

**Faculteit der Letteren**  
BA Taalwetenschap

Studiejaar 2018-2019  
28/06/2019

# **Het effect van presentatiewijze op de beoordeling van een schrijfttekst**

Anne G. H. van Asseldonk, s4761022  
A, eerste begeleider  
B, tweede begeleider

## *Voorwoord*

Voor u ligt de scriptie 'Het effect van presentatiewijze op de beoordeling van een schrijfttekst'. Het is een onderzoek naar het effect van de presentatiewijze van een tekst op de beoordeling ervan. Deze scriptie is geschreven in het kader van de afronding van mijn bacheloropleiding Taalwetenschap aan de Radboud Universiteit. Van april 2019 tot en met juni 2019 ben ik bezig geweest met het onderzoek en het schrijven van deze scriptie.

Bij het opzetten en schrijven van mijn scriptie ben ik begeleid door A. Ik zou hem hiervoor graag willen bedanken. Door onze brainstormsessies kreeg mijn scriptie vorm en door zijn feedback kwam mijn scriptie steeds dichterbij het eindproduct. Daarnaast wil ik mijn tweede begeleider B bedanken voor het beoordelen van mijn scriptie.

Voor mijn onderzoek was het cruciaal dat drie mensen met uiteenlopende handschriften vier bepaalde teksten overschreven. Ik ben hen (C, D en E) daarom zeer dankbaar.

Ik was erg blij met iedereen die me heeft geholpen bij het werven van proefpersonen. Ik wil ten eerste F, G en H bedanken. Dankzij hen kon ik mijn praatje doen bij de masterstudenten en de vierdejaars bachelorstudenten die opgeleid werden tot docent Engels. Verder wil ik I, J en H bedanken, die hun medestudenten of collega's hebben gevraagd om mee te doen aan mijn onderzoek. Ten slotte ben ik natuurlijk ook alle studenten die hebben deelgenomen erg dankbaar.

Het schrijven van deze scriptie was een leerzaam proces. Ondanks het tegenvallende aantal proefpersonen, ben ik tevreden over wat ik heb neergezet. Ik wens u veel leesplezier.

Anne van Asseldonk  
Nijmegen, 28 juni 2019

## *Inhoudsopgave*

Voorwoord	I
Inhoudsopgave	II
Hoofdstuk 1 Inleiding	1
1.1 Validiteit en betrouwbaarheid	1
1.2 Invloed van persoonskenmerken	1
1.3 Invloed van tekstkenmerken	1
1.4 Presentatie-effect	2
1.4.1 Verklaringen presentatie-effect	2
1.4.2 Presentatie-effect bij verschillende teksten	2
1.4.3 Presentatie-effect bij verschillende beoordelaars	3
1.4.4 Presentatie-effect bij verschillende criteria	3
1.4.5 Verminderen presentatie-effect	3
1.4.6 Inconsistente resultaten	4
1.4.7 Effect type handschrift	4
1.5 De tweedetaalleerder als docent	5
1.6 Huidige studie	6
1.6.1 Onderzoeksvraag	6
1.6.2 Verwachtingen	6
1.6.3 Relevantie	7
Hoofdstuk 2 Methode	7
2.1 Onderzoeksontwerp	7
2.2 Materiaal	8
2.3 Proefpersonen	9
2.4 Procedure	9
2.5 Data-analyse	9
Hoofdstuk 3 Resultaten	10
3.1 Gemiddelden en standaarddeviaties	10
3.2 Het effect van presentatie	11
3.3 Het interactie-effect van presentatie en criterium	11
3.4 Het interactie-effect van presentatie en taal	12
3.5 Het interactie-effect van presentatie en criterium in de T1 en T2	13
Hoofdstuk 4 Discussie	15
4.1 Bespreking resultaten	15
4.2 Mogelijke verklaringen resultaten	16
4.2.1. Steekproefgrootte	16
4.2.2. Bewustzijn	16
4.2.3. Homogeniteit beoordelaars	16
4.2.4. Analyse resultaten	16
4.3 Conclusie	17
Literatuurlijst	18
Bijlage I Teksten uit het onderzoek	20
Bijlage II Voorbeelden van de drie handschriften	22
Bijlage III Instructie voor het experiment	23
Bijlage IV Beoordelingsformulier voor het experiment	24

## *Abstract*

In de huidige studie is onderzocht of teksten anders beoordeeld worden afhankelijk van de manier waarop de tekst gepresenteerd wordt. Er waren vier presentatiewijzen: mooi handgeschreven, gemiddeld handgeschreven, lelijk handgeschreven en getypt. Daarnaast is onderzocht of het presentatie-effect interacteert met de taal waarin de tekst geschreven is (de T1 of de T2 van de beoordelaar) en het criterium waarop de tekst wordt beoordeeld (inhoud of taalgebruik). Om dit te onderzoeken hebben 12 studentdocenten een set van vier teksten beoordeeld. Uit de huidige studie is gebleken dat er geen sprake is van het presentatie-effect. De invloed van presentatiewijze bleek niet significant te verschillen tussen de twee niveaus van criterium en taal, ook niet wanneer de niveaus van criterium per taal werden getoetst. Er kan dus geconcludeerd worden dat teksten niet aantoonbaar anders beoordeeld worden op basis van de presentatiewijze van de tekst en dat er geen interactie-effect met taal en criterium is.

