variantie is tussen de leerlingen en in model 2 voegen we het klaslevel toe, zodat we weten wat de variantie is tussen de klassen. In model 3 voegen we vervolgens de predictoren leerjaar (0 = 2-havo, 1 = 5-havo), conditie (0 = controle, 1 = interventie) en meetmo-

³⁶ Meetmoment hebben we niet als level maar als predictor in de analyses verdisconteerd.

³⁷ Daar leerlingen slechts twee teksten schreven (een voormeting en een nameting), vallen het tekst- en leerlinglevel samen in de analyses.

ment (0 = voormeting, 1 = tussenmeting, 2 = nameting) toe³⁸ en de interactietermen leerjaar*meetmoment en leerjaar*conditie. Hierbij zijn leerjaar en conditie predictoren op het niveau van de klas en is meetmoment een predictor op het laagste niveau. In model 4 voegen we de interactie conditie*meetmoment toe om te zien of het effect van het meetmoment verschilt tussen de condities. Tot slot voegen we in model 5 de interactieterm leerjaar*conditie*meetmoment toe waardoor zichtbaar wordt of de eventuele interactie tussen conditie en meetmoment verschilt per leerjaar. De vergelijking tussen de modellen 4 en 3 laat zien wat het effect is geweest van het onderwijs met de toolbox. De vergelijking tussen de modellen 5 en 4 laat zien of dit eventuele effect tussen de leerjaren verschilt. *Random slope*-varianties zijn niet aan de modellen toegevoegd, omdat ze het *fixed part* van het model niet veranderen (Hox et al., 2018).

3.6.2 Vaardigheden

Bij de analyse van de alinea's hebben we gewerkt met de elf alineanormen als afhankelijke variabelen:

- 1a: de tekst bevat alinea's die consistent worden gemarkeerd
- 1b: de tekst bevat alinea's
- 2: de alinea bestaat uit twee of meer zinnen
- 3: de alinea heeft een herkenbare functie
- 4: de alinea heeft samenhang met centrale boodschap
- 5: de kernzin in de alinea ondersteunt de centrale boodschap
- 6: in de alinea is één deelonderwerp aanwezig
- 7: in de alinea is een kernzin aanwezig
- 8: overige informatie in de alinea sluit aan bij de kernzin
- 9a: nieuwe informatie in de alinea wordt gekoppeld aan kennis van woordbetekenissen
- 9b: nieuwe informatie in de alinea wordt gekoppeld aan kennis van de wereld
- 10a: de alinea bevat geen fouten met verwijzwoorden
- 10b: de alinea bevat geen fouten met verbindingswoorden
- 11: de alinea heeft een piramideopbouw

Deze afhankelijke variabelen zijn dichotoom (een alinea voldoet niet (0) of wel (1) aan een norm) en volgen daardoor niet de normaalverdeling. Om deze reden hebben we gebruikgemaakt van de logit transformatie (in plaats van de proportie 'p' wordt de natuurlijke logaritme van $p/(1-p)$ gebruikt). Zoals gebruikelijk in MLwiN hebben we in de logistische regressieanalyses eerst first order MQL (marginal quasi-likelihood) geschat om startwaarden te verkrijgen voor analyses met PQL (predictive quasi-likelihood). Indien bij second order PQL het model niet convergeert, gebruiken we de ruwere schatters van MQL. Dit bleek het geval bij de alineanormen 2, 9a en 9b binnen het genre betogende tekst. Om voor deze gevallen te kunnen bepalen hoe ver de

³⁸ De variabelen Definitiekwaliteit en Onderkenning van het belang (afgeleid uit de opdrachten uit de vragenlijst) kennen twee meetmomenten: voormeting (0) en nameting (1).

MQL-schattingen afwijken van de PQL-schattingen hebben we voor een qua verdeling vergelijkbare alineanorm die wel PQL geschat kon worden ook de MQL-schattingen uitgevoerd. De verschillen tussen beide schattingen bleken klein. Het eindmodel liet bij beide schattingsmethoden voor steeds dezelfde predictoren vergelijkbare effecten zien. In tegenstelling tot de analyse van de gegevens uit de voormeting die we gebruikten als nulmeting (zie hoofdstuk 3) hebben we de gegevens voor de beide genres apart geanalyseerd.

We analyseren de variabelen 1a en 1b via unilevel logistische regressieanalyse in SPSS (versie 26), nadat we hebben vastgesteld dat er in beide genres voor beide normen geen sprake is van significante klasegebonden variantie. De analyses vertrekken per genre vanuit een nulmodel waarin alleen de intercept geschat wordt en geen predictoren zijn opgenomen. In model 1 wordt de voormeting aan de vergelijking toegevoegd, in model 2 de predictoren conditie en leerjaar en in model 3 de interactie conditie*leerjaar. Als conditie significant is, is er verschil in vooruitgang op deze variabelen tussen beide condities. Als leerjaar significant is, is er verschil in de leerwinst tussen 2-havo en 5-havo. Als de interactie significant is, verschilt het effect van de conditie voor beide leerjaren.

De overige variabelen (alineanorm 2 tot en met 11) analyseren we op alineaniveau en, zoals eerder aangegeven, in MLwiN. Voor elke variabele hebben we per genre vier modellen gespecificeerd. In model 0 schatten we de gemiddelde score (de intercept) en de variantie tussen de scores. Daarna voegen we in model 1 het leerlinglevel toe om te zien wat de variantie is tussen de leerlingen en in model 2 voegen we het klaslevel toe, zodat we weten wat de variantie is tussen de klassen. In model 3 voegen we vervolgens de predictoren leerjaar (0 = 2-havo, 1 = 5-havo), conditie (0 = controle, 1 = interventie) en meetmoment (0 = voormeting, 1 = nameting) toe en alle interactietermen. Hierbij zijn leerjaar en conditie predictoren op klasniveau en meetmoment is een predictor op alineaniveau. De interactie conditie*meetmoment laat zien wat het effect is geweest van het onderwijs met de toolbox. De interactie leerjaar*conditie*meetmoment in model 3 laat zien of dit eventuele effect tussen de leerjaren verschilt. *Random slope*-varianties zijn niet aan de modellen toegevoegd, omdat ze het *fixed part* van het model niet veranderen (Hox et al., 2018).

Voor de goede orde merken we op dat onze analysestrategie bij de vaardigheden afwijkt van de aanpak bij kennis en *self-efficacy beliefs*. De aanpak bevat een kleiner aantal stappen en laat zich korter rapporteren, wat gezien het grote aantal afhankelijke variabelen aantrekkelijk is. Voor de uiteindelijke uitkomsten en de te trekken conclusies maakt dit geen verschil.

3.6.3 Significantiebepalingen

Voor het bepalen van de significantie van de levels hebben we bij de analyses van de vier variabelen voor kennis en *self-efficacy beliefs* gewerkt met de likelihoodratietest. Bij de analyses van de alinea's is de likelihoodratietest niet mogelijk, omdat deze variabelen dichotoom zijn. Om die reden hebben we de significantie van de levels getest via het *Intervals and tests*-scherm in MLwiN. Voor het bepalen van de significantie van de regressiecoëfficiënten hebben we gebruikgemaakt van de Waldtest. Als $N > 50$, dan interpreteren we de uitkomsten van de Waldtest als z-score. N is hierbij het aantal waarnemingen binnen het level waarop de variabele varieert (alinea's, leerlingen, klassen). Als $N < 50$ (alleen op het niveau van de klassen) is de uitkomst geïnterpreteerd als t-waarde met als aantal vrijheidsgraden het aantal waarnemingen minus het aantal predictoren minus 1.

4 Resultaten

Deze paragraaf bevat de resultaten van de analyses. In paragraaf 4.1 beschrijven we de uitkomsten van de multilevel regressieanalyses voor de variabelen Kennis, Definitiekwaliteit, Onderkenning belang en Self-efficacy. In paragraaf 4.2 volgen de resultaten van de unilevel logistische regressieanalyses voor norm 1a en 1b op tekstniveau, gevolgd door die van de multilevel logistische regressieanalyses voor de alinea-normen 2 tot en met 11 voor de alinea's. In paragraaf 4.3 beschrijven we enkele nadere exploraties van de data.