## *1.0 Inleiding*

### *1.1 Validiteit en betrouwbaarheid*

Een belangrijke vaardigheid waarover een docent moet beschikken is het correct kunnen beoordelen van de prestaties van leerlingen. De beoordelingen kunnen namelijk de studieloop, of zelfs de carrière, van een leerling beïnvloeden. Het is daarom van belang dat een beoordeling valide en betrouwbaar is. Een beoordeling is valide wanneer beoordeeld wordt wat er gemeten moet worden. Het gaat erom of de beoordeling daadwerkelijk de leerdoelen toetst en of alle leerdoelen voldoende gerepresenteerd zijn in de beoordeling. Daarnaast moet een beoordeling betrouwbaar zijn: een consistente score laat zien dat de beoordelingen betrouwbaar zijn. Wanneer twee verschillende docenten dezelfde tekst beoordelen, moeten zij op dezelfde score uitkomen. De validiteit en betrouwbaarheid van een beoordeling kunnen in het gedrang komen wanneer andere factoren dan de kwaliteit van een tekst een rol beginnen te spelen.

### *1.2 Invloed van persoonskenmerken*

Een beoordelaar kan zich onbewust door talloze persoonskenmerken, zoals de etniciteit of het gender van een leerling, laten beïnvloeden. Onderzoek naar deze factoren is wijdverbreid. Zo bleek uit onderzoek dat de lees- en schrijfteksten van studenten die tot een etnische minderheid behoren minder hoog beoordeeld worden dan de overige studenten (Meissel, Meyer, Yao, & Rubie-Davies, 2017). Ook op het gebied van *bias* op basis van gender is veel onderzoek gedaan. De resultaten indiceren over het algemeen dat het gender van de student de beoordeling van de docent niet significant beïnvloedt (o.a. Hecht & Greenfield, 2002), maar ook het tegendeel is gevonden: Ready en Wright (2011) vonden dat jongens negatiever beoordeeld werden dan meisjes op het gebied van taalvaardigheid.

### *1.3 Invloed van tekstkenmerken*

Naast persoonskenmerken, kan een beoordelaar zich laten beïnvloeden door kenmerken van de tekst, zoals het soort voorbeelden dat gebruikt wordt. Studenten die in hun essays voorbeelden op basis van hun persoonlijke ervaring gebruikten, kregen lagere scores dan studenten die wetenschappelijke voorbeelden aanhaalden (Kobrin, Deng, & Shaw, 2011). De studenten die beide typen voorbeelden aanhaalden, kregen de hoogste scores. Ook het vijfparagrafenmodel, een essayvorm van vijf paragrafen met een bepaalde opbouw, wordt genoemd als een factor die van invloed is op de beoordeling van een essay. In Kobrin et al. (2011) werd weliswaar gevonden dat essays in deze vorm hogere scores kregen, maar wegens interactie met andere factoren kunnen geen duidelijke conclusies worden getrokken.

Het woordaantal beïnvloedt de beoordeling van een essay ook: er is een positieve relatie gevonden tussen woordaantal en essayscore (o.a. Kobrin, Deng, & Shaw, 2007; Mattern, Camara, Kobrin, 2007). Hoe hoger het woordaantal was, hoe hoger de essayscore. Dit is niet zo verrassend gegeven het feit dat woorden nodig zijn om gedachten uit te drukken en argumenten te onderbouwen. Desondanks werd door onder andere Perelman (2005) geclaimd dat essaylengte boven inhoud wordt beloond in het scoringsproces.

#### *1.4 Presentatie-effect*

Met de opkomst van de computer in het klaslokaal is een nieuw tekstenkenmerk van invloed ontstaan: ook de manier waarop een tekst gepresenteerd wordt, blijkt namelijk de beoordeling van een tekst te beïnvloeden. De verschillen in beoordeling van handgeschreven en getypte teksten is al veelvuldig onderzocht. Zo vonden Arnold en collega's (1990) dat getypte versies van schrijftaken lager werden beoordeeld dan hun handgeschreven equivalent. Lezers gaven de voorkeur aan handgeschreven teksten, zelfs al waren die teksten soms slechter leesbaar dan de getypte teksten. Onderzoeken van onder anderen Powers, Fowles, Farnum en Ramsey (1994), Russell en Tao (2004) en Kohler (2015) ondersteunen dit resultaat. In Russell & Tao (2004a/b) wordt het verschil in beoordeling door de manier waarop een tekst gepresenteerd wordt het presentatie-effect genoemd. In de huidige studie staat dit effect centraal. Het presentatie-effect bleek niet te interacteren met het gender, de etniciteit en het taalniveau van studenten (Bridgeman en Cooper, 1998).

##### *1.4.1 Verklaringen presentatie-effect*

Er worden verschillende verklaringen voor het presentatie-effect gegeven. Zo suggereerden Arnold en collega's (1990) dat lezers hogere verwachtingen hadden bij getypte essays, dat lezers zich beter met de auteur konden identificeren bij handgeschreven essays en dat lezers studenten het voordeel van de twijfel gaven wanneer ze hun handschrift niet goed konden lezen. Andere genoemde redenen voor de lagere beoordeling van getypte teksten zijn het lengte-effect (getypte essays leken korter dan handgeschreven essays) en het feit dat fouten beter zichtbaar zijn in een getypte tekst (Powers et al., 1994).

##### *1.4.2 Presentatie-effect bij verschillende teksten*

Het presentatie-effect is onderzocht bij teksten met verschillende thema's. In het onderzoek van Powers et al. (1994) werden bijvoorbeeld essays over persoonlijke ervaringen en essays over algemene zaken gebruikt. MacCann, Eastment en Pickering (2002) maakten zelfs gebruik van vijf verschillende schrijfoopdrachten, waarnaar ze verwijzen met secties A, B, C, D en E. Scholieren van 14 en 15 jaar maakten deze opdrachten. De teksten in secties A, B en C waren antwoorden op open vragen. In sectie A kreeg de scholier 15 minuten de tijd om te reageren op een vraag. In secties B en C kreeg de scholier langere schrijftijd: 30 minuten om te reageren op een langere vraag (sectie B) of een vraag naar keuze (sectie C). De teksten in secties D en E waren essays van een externe toets. In sectie D kreeg de scholier 15 minuten om een formele brief te schrijven aan een redacteur van een krant. In deze brief moest de scholier argumenten vóór of tegen een standpunt in het stimulusmateriaal geven. In sectie E, ten slotte, moest de scholier een formele toespraak uitschrijven. De toespraak was erop gericht om een commissie te overtuigen. De resultaten van dit onderzoek waren gemengd: bij secties A, B en D werd de handgeschreven tekst hoger beoordeeld dan de getypte tekst, maar bij sectie C was het omgekeerde van toepassing. De onderzoekers wijzen dit toe aan de extreme openheid van de vraag: de scholieren werden gevraagd om een verhaal te schrijven over één van negen verschillende stimulusitems. Hierdoor zouden ze volgens de onderzoekers meer gefocust hebben op

creativiteit dan op het plannen en integreren van informatie. Een andere reden die genoemd wordt, is het feit dat de teksten in sectie C langer waren, wat ertoe kan leiden dat het lengte-effect niet meer van toepassing was. Uit het onderzoek van MacCann en collega's blijkt dus dat het thema van een tekst invloed kan hebben op het voorkomen van het presentatie-effect. In de huidige studie zullen teksten worden gebruikt waarin scholieren hun mening over een algemene zaak geven.

#### *1.4.3 Presentatie-effect bij verschillende beoordelaars*

In de meeste studies beoordeelden ervaren docenten de teksten van de leerlingen. In de studie van Markham (1976) werden zowel docenten als studentdocenten ingezet als beoordelaars. De docenten en de studentdocenten werden niet significant anders beïnvloed door de presentatiewijze van een tekst. Het bleek alleen dat de studentdocenten hogere cijfers gaven. Een studie van Sweedler-Brown (1985) bevestigde dat onervaren beoordelaars hogere cijfers geven dan ervaren beoordelaars.

Het presentatie-effect is gevonden bij docenten van verschillende schoolniveaus. Zo toonden onder anderen Bridgeman en Cooper (1998) het presentatie-effect aan bij docenten in het hoger onderwijs en Russel en Tao (2004a) bij docenten op de middelbare school en de basisschool. In de huidige studie zullen toekomstige middelbareschooldocenten teksten van middelbare scholieren beoordelen.