4.1 Kennis en self-efficacy beliefs

In tabel 22 staan per geanalyseerd model de fit en passingsvergelijking voor de afhankelijke variabelen Kennis, Definitiekwaliteit, Onderkenning van het belang en Self-efficacy beschreven. Uit de tabel wordt duidelijk dat toevoeging van het leerlinglevel in model 1 ten opzichte van model 0 steeds een verbetering van de fit betekent, behalve voor Onderkenning belang³⁹. Hetzelfde geldt voor toevoeging van het klaslevel in model 2. Voor alle vier de afhankelijke variabelen verdient model 3 de voorkeur boven de andere modellen. We hebben daarom alle vervolganalyses uitgevoerd met modellen waarin zowel het leerling- als het klaslevel zijn opgenomen. Toevoeging van de interactieterm *conditie*meetmoment* in model 4 levert voor geen van de variabelen een significante verbetering van de passing van het model op. Ook de toevoeging van de drieweginteractieterm in model 5 levert voor de vier variabelen geen significante verbetering van de fit op.

³⁹ Bij de variabele Onderkenning belang hebben we het klas- en leerlinglevel (ondanks niet significante resultaten in model 1 en 2) behouden, omdat de predictoren die we vervolgens opnemen zich bevinden op klasniveau.

TABEL 22

Vergelijking van de passinggegevens voor de afhankelijke variabelen Kennis, Definitiekwaliteit, Onderkenning belang en Self-efficacy

Variabele en omschrijving van modellen	-2loglikelihood	Vergelijking			
		Modellen	ΔX^2	Δdf	p
<i>Kennis</i>					
Model 0 alleen intercept	7371.156				
Model 1 met leerlinglevel	7176.896	1 vs 0	194.26	1	<.001
Model 2 met klaslevel	7098.368	2 vs 1	78.528	1	<.001
Model 3 met alle predictoren en interacties L*C en L*M	7045.314	3 vs 2	53.054	5	<.001
Model 4 met interactie C*M	7042.696	4 vs 3	2.618	1	.624 ^{ns} ⁴⁰
Model 5 met interactie C*L*M	7042.679	5 vs 4	.017	1	.999 ^{ns}
<i>Definitiekwaliteit</i>					
Model 0 alleen intercept	2067.895				
Model 1 met leerlinglevel	2058.024	1 vs 0	9.871	1	.002
Model 2 met klaslevel	2042.599	2 vs 1	15.425	1	<.001
Model 3 met alle predictoren en interacties L*C en L*M	2012.559	3 vs 2	30.040	5	<.001
Model 4 met interactie C*M	2010.315	4 vs 3	2.244	1	.134 ^{ns}
Model 5 met interactie C*L*M	2010.291	5 vs 4	.024	1	.877 ^{ns}
<i>Onderkenning belang</i>					
Model 0 alleen intercept	910.547				
Model 1 met leerlinglevel	907.998	1 vs 0	2.549	1	.110 ^{ns}
Model 2 met klaslevel	907.998	2 vs 1	0	1	1 ^{ns}
Model 3 met alle predictoren en interacties L*C en L*M	891.131	3 vs 2	16.867	5	.005
Model 4 met interactie C*M	890.298	4 vs 3	.833	1	.361 ^{ns}
Model 5 met interactie C*L*M	890.089	5 vs 4	.209	1	.648 ^{ns}
<i>Self-efficacy</i>					
Model 0 alleen intercept	6585.820				
Model 1 met leerlinglevel	6425.954	1 vs 0	159.866	1	<.001
Model 2 met klaslevel	6417.099	2 vs 1	8.855	1	.003
Model 3 met alle predictoren en interacties L*C en L*M	6300.446	3 vs 2	116.653	5	<.001
Model 4 met interactie C*M	6299.883	4 vs 3	.563	1	.453 ^{ns}
Model 5 met interactie C*L*M	6296.963	5 vs 4	2.920	1	.087 ^{ns}

⁴⁰ In alle tabellen zijn de niet significante resultaten gemarkeerd met de letters 'ns'.

TABEL 23

Proportie variantie op leerlingniveau (PCL) en op klasniveau (PCK), de intercept (I) en de regressiecoëfficiënten voor (interacties van) leerjaar (L), conditie (C) en meetmoment (M) (met tussen haakjes de *standard error*), de proportie verklaarde variantie op klasniveau (Nagelkerke R^2_{klas})

Variabelen	Model 1		Model 2		Model 3						R^2_{klas}
	PCL	PCL	PCK	I	L	C	M	L*C	L*M		
Kennis	.525	.314	.207	9.666 (4.144)	6.316 (1.036)	-12.604 (5.295)	7.282 (1.927)	3.482 (1.294)	-1.111 (.474)	1	
Definitiekwaliteit	.211	.137	.063	2.535 (.455)	.266 (.114)	1.749 (.365)	-.766 ^{ns} (.602)	-.434 (.091)	.236 ^{ns} (.147)	.707	
Onderkenning belang	.107	.107	0	.941 (.118)	.050 ^{ns} (.030)	-.112 ^{ns} (.147)	.464 (.131)	.041 ^{ns} (.036)	-.118 (.034)	0	
Self-efficacy	.491	.443	.046	65.395 (3.070)	.677 ^{ns} (.768)	-6.777 ^{ns} (4.084)	7.012 (1.165)	2.470 (.996)	-.661 (.287)	.926	

In tabel 23 zijn voor de vier variabelen de regressiecoëfficiënten en de verklaarde variantie op klasniveau uit model 3 weergegeven en de varianties op leerling- en klasniveau uit model 1, 2 en 3. Model 1 laat zien dat er vooral voor Kennis (.525 is de hoogste score) en Self-efficacy (.491) sprake is van substantiële variantie tussen leerlingen. Van variantie tussen klassen is vooral sprake bij Kennis (.207; de overige waarden zijn hier lager).

In tabel 23 zien we voor de variabelen Kennis en Definitiekwaliteit effecten van leerjaar en conditie. Leerlingen in 5-havo scoren gemiddeld beter dan de leerlingen in 2-havo. Leerlingen in de experimentele conditie scoren gemiddeld beter op Definitiekwaliteit en slechter op Kennis dan leerlingen uit de controleconditie. Voor de variabelen Onderkenning belang en Self-efficacy is er bij leerjaar en conditie geen sprake van een significant effect. Daarnaast zien we voor de variabelen Kennis, Onderkenning belang en Self-efficacy een effect van meetmoment. Voor deze variabelen geldt dus dat de resultaten in de nameting hoger zijn dan in de voormeting. Het effect van leerjaar*meetmoment is voor drie van de vier variabelen significant en negatief. Dat betekent dat de verschillen tussen de twee meetmomenten voor leerjaar 2 groter zijn dan voor leerjaar 5. Het effect van de interactieterm leerjaar*conditie is significant positief voor de variabelen Kennis en Self-efficacy en significant negatief voor de variabele Definitiekwaliteit. Dat betekent voor de variabelen Kennis en Self-efficacy dat de leerlingen in de experimentele groepen in 5-havo op beide meetmomenten hoger scoorden dan de overige groepen leerlingen (5-havo controlegroep en 2-havo controlegroep en experimentele groep), terwijl voor Definitiekwaliteit juist het omgekeerde het geval was. De varianties op klasniveau ten slotte worden in model 3 voor Kennis, Self-efficacy en Definitiekwaliteit grotendeels tot helemaal verklaard.

4.2 Vaardigheden

De unilevel logistische regressieanalyses in SPSS laten voor norm 1a ('De tekst bevat alinea's die consistent worden gemarkeerd') voor de e-mail en het betoog zien dat model 1 beter bij de data past dan model 0 ($\chi^2 = 10.467$; $df = 1$; $p = .001$ respectievelijk $\chi^2 = 29.966$; $df = 1$; $p < .001$). Model 2 en model 3 geven voor de e-mail geen verbetering van de fit ten opzichte van respectievelijk model 1 ($\chi^2 = 2.879$; $df = 2$; $p = .237$) en 2 ($\chi^2 = .441$; $df = 1$; $p = .507$). Voor het betoog is de fit van model 2 beter dan die van model 1 ($\chi^2 = 6.140$; $df = 2$; $p = .046$). Model 3 past echter niet beter dan model 2 ($\chi^2 = .349$; $df = 1$; $p = .555$).