#### *1.4.4 Presentatie-effect bij verschillende criteria*

In een groot deel van de onderzoeken naar het presentatie-effect werd aan de beoordelaars gevraagd of ze de teksten een algemene score wilden geven. In Russell & Tao (2004b) werd de beoordeling opgesplitst in twee categorieën: Engelse conventies en topicontwikkeling. Aan de hand van omschrijvingen moesten de beoordelaars de topicontwikkeling in de teksten een score van 1 tot en met 6 geven en de Engelse conventies een score van 1 tot en met 4. De beoordelaars kregen handgeschreven voorbeeldessays voor elke score. De handgeschreven teksten bleken op beide categorieën significant hogere scores te krijgen dan de getypte teksten. Het verschil in de beoordeling van handgeschreven en getypte teksten was het grootst op het gebied van topicontwikkeling. Het onderzoek richtte zich echter niet op het verschil tussen de twee categorieën, dus het is onbekend of dit verschil significant was.

#### *1.4.5 Verminderen presentatie-effect*

In verschillende experimenten is geprobeerd het presentatie-effect te verminderen, onder andere door middel van het trainen van de beoordelaars. In de studie van Powers en collega's (1994) werd tijdens deze training benadrukt dat handgeschreven en getypte teksten verschillende indrukken geven en de beoordelaars werden aangemoedigd om verder te kijken dan de manier waarop het essay was gepresenteerd. De onderzoekers bespraken tijdens deze training de resultaten van een onderzoek dat het presentatie-effect aantoonde. Ook de invloed van de waargenomen lengte van een essay op de beoordeling werd besproken en er werd een poging gedaan om deze invloed te verminderen door de getypte teksten in het onderzoek een vergrote regelafstand te geven. Het presentatie-effect was hierdoor 25 procent minder groot dan zonder training en aangepaste regelafstand. In Russell & Tao (2004a), daarentegen, had het aanpassen van de regelafstand geen significante impact op de scores van de beoordelaars. Dit duidt erop dat de afname van het presentatie-effect in Powers et al. (1994) toe te schrijven is aan de training en niet aan de aangepaste regelafstand. In het vervolgonderzoek van Russell en Tao (2004b) werd de regelafstand van de getypte teksten daarom niet aangepast. De training was verder vergelijkbaar: ook hier werden het lengte-effect en eerder onderzoek naar het presentatie-

effect besproken. Daarnaast wezen de onderzoekers op de verschillen in de zichtbaarheid van fouten tussen de twee typen teksten en discussieerden ze met de beoordelaars over de verschillen in scores met de focus op de invloed van uiterlijk. De beoordelaars die de extra training niet hadden gevolgd en niet bekend waren met het presentatie-effect beoordeelden de getypte tekst significant lager dan de handgeschreven tekst. De beoordelaars die extra training hadden gekregen en hierdoor bewust waren van het effect, beoordeelden de getypte en handgeschreven versies niet langer significant verschillend.

In hetzelfde onderzoek van Russell en Tao (2004b) werd ook onderzocht of een getypte tekst in een lettertype dat op handschrift lijkt het effect vermindert. Uit deze studie bleek dat deze tekst niet anders werd beoordeeld dan een handgeschreven tekst, maar wel significant hogere beoordelingen kreeg dan getypte teksten in een standaardlettertype. Beoordelaars gaven aan de tekst in dit lettertype moeilijker te kunnen lezen, waardoor ze de teksten minder grondig lazen.

Ten slotte werd in Russell & Tao (2004b) de invloed van de zichtbaarheid van fouten op de beoordeling onderzocht. Zowel Powers en collega's (1994) en Russell en Tao (2004a) noemden de verhoogde zichtbaarheid van fouten in een getypte tekst als mogelijke verklaring voor het presentatie-effect. Door spelfouten te verwijderen, hoopten de onderzoekers te ontdekken of de verhoogde zichtbaarheid van fouten daadwerkelijk bijdroeg aan het presentatie-effect. Het effect van de verhoogde zichtbaarheid van fouten was groter voor de categorie Engelse conventies dan voor topicontwikkeling. Bij beide categorieën leidde het verwijderen van de spelfouten echter niet tot een significant verschil.

#### *1.4.6 Inconsistente resultaten*

De resultaten van verschillende studies naar het presentatie-effect zijn niet consistent. Eerder beschreven onderzoeken van onder andere Arnold en collega's (1990) vonden namelijk dat handgeschreven teksten hoger beoordeeld werden dan getypte teksten. Zoals eerder genoemd gaf het onderzoek van MacCann et al. (2002) gemengde resultaten. Het onderzoek van Harrington, Shermis en Rollins (2000) gaf zelfs geen significante verschillen in de beoordeling van handgeschreven, getypte en omgezette (van handgeschreven naar getypt) teksten. Daarnaast vonden enkele oudere studies juist dat getypte teksten hoger werden beoordeeld dan handgeschreven teksten (Peacock, 1988; Sweedler-Brown, 1991).

#### *1.4.7 Effect type handschrift*

In de genoemde onderzoeken naar het verschil tussen getypte en handgeschreven teksten is bij handgeschreven teksten geen onderscheid gemaakt tussen mooie en lelijke handschriften. Onderzoeken die dit onderscheid wél hebben gemaakt, hebben gevonden dat dit een effect heeft op de beoordeling. Een essay geschreven in een mooi handschrift wordt namelijk significant hoger beoordeeld dan een essay in een lelijk handschrift (o.a. Briggs, 1970; Markham, 1976). In de studie van Klein en Taub (2005) werden verschillende typen handschriften onderscheiden, op basis van de leesbaarheid van het handschrift. Onder andere netheid, ruimtelijke ordening, uniformiteit en het soort lettertype dragen bij aan een goede leesbaarheid (Graham, Berninger, & Weintraub, 1998). In Klein & Taub (2005) werden goed leesbare getypte essays het best beoordeeld, gevolgd door goed leesbare met de pen geschreven essays. De goed leesbare essays werden in alle varianten beter beoordeeld dan de onduidelijke essays. In een recent onderzoek van Greifeneder en collega's (2010) werden vergelijkbare resultaten gevonden: de goed leesbare essays kregen hogere beoordelingen dan de slecht leesbare essays. De onderzoekers linkten dit aan vloeiendheid. Volgens hen worden goed leesbare teksten vloeiender verwerkt door de lezer, wat leidt tot de gedachte dat de tekst dan wel goed moet zijn.

Ook in de studie van Marshall en Powers (1969) werd een onderscheid tussen verschillende typen handschriften gemaakt. Zij gebruikten vier typen handschriften in hun experiment: één in een mooi handschrift geschreven, één in een gemiddeld handschrift geschreven, één in een slordig handschrift geschreven en één getypt met een typemachine. De vier typen handschriften werden gecombineerd met drie verschillende vormen en hoeveelheden fouten: geen fouten, 18 spelfouten of 18 grammaticale fouten. Op deze manier waren er uiteindelijk twaalf verschillende essays. De teksten werden beoordeeld door docenten in opleiding. Uit dit onderzoek bleek dat de essays zonder fouten significant hoger werden beoordeeld dan de essays met spel- of grammaticafouten. Bovendien bleek dat het essay in het nette handschrift het beste werd beoordeeld, gevolgd door respectievelijk het essay in het slordige handschrift, het getypte essay en het essay in het gemiddelde handschrift. Er was alleen een significant verschil tussen de beoordeling van het essay in het nette handschrift en het essay in het gemiddelde handschrift. Er bleek geen interactie-effect tussen fouten en handschrift te zijn. De auteurs noemen de resultaten op het gebied van handschrift verrassend en houden de optie open dat de resultaten het gevolg zijn van de ongewone beoordelingssituatie waarin de beoordelaars verkeerden. Ze beoordeelden de teksten namelijk tijdens een les in plaats van in hun eigen tijd. Daarbij kregen ze specifieke instructies mee, zoals het moeten gebruiken van een 9-puntsschaal en het moeten beoordelen puur op de inhoud van essay.

### *1.5 De tweedetaalleerder als docent*

Het experiment in het huidige onderzoek zal afgenomen worden bij tweedetaalleerders van het Engels die opgeleid worden tot docent Engels. Zij zullen in dit onderzoek zowel teksten in hun moedertaal als in hun tweede taal beoordelen. Er is reeds veel onderzoek gedaan naar de verschillen tussen docenten die hun moedertaal doceren en docenten die hun tweede taal doceren. Medgyes (1994) stelt dat deze twee groepen docenten verschillen in hun onderwijsgedrag en dat deze verschillen grotendeels voortkomen uit de discrepantie in hun taalvaardigheid: moedertaalsprekers zijn vaardiger dan tweedetaalleerders. Eén verschil dat de onderzoeker vond is dat moedertaaldocenten meer focusten op betekenis, hoewel de tweedetaaldocenten zich vooral richtten op vorm en grammaticaregels. Árva en Medgyes (2000) wilden de claims die gemaakt werden in het onderzoek van Medgyes heroverwegen. Zij bevestigden dat de focus van de twee groepen docenten anders ligt: de moedertaalsprekers gaven voornamelijk conversatielessen gericht op communicatie, terwijl de tweedetaalleerders beter inzicht hadden in de grammatica en fouten strenger beoordeelden. Braine (2010) geeft in zijn boek een overzicht van onderzoeken naar de zelfpercepties van docenten die lesgeven in hun tweede taal. In de conclusies die hij trekt over deze onderzoeken, bevestigt hij dat deze docenten gericht zijn op de meer structurele en stilistische conventies van een taal, ten koste van communicatief taalonderwijs.