Voor norm 1b ('De tekst bevat alinea's') laten de unilevel logistische regressieanalyses in SPSS zien dat voor de e-mail model 1 beter past dan model 0 ($\chi^2 = 12.747$; $df = 1$; $p < .001$) en dat model 2 beter past dan model 1 ($\chi^2 = 10.344$; $df = 2$; $p = .006$). Model 3 levert geen verbetering op ten opzichte van model 2 ($\chi^2 = .767$; $df = 1$; $p = .381$). Voor de betogende teksten convergeren de modellen niet, doordat er vrijwel geen variantie is in de afhankelijke variabele. Op een na bevatten alle teksten alinea's.

De resultaten van deze analyses zijn in tabel 24 samengevat. Bij beide genres is voor norm 1a ('De tekst bevat alinea's die consistent gemarkeerd zijn') sprake van significante effecten van de voormeting. Bij het genre van de betogende teksten zien we dat er ook voor de predictor leerjaar een significant effect optreedt. Voor norm 1b ('De tekst bevat alinea's') geldt dat bij de e-mail significante resultaten zichtbaar zijn voor de voormeting en de conditie. Het negatieve hoofdeffect van conditie laat zien dat de experimentele groepen (in beide leerjaren) na controle voor de voormeting slechter in staat zijn alinea's aan te brengen dan de controlegroepen.

TABEL 24

Intercept (I) en statistisch significante regressiecoëfficiënten voor voormeting en (interactie van) conditie (C) en leerjaar (L) (met tussen haakjes de *standard error*) en de proportie verklaarde variantie op leerlingniveau (Nagelkerke R^2) bij unilevel logistische regressieanalyse in SPSS; schattingen uit model 1 voor e-mail norm 1a en uit model 2 voor e-mail en betoog norm 1b

Genre	Norm	I	voormeting	L	C	L*C	R^2 - leerling
E-mail	1a	-.059 ^{ns} (.243)	1.074 (.337)				.085
	1b	3.060 (.955)	1.777 (.638)	-1.470 ^{ns} (.828)	-1.523 (.663)		.289
Betoog	1a	-.682 ^{ns} (.455)	2.527 (.433)	-1.009 (.429)	.164 ^{ns} (.395)		.290

In tabel 25 zijn de resultaten van de multilevel logistische regressieanalyses van de normen 2 tot en met 11 samengevat voor de e-mails. Deze resultaten zijn ontleend aan de modellen 1, 2 en 3. Voor alle variabelen geldt dat model 1 beter paste dan model 0, model 2 beter dan 1 en model 3 beter dan 2. Zie voor een toelichting hierop bijlage 5. Uit de modelpassingen kunnen we opmaken dat er voor alle alineanormen in de e-mails significante variantie is op het leerling- en klasniveau. Tabel 25 laat zien dat er in model 2 na toevoegen van de significante klasniveaus de variantie op leerling- en klasniveau ongeveer gelijk is. We kunnen de verschillen tussen leerlingen dus voor ongeveer de helft toeschrijven aan het feit dat zij in verschillende klassen zitten.

TABEL 25 E-mails:

proportie variantie op leerlingniveau (PCL) en op klasniveau (PCK), de intercept (I) en de statistisch significante regressiecoëfficiënten voor (interacties van) leerjaar (L), conditie (C) en meetmoment (M) (met tussen haakjes de *standard error*) en de totale door model 3 verklaarde proportie variantie (Nagelkerke R²)

Nr.	Model 1			Model 2			Model 3					R ²
	PCL	PCL	PCK	I	L	C	M	L*C	L*M	C*M	L*C*M	
2	.499	.185	.293	.659 (.248)	2.373 (.373)							.226
3	.271	.127	.126	-.086 (.240)	2.292 (.339)		.541 (.251)		-1.497 (.361)			.151
4	.611	.238	.348	.443 (.367)	2.421 (.524)		-1.071 (.325)		1.531 (.504)			.301
5	.235	.072	.134	-.969 (.243)	1.585 (.281)		-.900 (.161)					.186
6 ⁴¹	.187	.086	.092	-.504 (.220)	1.440 (.276)		-.019 ^{ns} (.249)		-.499 ^{ns} (.316)			.093
7	.189	.069	.113	-1.163 (.186)	1.301 (.232)							.100
8	.175	.062	.106	-1.176 (.174)	1.274 (.217)							.098
9a	.486	.176	.292	.636 (2.315)	2.315 (.383)							.220
9b	.487	.167	.301	.594 (.277)	2.356 (.399)							.227
10a	.127	.059	.060	-.276 (.212)	1.465 (.271)		.262 ^{ns} (.238)		-1.002 (.310)			.196
10b	.249	.127	.109	-.155 (.167)	1.356 (.223)							.103
11	.157	.035	.110	-1.369 (.179)	1.289 (.219)							.101

Uit de resultaten van model 3 blijkt dat er voor alle alineakenmerken (norm 2 tot en met 11) een significant effect is van leerjaar; de alinea's van leerlingen uit 5-havo voldoen vaker aan de normen dan die van leerlingen uit 2-havo. Meetmoment heeft driemaal een significant effect; leerlingen gaan van de voor- naar de nameting vooruit op norm 3 en

41 Bij norm 6 zijn in model 3 meetmoment en leerjaar*meetmoment significant. Na verwijdering van de niet-significante predictoren conditie en leerjaar*conditie worden de predictoren M en L*M niet-significant.

achteruit op norm 4 en 5. De interactie leerjaar*meetmoment heeft ook driemaal een significant effect; het verschil tussen voor- en nameting is tweemaal groter voor 2-havo (norm 3 en 10a) en eenmaal voor 5-havo (norm 4). Daarnaast zien we in model 3 geen (interactie-)effecten van conditie. Dat betekent dat er geen effect te zien is van het werken met de toolbox.

In tabel 26 zijn de resultaten van de multilevel logistische regressieanalyses van de normen 2 tot en met 11 samengevat voor de betogende teksten. Voor alle variabelen geldt dat model 1 beter past dan model 0, maar dat model 2 niet beter past dan model 1. Zie voor een toelichting op de fit en passingsvergelijking bijlage 5. De p-waarden voor de modelvergelijkingen zijn niet significant. Daarom ontbreken in de tabel voor model 2 de schattingen van de PCL en PCK en hebben we voor alle normen in model 3 het klasniveau achterwege gelaten. Model 3 past voor alle variabelen beter dan model 2.

TABEL 26 Betogende teksten:

proportie variantie op leerlingniveau (PCL) en op klasniveau (PCK), de intercept (I) en de statistisch significante regressiecoëfficiënten voor (interacties van) leerjaar (L), conditie (C) en meetmoment (M) (met tussen haakjes de *standard error*) en de totale door model 3 verklaarde proportie variantie (Nagelkerke R²)⁴²

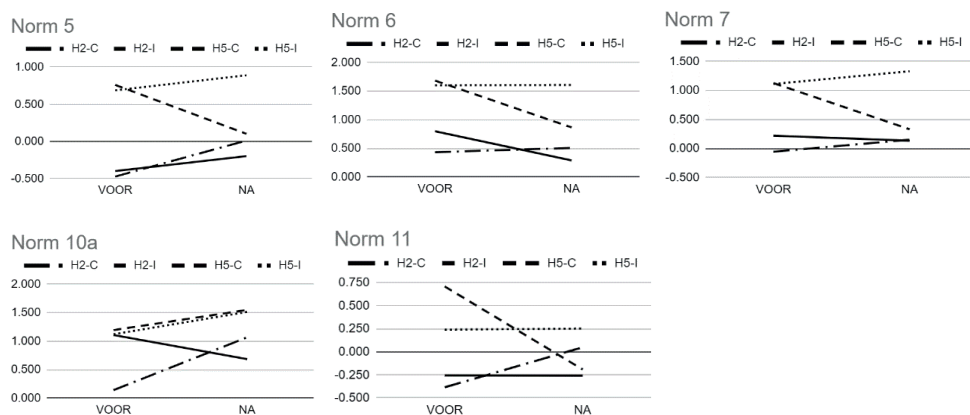
Nr.	Model 1			Model 2			Model 3					R ²	
	PCL	PCL	PCK	I	L	C	M	L*C	L*M	C*M	L*C*M		
2	.486	-	-	1.470 (.185)	1.889 (.318)							.138	
3	.257	-	-	.584 (.228)	1.852 (.378)		.397 ^{ns} (.331)		-1.100 (.501)			.098	
4	.339	-	-	.852 (.229)	1.431 (.346)							.068	
5	.228	-	-	-.473 (.277)	1.233 (.405)		.074 ^{ns} (.387)	.484 ^{ns} (.289)	-.147 ^{ns} (.549)	-1.143 (.416)	-.283 ^{ns} (.410)	1.145 (.570)	.047
6	.210	-	-	.432 (.257)	1.252 (.399)		.366 ^{ns} (.363)	.074 ^{ns} (.282)	-.448 ^{ns} (.541)	-.893 (.429)	-.584 ^{ns} (.401)	1.410 (.592)	.059
7	.193	-	-	-.056 (.247)	1.183 (.367)		.276 ^{ns} (.344)	.208 ^{ns} (.273)	-.291 ^{ns} (.498)	-1.004 (.399)	-.294 ^{ns} (.387)	1.310 (.552)	.055
8	.179	-	-	.439 (.100)	1.802 (.308)							.156	
9a	.472	-	-	1.466 (.181)	1.802 (.308)							.156	
9b	.399	-	-	1.399 (.170)	1.328 (.268)							.080	
10a	.147	-	-	.139 (.222)	1.055 (.333)		.969 (.325)	.928 (.274)	-1.042 (.460)	-.573 ^{ns} (.408)	-1.352 (.398)	1.390 (.561)	.159
10b	.221	-	-	.726 (.181)	.628 (.240)			.794 (.160)					.045
11	.135	-	-	-.384 (.222)	1.092 (.319)		.128 ^{ns} (.309)	.433 ^{ns} (.262)	-.597 ^{ns} (.430)	-1.332 (.369)	-.435 ^{ns} (.374)	1.347 (.505)	.031

42 Bij de normen 3, 5, 6, 7, 10a en 11 ontstaan in model 3 bij verschillende predictoren en interactie-termen niet significante regressiecoëfficiënten die wel in het model behouden blijven, omdat de bovenliggende interactie-termen wel significant zijn.

Uit de resultaten van model 3 blijkt dat er voor vrijwel alle alineakenmerken (behalve norm 8) een significant effect is van leerjaar; de alineas van leerlingen uit 5-havo voldoen vaker aan de normen dan die van leerlingen uit 2-havo. Conditie heeft eenmaal een significant effect; de alineas van de leerlingen uit de experimentele conditie voldoen vaker aan norm 10a dan die van de leerlingen uit de controleconditie. Meetmoment heeft tweemaal een significant effect; leerlingen gaan van de voor- naar de nameting vooruit op norm 10a en 10b. De interactie leerjaar*conditie heeft tweemaal een significant effect; de verschillen tussen de condities zijn voor de normen 3 en 10a in 5-havo kleiner dan in 2-havo. De interactie leerjaar*meetmoment heeft viermaal een significant effect; het verschil tussen voor- en nameting is groter voor norm 5, 6, 7 en 11 groter voor 2-havo dan voor 5-havo. Voor conditie*meetmoment zien we voor norm 10a een significant negatief effect. Dit interactie-effect moeten we interpreteren in samenhang met de bijbehorende positieve interactie leerjaar*conditie*meetmoment. Voor norm 10a heffen deze coëfficiënten elkaar op, wat aangeeft dat het werken met de toolbox voor norm 10a negatief heeft uitgedrukt voor 2-havo en geen effect had voor 5-havo. Voor de alineakenmerken 5, 6, 7 en 11 vinden we ook een significant positief interactie-effect tussen leerjaar, conditie en meetmoment; dit betekent dat de resultaten van de controlegroep na de interventie verslechterd is en dat de resultaten voor de experimentele groep min of meer stabiel blijven. In figuur 9 hebben we voor de normen 5, 6, 7, 10a en 11 de resultaten visueel weergegeven.

FIGUUR 9

Grafische weergave van de gemiddelde resultaten van de voor- en de nameting van vijf alineanormen in de betogende teksten, waarbij H2 = leerlingen van 2-havo, H5 = leerlingen van 5-havo, I = interventiegroep en C = controlegroep



Uit de tabellen en grafieken zien we geen overtuigende aanwijzingen voor verschillen tussen de experimentele conditie en de controleconditie. In de experimentele conditie

van 5-havo kunnen we een klein effect zien, maar deze lijn voor 5-havo loopt (vrijwel) parallel met de lijnen voor de beide condities voor 2-havo. Opvallend is vooral dat de controleconditie van 5-havo bij vier van de vijf normen een duidelijke achteruitgang laat zien. De verschillen tussen de twee condities in 2-havo zijn minder groot dan bij 5-havo.

4.3 Vragen bij deze resultaten

We kunnen op basis van de hierboven vermelde resultaten niet constateren dat het gebruik van de ontwikkelde genrespecifieke alineatoolboxen effectief is gebleken. We zien bij de e-mail geen significante interactie-effecten waarin conditie een rol speelt en bij het betoog hebben de significante interacties van conditie met meetmoment en leerjaar vooral te maken met dalende scores op de nameting voor de controlegroep in 5-havo. Er zijn bij deze resultaten verschillende vragen te stellen. We exploreren er enkele.

4.3.1 Exploratie 1; wat is het effect van de ontbrekende data?

Hoewel van de meeste deelnemers aan deze effectstudie zowel teksten in beide genres en van beide meetmomenten als vragenlijsten in de dataset opgenomen zijn, waren er ook ontbrekende data, doordat leerlingen ziek of afwezig waren.

Er zijn enkele verklaringen voor de ontbrekende data te geven. Ten eerste heeft 2-havo in de voormeting van de e-mail moeten wennen aan de werkwijze. Het digitaal inleveren via de elektronische leeromgeving van de eerste versies was niet voor elke leerling duidelijk. Ondanks reminders heeft dat ervoor gezorgd dat er bij de eerste meting van 2-havo data verloren zijn gegaan. Daarnaast kreeg diezelfde jaarlaag tijdens het schrijven van de betogende teksten te maken met de uitbraak van de COVID-19-pandemie. De afronding van de herschrijvingen viel precies in de week dat het voortgezet onderwijs gesloten werd. Ook bij de leerlingen van 5-havo speelt de sluiting van de scholen een rol bij de klassikale afname van de vragenlijst in de nameting. We hebben alle leerlingen van wie de vragenlijst ontbrak, digitaal benaderd met een persoonlijk bericht met het verzoek de lijst alsnog in te vullen en terug te sturen. Daarop kwamen reacties, maar zeker niet van alle leerlingen. Uit de interviews die we gehouden hebben met de docenten na afloop van de lessen, bleek dat de aandacht waarmee de leerlingen de vragenlijsten invulden te wensen overliet.

We kunnen nagaan of de ontbrekende data effect gehad hebben op de resultaten door van het al dan niet ontbreken van de nameting een dichotome variabele te maken ('missing na', afgekort tot 'MN') die we gebruiken in analyses van de gegevens van de voormeting. De extra predictor voegen we toe aan model 3⁴³, zodat we een

43 De aanduidingen van de modellen bij de exploraties corresponderen met die van de modellen waarover gerapporteerd is in paragraaf 4.1 en 4.2.

mogelijk effect van deze predictor zichtbaar kunnen maken. Het resultaat hiervan is dat zowel bij de variabelen uit de vragenlijsten als bij de variabelen bij de e-mails en betogende teksten 'missing na' geen significante predictor vormt. Ook van een interactie tussen 'missing na' en conditie is geen sprake. Zo bezien heeft het ontbreken van de nameting geen invloed gehad op de vergelijking tussen de interventiegroep en de controlegroep. Het uitblijven van effecten van de interventie kan derhalve niet worden toegeschreven aan het ontbreken van gegevens op de nameting en aan de factoren die daaraan debet zijn. De tabellen die horen bij deze exploratie staan in bijlage 6. Daar staan zowel de tabellen met fit en passingsvergelijkingen voor alle normen als de tabellen met de bijbehorende regressiecoëfficiënten voor de e-mails en de betogende teksten.