De verschillen in onderwijsgedrag van docenten die in hun moedertaal en docenten die in hun tweede taal doceren, zijn ook terug te vinden in de manieren waarop zij schrijfteksten beoordelen. In een onderzoek van Lee (2009) werden docenten met een Engelse achtergrond vergeleken met docenten met een Koreaanse achtergrond bij het beoordelen van Engelse schrijfteksten. De onderzoeker concludeerde dat de moedertaalsprekers strikter waren bij het beoordelen van inhoud en de tweedetaalleerders bij het scoren van grammatica en zinsstructuren. De onderzoeker concludeerde ook dat de Koreaanse docenten inferieur waren bij het meten van linguïstische componenten en dat hun beoordelingen minder betrouwbaar waren dan die van de Engelse docenten. De beoordelingen van de Engelse docenten waren consistent. Ook Kim en Di Gennaro (2012) concludeerden dat de tweedetaalleerders meer varieerden in de strengheid van hun beoordelingen. De resultaten van het onderzoek van Lee zijn echter niet volledig in



overeenstemming met de resultaten van Kim en Di Gennaro. In het onderzoek van Kim en Di Gennaro en in dat van Park (2015) beoordeelden de tweedetaalleerders op alle facetten strenger dan de moedertaalsprekers. Johnson en Lim (2009) concludeerden daarentegen dat de tweedetaalleerders en moedertaalsprekers in hun onderzoek even effectief waren in het beoordelen van schrijfteksten.

## *1.6 Huidige studie*

### *1.6.1 Onderzoeksvraag*

In de huidige scriptie wordt het onderscheid tussen mooie en lelijke handschriften gemaakt. Er worden vier typen teksten meegenomen: mooi handgeschreven, gemiddeld handgeschreven, lelijk handgeschreven en getypt. Daarbij wordt ten eerste een onderscheid gemaakt tussen de beoordeling op inhoud en de beoordeling op taalgebruik en ten tweede tussen de beoordeling van een tekst in de eerste taal (T1) en in de tweede taal (T2) van de beoordelaar. Mijn onderzoeksvraag is dus als volgt: 'In hoeverre wordt een beoordelaar beïnvloed door de manier waarop een tekst gepresenteerd wordt bij het beoordelen ervan en is er een verschil tussen de invloed op de beoordeling van inhoud of taalgebruik en de beoordeling van teksten in de T1 en T2?'

De deelvragen die daaruit volgen, zijn:

1. Verschillen de beoordelingen van een mooi handgeschreven, een gemiddeld handgeschreven, een lelijk handgeschreven en een getypte tekst van elkaar?
2. Zijn de verschillen tussen de beoordelingen van de teksten in de vier presentatiewijzen anders voor inhoud en taalgebruik?
3. Zijn de verschillen tussen de gemiddelde beoordelingen van de teksten in de vier presentatiewijzen anders voor teksten in de T1 en T2 van de beoordelaar?
4. Zijn de verschillen tussen de beoordelingen van de teksten in de vier presentatiewijzen anders voor inhoud en taalgebruik afhankelijk van de taal waarin de tekst is geschreven?

### *1.6.2 Verwachtingen*

Bij de eerste deelvraag verwacht ik verschillen in de beoordelingen van de teksten in de vier presentatiewijzen te vinden. Ik verwacht dat de teksten die in een mooi handschrift geschreven zijn en de teksten die getypt zijn beter beoordeeld zullen worden dan de teksten in een lelijk handschrift. Dit is niet in overeenstemming met het onderzoek van Arnold en collega's (1990), waar de lezers de voorkeur gaven aan handgeschreven teksten, ook al waren die minder goed leesbaar. Dit onderzoek is echter al 29 jaar geleden gepubliceerd en ik verwacht dat docenten nu meer gewend zijn aan getypte teksten en hier daarom de voorkeur aan zullen geven boven teksten die minder goed leesbaar zijn door een slordig handschrift. Dit bleek ook uit het recentere onderzoek van Klein en Taub (2005). Ik verwacht daarnaast dat er een verschil in beoordeling zal zijn tussen een tekst in mooi handschrift en een getypte tekst.

Ik vermoed dat de verschillen in de beoordeling van de teksten in de vier presentatiewijzen anders zullen zijn op het gebied van taalgebruik dan op het gebied van inhoud. Dit sluit aan bij de resultaten van Russell en Tao (2004b): het effect van de verhoogde zichtbaarheid van fouten was daar groter voor de categorie Engelse conventies dan voor topicontwikkeling. Spel- en typfouten zullen dus milder of strenger worden beoordeeld op basis van de conditie waarin de tekst zich bevindt. Echter was het verschil in de beoordeling van handgeschreven en getypte teksten het grootst op het gebied van topicontwikkeling. Het presentatie-effect zou dus een groter effect kunnen hebben op de beoordeling van één van de twee criteria.

Daarnaast verwacht ik dat de teksten in de vier presentatiewijzen anders beoordeeld zullen worden op basis van de taal waarin ze geschreven zijn: de T1 of de T2 van de beoordelaar. Dit is nog niet eerder onderzocht. Ik verwacht dat bij de teksten in de T1, de Nederlandse teksten, de presentatiewijze een minder groot effect heeft op de beoordeling dan bij de teksten in de T2, de Engelse teksten. Uit de onderzoeken van Lee (2009) en Kim en Di Gennaro (2012) bleek dat tweedetaalleerders minder consistent zijn in het beoordelen van schrijftaken dan moedertaalsprekers. Tijdens het beoordelen van een tekst in hun T2 laten zij zich wellicht meer leiden door andere factoren, waardoor de presentatiewijze een groter effect zou kunnen hebben.

Ten slotte is mijn verwachting dat de beoordelaars bij de Nederlandse teksten zich meer zullen focussen op de inhoud, terwijl ze bij de Engelse teksten zich zullen richten op het taalgebruik. Dit sluit aan bij de verschillen in onderwijsgedrag die Medgyes (1994) vond. Hierdoor zullen de beoordelaars zich wellicht meer door de presentatiewijze laten leiden bij het criterium waar de focus niet op ligt. Ik verwacht daarom dat het presentatie-effect in de T2 groter is bij inhoud dan bij taalgebruik. Andersom verwacht ik dat het presentatie-effect in de T1 groter is bij taalgebruik dan bij inhoud.

### *1.6.3 Relevantie*

De relevantie van dit onderzoek zit ten eerste in het feit dat er weinig recent onderzoek naar dit fenomeen is gedaan. Aangezien het gebruik van de computer in het klaslokaal sterk gegroeid en genormaliseerd is, is het mogelijk dat de beoordelingen van de verschillende versies van de essays veranderd zijn. Opdrachten die beoordeeld moeten worden, worden tegenwoordig bijna altijd getypt. Dit kan er eventueel toe leiden dat docenten anders tegen getypte en handgeschreven teksten aankijken. Dit onderzoek is ook relevant, omdat er twee nieuwe aspecten worden onderzocht, namelijk: is er een verschil tussen beoordeling op taalgebruik en inhoud én is er een verschil tussen beoordeling op een tekst in de T1 en T2 van de beoordelaar? Antwoorden op deze vragen zullen leiden tot nieuwe inzichten op het gebied van taaltoetsing. Docenten kunnen de resultaten van dit onderzoek meenemen in de praktijk. Tijdens een training kan bijvoorbeeld gericht getraind worden op het presentatie-effect.

## 2.0 Methode

### *2.1 Onderzoeksontwerp*

Om de onderzoeksvraag te kunnen beantwoorden, is er een experiment opgezet. In dit experiment kregen beoordelaars de opdracht om vier teksten te beoordelen. In het experiment werden telkens dezelfde vier teksten aangeboden aan de beoordelaars. De volgorde van de teksten bleef gelijk. De eerste twee teksten waren telkens in de tweede taal van de beoordelaar geschreven: het Engels. De laatste twee teksten waren in de eerste taal van de beoordelaar geschreven: het Nederlands. De vier teksten staan in het getypte format in Bijlage I, in de volgorde waarin ze ook in het experiment voorkwamen. De beoordelaars beoordeelden elke tekst op zowel inhoud als taalgebruik. De presentatieconditie waarin zij elke tekst moesten beoordelen, verschilde. Hierdoor kwam elke presentatiewijze in combinatie met elke tekst voor. Hier is voor gekozen om effect van volgorde van presentatiewijze te voorkomen én om extra te controleren voor een eventueel effect van de kwaliteit van individuele teksten. Er waren vier verschillende presentatiewijzen, waardoor er in totaal 16 verschillende teksten waren. De tekst kon met de hand geschreven zijn, waarbij onderscheid werd gemaakt tussen teksten in een mooi, gemiddeld en lelijk handschrift, of getypt zijn. De beoordelaars werden willekeurig over de verschillende combinaties verdeeld.

De beoordelaars beoordeelden de teksten in de vier presentatiewijzen, in de twee talen en op de twee criteria, waardoor de scores op de verschillende condities niet

onafhankelijk zijn. Hierdoor wordt deze assumptie van de eenweg- en tweeweg-ANOVA's geschonden. In de resultatensectie wordt hier nog op ingegaan.