4.3.2 Exploratie 2; wat is het effect van de hoeveelheid revisie?

In de interviews met de docenten van 5-havo merkten zij op dat leerlingen het herzien van de eerste versie van hun tekst niet zo serieus namen. Deze opmerking roept de vraag op of én op welke manier de leerlingen hun teksten voor de nameting gereviseerd hebben. Mogelijk zien we daar een verklaring voor het niet optreden van effecten van de interventie.

Om dit te onderzoeken hebben we samen met zes studenten in een onderzoeksatelier uit het tweede jaar van de bachelor Nederlandse Taal en Cultuur aan de RUG in kaart gebracht of alle als nameting ingeleverde teksten daadwerkelijk gereviseerd zijn. We hebben hiervoor een tekstvergelijking gemaakt tussen de eerste versie en de revisie met behulp van het programma Compare text online (www.countwordsfree.com). Zo hebben we het percentage verschil tussen beide teksten per leerling in kaart gebracht. Daarbij heeft een van de supervisors een controle uitgevoerd waaruit bleek dat de overeenkomst tussen de eerste en de tweede beoordelaar zeer hoog was ($R = .95$, $N = 24$, $p = .000$).

Het percentage verschil ('percent different' afgekort tot 'PD') hebben we zowel voor de tekstdata (alineanorm 1) als voor de alineadata (alineanormen 2-11) als predictor op het niveau van de leerlingen aan model 3 toegevoegd om te kunnen zien of er interactie-effecten optreden met het meetmoment. Het valt immers te verwachten dat leerlingen die veel gereviseerd hebben betere alinea's schrijven op de nameting dan op de voormeting.

In de resultaten zien we dat de mate van reviseren van de leerlingen in de nameting geen effect heeft op de scores voor de afzonderlijke alineanormen. Zowel bij de e-mails als bij de betogende teksten vormt het percentage revisie geen significante predictor. Zo bezien kan het uitblijven van effecten van de interventie daarom niet

worden toegeschreven aan de mate van revisie door de leerlingen. In bijlage 6 van dit proefschrift staan de bij deze exploratie horende tabellen met fit en passingsvergelijkingen en met regressiecoëfficiënten voor alle normen in beide genres.

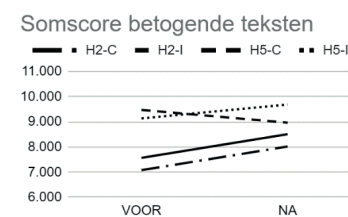
4.3.3 Exploratie 3; wat is het effect van werken met somscores?

Tot nu toe hebben we de alineanormen steeds afzonderlijk bekeken. We hebben geredeneerd dat elke norm een andere deelvaardigheid van het schrijven van een alinea bevat. Hier exploreren we wat we over de effectiviteit van de interventie kunnen zeggen als we de alineanormen sommeren. Hiervoor hebben we per genre een somscore samengesteld die bestaat uit de som van de losse scores van de alineanormen 2 tot en met 11. De maximaal te behalen score daarbij bedraagt 12. De alineanormen blijken sterk samen te hangen (e-mails: $\alpha = .94$; betogende teksten: $\alpha = .92$). De resulterende somscore is een afhankelijke variabele die in MLwiN geanalyseerd is⁴⁴.

Voor de e-mails zien we bij de analyse van de somscore dat er alleen een leerjaar-effect is. Dit sluit aan bij de resultaten van de afzonderlijke normen. In de betogende teksten zien we zowel een effect van leerjaar als van meetmoment en ook interactie-effecten van leerjaar*meetmoment en van leerjaar*conditie*meetmoment. In bijlage 6 bij dit proefschrift staat de tabel met de fit en passingsvergelijking voor de somscores en daarnaast ook de bijbehorende regressiecoëfficiënten voor de somscores in beide genres en de onder model 3 verwachte gemiddelde somscore voor beide genres.

FIGUUR 10

Grafische weergave van de gemiddelde somscores van de betogende teksten



De analyses resulteren in een gemiddelde somscore per genre. In 2-havo zien we dat de nametingen gemiddeld één scorepunt toenemen ten opzichte van de voormetingen.

44 Omdat we hier te maken hebben met intervalvariabelen zijn de regressieanalyses in MLwiN niet logistisch. In dergelijke modellen toetsen we fitverbetering door na te gaan of het verschil in deviance ($-2 \times \log$ likelihood) tussen de beide geneste modellen (met en zonder het betreffende effect) significant is. Het verschil in deviance is chi-kwadraat verdeeld met als aantal vrijheidsgraden het verschil in aantal geschatte parameters tussen beide modellen. De gevonden overschrijdingskans moet bij deze toetsing van variantielevels door twee gedeeld worden, omdat varianties niet negatief kunnen worden (Hox et al., 2018).

Daarnaast scoort de experimentele groep (8.506) dus gemiddeld hoger op de somscore dan de controlegroep (8.017). In 5-havo zien we dat de interactie-effecten een rol spelen; bij de experimentele groep neemt de gemiddelde score toe in de nameting (9.683), terwijl bij de controlegroep de nameting juist gemiddeld lager scoort (8.962). De alinea's van de leerlingen uit de interventiegroep gaan vooruit, terwijl de alinea's van de leerlingen uit de controlegroep achteruitgaan.

De alinea's van de leerlingen in 2-havo die een betogende tekst schrijven gaan allemaal vooruit, ongeacht de conditie. Bij de alinea's van de leerlingen in 5-havo is de winst in de interventiegroep kleiner dan in 2-havo, terwijl de leerlingen van 5-havo in de controlegroep achteruitgaan. Het effect kan hier dus niet toegeschreven worden aan de interventie. In figuur 10 hierboven zijn de resultaten voor de somscore van de betogende teksten visueel gemaakt.

4.3.4 Exploratie 4; wat is het effect van de wijze waarop leerlingen met de zelfevaluaties en hulp op maat gewerkt hebben?

Tot slot gaven de docenten in de interviews die we na afloop van de lessenreeksen gehouden hebben aan dat de schrijfvaardigheidslessen nogal onbevredigend waren, omdat het de leerlingen ontbrak aan interne motivatie. Tijdens de lessen werkten de leerlingen van de experimentele groepen met een werkboekje waarin zij oefenden met zelfevaluaties en hulp op maat. De docenten constateerden dat de leerlingen met name de hulp-op-maatopdrachten niet of beperkt maakten.

Om na te gaan of de mate van invulling van de zelfevaluaties en de hulp-op-maatopdrachten inderdaad effect gehad heeft op de resultaten hebben we een vierde exploratie uitgevoerd. Hiervoor hebben we alle ingeleverde werkboekjes van de leerlingen uit de experimentele groepen ($N = 171$) voor zowel de e-mails als voor de betogende teksten gecontroleerd op de wijze van invulling van de zelfevaluatie en de hulp op maat. Deze percentages zelfevaluatie en hulp op maat ('PZenH') hebben we zowel voor de tekstdata (alineanorm 1) als voor de alineadata (alineanormen 2-11) voor beide genres als predictor toegevoegd in model 3 om te kunnen zien of er interactie-effecten optreden met het meetmoment. Het valt immers te verwachten dat leerlingen die een hoge score hebben op het percentage zelfevaluatie en hulp op maat betere alinea's schrijven in de nameting.

De resultaten van deze exploratie (zie de tabellen in bijlage 6 van dit proefschrift) laten zien dat er geen effect is van het maken van de zelfevaluaties en hulp-op-maatopdrachten bij de e-mails en de betogende teksten, behalve bij de normen 7 ('De belangrijkste uitspraak over het deelonderwerp staat in de kernzin'), 8 ('De overige informatie sluit aan bij de kernzin') en 10a ('Binnen de alinea worden de zinnen

verbonden tot een samenhangend geheel door het gebruik van verwijswaarden') van dit laatste genre. De normen 7 en 8 laten in de analyses zien dat de interactie-term meetmoment* PZenH significant is. Dit wil zeggen dat leerlingen die meer werk hebben geleverd in de zelfevaluatie en hulp op maat in de nameting van de betogende teksten betere alinea's schreven dan leerlingen die hun werk minder serieus deden.

5 Conclusie en discussie

Deze deelstudie beschrijft het interventieonderzoek waarin we hebben onderzocht wat het effect is van het werken met een alineatoolbox op de kennis, *self-efficacy beliefs* en vaardigheden van de leerlingen bij het schrijven van begrijpelijke alinea's. Eventuele effecten van het werken met de alineatoolboxen zouden in de resultaten zichtbaar moeten worden in de significante positieve interacties tussen conditie en meetmoment. Daaruit blijkt immers dat er in de nameting rekeninghoudend met de voormeting een beter resultaat is in de experimentele groep dan in de controlegroep.

5.1 Conclusie

Onze hypothese was dat leerlingen in beide leerjaren voor beide genres baat zouden hebben bij de interventie en dat de experimentele groep dus beter zou scoren dan de controlegroep. Deze hypothese vindt geen steun in de data. Op basis van de vermelde resultaten kunnen we niet constateren dat het gebruik van de ontwikkelde genrespecifieke alineatoolboxen binnen dit onderzoek effectief is gebleken voor de kennis, *self-efficacy beliefs* en vaardigheden van leerlingen bij het schrijven van alinea's. We zien namelijk voor geen enkele variabele een overtuigend verschil tussen de experimentele groep en de controlegroep. Voor kennis en *self-efficacy beliefs* zien we helemaal geen effect van het werken met de alineatoolboxen. Voor de vaardigheden zien we ook geen effecten, al is er een verschil zichtbaar tussen de onderzochte genres. Bij het schrijven van de e-mails hebben de leerlingen nergens profijt gehad van het werken met de alineatoolboxen; bij de betogende teksten zien we bij de normen 5 ('De kernzin ondersteunt de centrale boodschap van de tekst'), 6 ('De alinea beslaat één deelonderwerp van een tekst'), 7 ('De belangrijkste uitspraak over het deelonderwerp staat in de kernzin'), 10a ('Binnen de alinea worden de zinnen verbonden tot een samenhangend geheel door het gebruik van verwijswaarden') en 11 ('De alinea is opgebouwd als piramide') wel verschillen tussen de condities. Deze lijken vooral te maken te hebben met achteruitgang van de scores voor de controleconditie in 5-havo.

Bij individuele variabelen zien we wel effecten van leerjaar, van meetmoment en van interacties van de verschillende predictoren. We bespreken deze effecten hier achtereenvolgens voor de variabelen uit de vragenlijst en voor de alineanormen per genre.

Er zijn verschillen zichtbaar tussen de kennis en het zelfvertrouwen van leerlingen in 2-havo en 5-havo. Onafhankelijk van het reguliere of experimentele onderwijs dat de leerlingen kregen, weten leerlingen in 5-havo meer over de alinea en ze zijn beter in staat zelf onder woorden te brengen wat een alinea is dan de leerlingen in 2-havo. Kennelijk is er door de jaren heen ontwikkeling in hun kennis over alinea's en leren leerlingen ook over alinea's van het lezen van teksten, wat zij in het onderwijs veelvuldig doen. Daarnaast hebben leerlingen in 5-havo ook meer zelfvertrouwen in het schrijven van alinea's dan de leerlingen in 2-havo. Zowel voor de e-mails als de betogende teksten geldt dat er voor alle alineanormen⁴⁵ een hoofdeffect van leerjaar zichtbaar is; dat wil zeggen dat de leerlingen van 5-havo de alineanormen in beide genres beter gebruiken en beheersen dan de leerlingen in 2-havo. Voor de betogende teksten geldt bij norm 1a ('De tekst bevat alinea's die consistent gemarkeerd zijn) dat het leerjaareffect negatief is. Dat betekent dat de leerlingen van 2-havo de alinea's van de betogende teksten consistent markeren dan de leerlingen van 5-havo. Bij norm 3 ('De functie van de alinea binnen de tekst is herkenbaar') zien we een negatief interactie-effect van leerjaar en conditie. Deze verschillen bij het schrijven van alinea's zijn echter niet toe te schrijven aan de interventie, want ze doen zich al voor bij de voormeting.

Er zijn verschillen zichtbaar tussen de voor- en de nameting. Alle leerlingen (ongeacht het leerjaar of de klas waarin zij zaten) weten in de nameting meer over alinea's en zijn beter in staat onder woorden te brengen wat een alinea is. Mogelijk zijn leerlingen bewust bezig geweest met het onderwijs dat zij kregen – ongeacht de vraag of ze gewerkt hebben met de toolbox, is hun aandacht gevestigd op de alinea en hebben ze van de lessen iets opgestoken. Daarnaast hebben de leerlingen in beide leerjaren in de nameting meer zelfvertrouwen in het schrijven van alinea's. Voor de e-mails is er een interactie-effect leerjaar*meetmoment zichtbaar bij drie alineanormen die zowel positief als negatief zijn. In bijlage 7 hebben we voor het overzicht ook de onder model 3 verwachte gemiddelden voor Kennis, Definitiekwaliteit, Onderkenning belang en Self-efficacy en de onder model 3 verwachte proporties voor de tekst- en alinea-kenmerken opgenomen.

5.2 Discussie

Binnen het onderzoek naar schrijfvaardigheidsonderwijs voor het voortgezet onderwijs is er niet eerder uitgebreid aandacht voor de alinea geweest. We constateerden dat dit opmerkelijk is, omdat alinea's een belangrijke functie hebben in het aanbrengen van samenhang in teksten. Deze eerste verkennende interventiestudie

⁴⁵ Uitzondering zijn norm 1b ('De tekst bevat alinea's') en norm 8 ('De overige informatie sluit aan bij de kernzin'). Voor norm 1b geldt dat het model niet geschat kon worden, omdat er geen variantie was. Dit komt doordat op een na alle betogende tekst alinea's bevatten. Norm 8 geeft in model 3 geen enkele predictor als resultaat.

brengt beperkingen met zich mee en roept verschillende vragen op. We presenteren er hier enkele.

Om te beginnen richten we de aandacht op de betrouwbaarheid, de validiteit en de power van de metingen. Zijn de metingen voldoende betrouwbaar geweest en waren de meetinstrumenten voldoende valide? Voor het analyseren van de twee opdrachten uit de vragenlijst over kennis en *self-efficacy beliefs* hebben we gewerkt met twee beoordelaars. De overeenstemmingspercentages per concept bij het scoren van de definities uit de vragenlijst zijn hoog te noemen, namelijk boven de 89%. De gesloten vragen uit de vragenlijst zijn op betrouwbaarheid geanalyseerd door voor de drie meetmomenten de interne consistentie te meten. Voor het gedeelte van de kennisitems bedraagt de Cronbach α gemiddeld over drie meetmomenten .69. Dat is voldoende betrouwbaar te noemen. De Cronbach α voor het gedeelte over de *self-efficacy beliefs* is hoog; gemiddeld over drie meetmomenten bedraagt deze .91. Ook bij het coderen van de alineanormen in de teksten van de leerlingen hebben we gewerkt met twee beoordelaars. De overeenstemmingspercentages tussen beide beoordelaars zijn gemeten na het beoordelen van een steekproef (van 96 teksten met 327 alinea's) uit de voormeting en een steekproef (van 97 teksten met 384 alinea's) uit de nameting. Het percentage overeenstemming tussen de beoordelaars is goed te noemen; het ligt voor alle alineanormen boven de 84%. Cohens kappa is als maat voor de interbeoordelaarsbetrouwbaarheid voldoende tot goed beoordeeld. De scores liggen tussen .69 en .95. De betrouwbaarheid van al deze metingen is al met al voldoende tot goed te noemen.

De vraag naar de validiteit van de metingen is te beantwoorden door de conceptuele basis van de afhankelijke variabelen in dit onderzoek te bekijken. Deze afhankelijke variabelen komen overeen met de alineanormen die we verantwoord hebben vanuit de wetenschappelijke literatuur. Vervolgens hebben we deze afhankelijke variabelen in de vragenlijsten naar de kennis, *self-efficacy beliefs* en vaardigheden en ook in de ontworpen toolboxes en tekstanalyses laten terugkomen. Deze continuïteit is een waarborg geweest voor de validiteit in dit onderzoek. Binnen het tekstanalytisch deel van dit onderzoek hebben we de alinea's uit de teksten van de leerlingen gescoord op de aanwezigheid van de elf alineanormen. Deze alineanormen laten allemaal een ander aspect van de alinea zien; we hebben dus gewerkt met veel verschillende afhankelijke variabelen. Het werken met veel verschillende variabelen kan tot gevolg hebben dat het geheel uit het oog verloren wordt. Om deze reden hebben we in de derde exploratie gekeken of het werken met somscores de conclusies verandert. Dat is niet het geval. Het voordeel van het analyseren van de losse alineanormen als variabelen heeft ervoor gezorgd dat zichtbaar geworden is welke aspecten van een alinea leerlingen wel of niet beheersen. Dit kan onderzoekers in de toekomst houvast geven bij het uitvoeren van vervolgonderzoek.

Een volgende vraag is of de power van het onderzoek voldoende was. Voorafgaand aan de uitvoering van het onderzoek hebben we met behulp van *GPower 3.1* berekend hoe groot de steekproef moest zijn om klein tot medium effect van .15 te kunnen tonen. Het resultaat van de analyse liet zien dat bij een regressieanalyse met drie predictoren (leerjaar, voormeting en conditie) en een correlatie van .50 tussen vooren nameting een effectieve steekproef van 49 leerlingen per conditie volstond voor een power van .90 en een $\alpha = .05$. De feitelijke steekproefomvang was 137 in de controlegroep en 167 in de experimentele groep. Zelfs bij een intraklassecorrelatie van .06 was de effectieve steekproef in beide condities groter dan de benodigde 49. De power van deze studie was daarmee voldoende groot om een effect met een grootte van .15 aan te kunnen tonen.

Vervolgens kunnen de beperkingen van het onderzoek ook op andere niveaus aanwijsbaar zijn. We hebben geen effecten gevonden van de interventie. Daarmee zouden we in de eerste plaats het theoretisch kader ter discussie kunnen stellen; zijn de geïdentificeerde alineanormen wel van invloed op de kwaliteit van de gehele tekst? In hoeverre zijn alle alineanormen relevant voor het schrijven van een alinea? Gezien de onderbouwde reconstructie van het alineanormenkader in hoofdstuk 2 hebben we geen reden om aan te nemen dat de theoretische achtergrond niet voldoet. Wel zien we dat we er nog niet voldoende in geslaagd zijn om de alineanormen over te brengen op leerlingen en ervoor te zorgen dat leerlingen de normen ook daadwerkelijk toepassen in hun teksten. We onderzoeken hiervoor de implementatietrouw (O'Donnell, 2008); is de interventie door de docenten uitgevoerd zoals deze bedoeld was en in hoeverre hebben de leerlingen de leerprocessen doorlopen zoals beoogd was?

We hebben alle redenen om aan te nemen dat de participerende docenten de lessen uitgevoerd hebben zoals deze van tevoren bedacht waren. Zowel voor de experimentele groepen als voor de controlegroepen zijn voorafgaand aan het onderzoek de lesplannen geheel uitgeschreven. De docenten die participeerden in het onderzoek kregen de lesplannen op papier en de docent-onderzoeker heeft deze mondeling toegelicht. Omdat de interventie plaatsvond op de school van de docent-onderzoeker, was het een bewuste keuze om de lessen niet te observeren om zo controlegedrag te vermijden. Het was onze inschatting dat observaties de leerlingen en de betreffende docenten zouden afleiden. In plaats daarvan is gekozen voor interviews na afloop van een lessenreeks met de docenten en met groepjes leerlingen om zo wel zicht te hebben op de implementatie.

Uit de interviews met de docenten bleken drie interessante punten. Ten eerste waren de begeleidende docenten zich zeer bewust van het belang van de lessen op de voorgescreven manier uit te voeren:

- *'omdat het voor een onderzoek is, wil je je heel strikt houden aan het lesplan'*
- *'ik was me heel erg bewust dat het voor een onderzoek was; ik merkte dat ik heel erg bezig was met de juiste woorden kiezen en dat vond ik intensiever dan gewoon lesgeven. Nu bereidde ik echt alles stap voor stap voor. Als ik iets verkeerd zou doen, zou jij mijn klas niet meer kunnen gebruiken'*
- *'het gaf me een vervelend gevoel [dat leerlingen in de klas niet goed meededen]. Ik heb druk gevoeld vanwege de korte periodes en het gevoel dat ik niet ziek mocht zijn. Er moesten immers wel resultaten van de hele klas uitkomen, want het is voor je onderzoek. Vorig jaar zat ik er meer ontspannen in. Mijn enthousiasme [voor de lessen] was wel hetzelfde en misschien nog wel meer, want vorig jaar was ik heel enthousiast over de lessen'*

(Citaten uit de interviews met de docenten van 5-havo)

Ten tweede hebben de docenten gezocht naar verklaringen voor het feit dat leerlingen (met name in 5-havo) niet goed meededen en hun werk niet inleverden. Ze noemen de volgende verklaringen:

- *'gemakzucht; leerlingen hebben geen zin om er tijd aan te besteden'*
- *'het is geen toets, dus het telt niet mee, er zitten geen consequenties aan'*
- *'de werkhouding is extreem slecht, omdat ze het grotere nut er niet van zien. De interne motivatie om goed werk af te leveren en iets te leren over het schrijfonderwijs lijkt dus te ontbreken'*

(Citaten uit de interviews met de docenten van 5-havo)

Ook de leerlingen werden gevraagd naar verklaringen voor de werkhouding van de leerlingen in 5-havo. Zij bevestigden in de focusgroepinterviews de verklaringen van de docenten:

- *'we hebben het niet nodig. Het is wel belangrijk om een mail te schrijven, maar over het algemeen wordt het wel veel minder gedaan. Ik denk dat je alleen een mail stuurt echt als het moet naar een docent of een sollicitatiegesprek'. Aanvulling van een andere leerling: 'Ja, sinds kort hoef je eigenlijk ook alleen nog maar zo'n vragenlijst in te vullen als je wilt solliciteren bij Albert Heijn ofzo.'*
- *'honderd procent luiheid'. (...) 'Ik denk niet dat het bij elke leerling zo zit maar je hebt hier ook bij ons mensen in de klas die echt gewoon helemaal niks doen. En dat merk je ook niet alleen bij Nederlands maar eigenlijk gewoon bij alles en dan staan ze al helemaal niet te springen om een tekst te gaan lezen. Klinkt eigenlijk heel sneu, maar ja.'*

(Citaten uit de focusgroepinterviews met leerlingen van 5-havo)

Ten derde zagen de docenten zelf vanuit didactisch oogpunt de waarde van het werken met een zelfevaluatie-instrument, maar zij merkten ook dat leerlingen kritisch zijn op hun eigen werk moeilijk vinden:

- *'een leerling neemt zelf verantwoordelijkheid voor zijn werk'*
- *'een leerling werkt alleen aan zijn eigen verbeterpunten; ze zouden echt aan de slag kunnen gaan met het onderdeel waar ze problemen mee hebben en dat zorgt ervoor dat ze dus echt werken aan datgene dat voor hen belangrijk is'.* Dat beeld wordt in de interviews met leerlingen bevestigd. *'Deze manier van werken heeft voordelen; iedereen doet wat hij lastig vindt. Dan zit niet iedereen opgescheept met wat hij misschien makkelijk vindt'.*
- *'leerlingen blijken het gewoon heel erg ingewikkeld te vinden om kritisch naar hun eigen teksten te kijken. Ze zijn snel tevreden en snappen vaak niet waarom ze nog een keer grondig naar hun eigen werk moeten kijken. Ik vraag me af of we hier niet juist wat mee moeten. De vragen die bij de zelfanalyse staan zijn wel concreet, maar voor leerlingen vaak nog heel moeilijk'.* Leerlingen reageren ook uitgebreid op dit punt:
- *'Ik merkte dat ik niet kritisch genoeg was op mijn teksten. Ik dacht gewoon dat ik het allemaal wel wist. Dat komt omdat de alineanormen heel herkenbaar waren. Dat zag ik ook in mijn klas: alle leerlingen zeiden steeds: 'Ja mevrouw, dat weet ik wel'.*
- *'Ik merkte dat ik niet kritisch genoeg ernaar (d.i. de zelfanalyse) had gekeken. Ik vind het heel lastig om dat van mezelf te bepalen. Ik vind het makkelijker om van iemand anders de tekst te beoordelen in plaats van die van jezelf. Ik denk dat je toch altijd minder kritisch naar jezelf kijkt. Ik denk dat dat bij alles is en vooral ook bij Nederlands. De spelfoutjes kan je er wel uithalen maar die opbouw... je denkt dat het echt klopt anders heb je het ook niet zo geschreven dus dan vind je automatisch dat het goed is'.*
- *'Ook de anderen vonden het lastig om zichzelf te beoordelen en dan ook aan een moeilijke opdracht te gaan. Dat vinden mensen toch altijd lastig. Dat zitten ze toch wel een beetje van 'dan kan ik beter wat anders gaan doen of zo'. Het is denk ik confronterend of zo dat je echt denkt ik ben hier slecht in en er zit iemand naast je en die heeft die opdracht echt foutloos gedaan'.*

(Citaten uit de interviews met de docenten van 5-havo en uit de focusgroep-interviews met de leerlingen van 5-havo)

Gegevens over de implementatie kunnen we niet alleen ontleen aan de docent- en leerlinginterviews, maar ook aan de werkboekjes die de leerlingen uit de experimentele groepen invulden. Deze leerlingen hebben zoals eerder aangegeven in een werkboekje (per genre) aangegeven hoe zij zichzelf beoordeelden in de zelfevaluatie, welke conclusie ze daarna getrokken hebben, welke hulp op maat zij gekozen hebben en in welke mate ze deze hulp op maat ook daadwerkelijk gemaakt hebben. Om een uitspraak te kunnen doen over de mate waarin leerlingen de zelfevaluaties en de hulp op maat daadwerkelijk gemaakt hebben, hebben we de zelfevaluaties en hulp-op-maatopdrachten gecontroleerd. Zelfevaluaties en hulp-op-maatopdrachten die gedeeltelijk ingevuld waren, telden mee als ingevuld wanneer zij voor meer dan de helft ingevuld waren, daarvoor ontbrak het ons aan tijd. Ter controle van de betrouw-

baarheid is een steekproef uit het materiaal door twee beoordelaars beoordeeld. De mate van overeenstemming is voor de beoordeling van de zelfevaluaties 99% en voor de hulp-op-maatopdrachten 96%.

Het percentage ingevulde zelfevaluaties was voor de e-mail 80,0% en voor de betogende teksten 71,9%. Het percentage gemaakte hulp-op-maatopdrachten ligt lager, namelijk voor de e-mail op 57,6% en voor de betogende teksten op 41,7%. De laatste twee percentages roepen (opnieuw) vragen op over de inzet van de leerlingen.

Over de manier waarop leerlingen gewerkt hebben met de toolboxes zijn twee opmerkingen hier relevant. Omdat we in dit onderzoek relatief weinig gegevens verzameld hebben over de werkwijze van de leerlingen, is vervolgonderzoek hiernaar zeer wenselijk. Waren de leerlingen wel in staat om te werken met de toolboxes? We willen graag dat leerling hun eigen teksten formatief evalueren tijdens de lessen, maar uit de interviews met de docenten bleek dat leerlingen hier veel ondersteuning bij nodig hebben. Een van de docenten (5-havo) formuleert dit als volgt: *"Deze lessen hebben mij laten zien dat leerlingen qua schrijfvaardigheid nog veel moeten leren. Het laat mij ook zien dat ze het best leren onder begeleiding. Wanneer ik feedback geef of één van de e-mails klassikaal bespreek, zie ik dat ze beter begrijpen wat er goed/fout is. Het laat mij nadenken over 'curling'-docenten ... hebben wij dit gecreëerd? Hoe zelfstandig zijn leerlingen en hoe groot is ons aandeel in deze onzelfstandigheid?"*

Daarnaast hebben we ons ook afgevraagd of de interventie bestond uit voldoende lessen waarin de leerlingen zich de kennis eigen konden maken. In het ontwikkelde materiaal zijn voor beide leerjaren alle alineanormen aan de orde geweest. De totale interventie bestond voor de leerlingen van 2-havo uit negen lessen en voor de leerlingen uit 5-havo uit acht lessen. Was het aanleren van elf normen niet een te omvangrijke taak voor de leerlingen in de beperkte tijd die ons ter beschikking stond? Mogelijk was meer tijd en oefening inbouwen voor de interventie wenselijk geweest en hadden de leerlingen dan in de nameting beter gescoord. Dit had het onderzoek ecologisch echter minder valide gemaakt. Een docent Nederlands is immers beperkt in zijn tijd. Het schrijven van alinea's maakt deel uit van de schrijfvaardigheidslessen en binnen het hele curriculum is dat slechts een beperkt gedeelte van alle te behandelen stof. Van de begeleidende docenten en van de deelnemende leerlingen hebben we geen opmerkingen gekregen over de 'volheid' van de interventielessen. We menen daarom dat het aantal bestede lessen voor het aanleren van de alineanormen binnen twee genres praktisch haalbaar en voldoende is geweest.

Een belangrijke vraag blijft dus: zijn de ontworpen alineatoolboxes adequaat geweest, of, met andere woorden, potentieel effectief? Gezien de literatuur waarop deze toolboxes

gebaseerd zijn en de eerder behaalde effecten met het werken met toolboxes, was het de verwachting dat we ook in het schrijfvaardigheidsonderwijs het werken met toolboxes konden integreren. Het opwerpen van deze laatste vraag brengt ons tot slot bij enkele suggesties voor vervolgonderzoek.

5.3 Suggesties voor vervolgonderzoek

Deze interventiestudie biedt de lezer, docent dan wel wetenschapper, een eerste uitgebreide en wetenschappelijk verantwoorde blik op het schrijven van alinea's van leerlingen. De uitkomsten van deze interventiestudie geven aanleiding tot vervolgonderzoek. We noemen hier twee mogelijkheden.

De eerste vraag die we ons gesteld hebben is: Worden alinea's die aan alle alinea-normen voldoen door lezers ook beoordeeld als kwalitatief betere alinea's? Vanuit het perspectief van het tekstanalytisch onderzoek waarin de tekst gescoord is op de alinea-kenmerken per alinea is dat zeker het geval. Immers, de alinea-normen zijn gebaseerd op tekstwetenschappelijke inzichten. Door een systematische analyse van de alinea's hebben we kunnen vaststellen aan welke alinea-normen de alinea's voldoen. Hoe meer alinea-normen een alinea bevat hoe hoger de kwaliteit van de alinea. In het onderzoek is echter een andere kwaliteitsmaatstaf gebruikelijk, namelijk het holistische oordeel. Het maken van een kwalitatieve analyse vanuit de holistische benadering kan een zinvolle aanvulling zijn op tekstanalyse die we nu gedaan hebben. Ligt volgens docenten de kwaliteit van de leerlingteksten waarin alinea's voorkomen die aan veel alinea-normen voldoen ook daadwerkelijk hoger dan die van teksten met alinea's waarin minder alinea-normen gevolgd worden? Een alternatief voor de holistische benadering die voor het analyseren van alinea's ook interessant kan zijn, is de *primary trait beoordeling* (Schoonen & De Glopper, 1992) waarbij de alinea beoordeeld wordt op het hoofdkenmerk, namelijk samenhang, ervan. De centrale vraag voor een alinea zou bijvoorbeeld kunnen zijn: Hoe samenhangend is deze alinea? Bij beide soorten beoordelingsonderzoek lijkt het gebruik van schaalbeoordeling (Pollmann et al., 2012) het meest voor de hand te liggen.

De tweede vraag die een vervolgonderzoek waard is, is die naar de verdeling van de alinea-normen over het curriculum. Hoe zou die zo efficiënt mogelijk kunnen zijn? In dit onderzoek hebben we ervoor gekozen de leerlingen van 2-havo en 5-havo alle normen aan te bieden, in niveau aangepast aan het leerjaar. Een alternatief zou het spreiden van de elf alinea-normen over meerdere leerjaren kunnen zijn, bijvoorbeeld in de brugklas en de tweede klas de meest basale normen. Dat kan uitgebouwd worden over de jaren heen. Dat roept ook weer andere vragen op. Wanneer we bijvoorbeeld in de brugklas en de tweede klas zouden beginnen met het aanleren van de afbakening van alinea's is daarbij ook de thematische samenhang nodig. Je kunt

immers pas goed afbakenen als je in staat bent te bepalen wat de deelonderwerpen zijn van het te beschrijven onderwerp.