De afhankelijke variabelen in dit onderzoek waren de (inhouds- en taalgebruik-)oordelen. Er waren daarnaast drie onafhankelijke variabelen: de manier waarop de tekst gepresenteerd werd, de taal waarin de tekst geschreven was en het criterium waarop de tekst beoordeeld werd.

## *2.2 Materiaal*

De vier teksten waren afkomstig uit een eerder onderzoek van Schoonen, Van Gelderen, Stoel, Hulstijn en De Glopper (2011). Deze teksten zijn geschreven door middelbare scholieren. In deze teksten geven de scholieren hun mening over een algemene zaak. In dit geval betrof het twee Engelse teksten over verplichte schooluniformen en twee Nederlandse teksten over het deel van de weg waar skaters mogen rijden.

Mijn keuze voor specifiek deze vier teksten is op een aantal zaken gebaseerd. Zo heb ik erop gelet dat de vier teksten in dezelfde presentatiewijze (in dit geval getypt) ongeveer gelijk zijn beoordeeld met een bovengemiddelde score. In het eerdergenoemde onderzoek van Schoonen en collega's (2011) waren de vier teksten met een score van circa één standaarddeviatie boven de gemiddelde score van 100 beoordeeld. De exacte scores van de teksten waren 113.5, 114, 115 en 115.5. In het huidige onderzoek is er om deze reden van uitgegaan dat de vier teksten kwalitatief vergelijkbaar zijn. Bij het kiezen van teksten is erop gelet dat het aantal spelfouten in de vier teksten ongeveer gelijk was. Ook is rekening gehouden met de lengte van de essays. De vier essays waren gemiddeld 200 woorden lang en varieerden van 144 tot 262 woorden. Ten slotte heb ik ervoor gekozen om telkens twee teksten (één Engels, één Nederlands) van één schrijver te nemen, zodat de schrijfstijl zoveel mogelijk gelijk bleef. De aangeleverde teksten waren ondertekend met initialen, waardoor de beoordelaars hadden kunnen ontdekken dat de vier teksten van dezelfde twee schrijvers afkomstig waren. Om eventuele vooroordelen tegenover een schrijver of tekst te voorkomen, zijn de initialen veranderd zodat het leek alsof er vier verschillende schrijvers waren.

Zoals vermeld, werden de teksten in vier verschillende condities aangeboden. De originele teksten waren in een getypt format. Deze teksten zijn overgeschreven door drie verschillende handschrijvers. Hun handschriften waren deel van een set van twaalf handschriften. Deze twaalf handschriften werden door 21 willekeurige personen op volgorde gezet van mooi tot lelijk. De handschriften die als mooist, gemiddeld en lelijkst beoordeeld werden, zijn gebruikt in dit onderzoek. Ter illustratie is een voorbeeld van elk handschrift toegevoegd. Deze zijn te vinden in Bijlage II. De handschrijvers kregen de teksten toegestuurd met een aantal instructies. Ze kregen de opdracht om de teksten op lijntjespapier over te schrijven, omdat dit gebruikelijk is bij geschreven schoolopdrachten. Daarnaast moesten ze ervoor zorgen dat ze de teksten exact overschreven zoals ze waren, inclusief fouten, witregels en spaties. Ten slotte werd tegen hen gezegd dat ze de teksten in hun normale handschrift moesten schrijven.

De teksten in de verschillende handschriften werden ingescand en samen met de getypte teksten samengesteld tot wisselende sets van vier teksten. De vier teksten werden telkens in dezelfde volgorde aangeboden, maar de presentatiewijze van de tekst kon verschillen. De volgorde van de vier presentatiewijzen was dus variabel. Elke set teksten werd voorafgegaan door een toestemmingsverklaring en een instructie. De instructie staat in Bijlage III.

### 2.3 Proefpersonen

In het experiment werden studentdocenten van de HAN ingezet als beoordelaars. Zij volgden op dat moment een opleiding tot docent Engels. Het betreft hier studentdocenten van verschillende jaarlagen. De masterstudenten werden opgeleid tot eerstegraadsdocenten en de bachelorstudenten waren bezig om hun tweedegraadsbevoegdheid te halen. In eerste instantie werden masterstudenten en vierdejaars bachelorstudenten benaderd. Na een wervingspraatje gaven 18 studenten aan mee te willen doen. Nadat bleek dat de respons laag was, zijn ook eerste- en tweedejaars bachelorstudenten benaderd. Allen waren moedertaalsprekers van het Nederlands en tweedetaalsprekers van het Engels. In totaal waren er 12 beoordelaars (drie mannen, negen vrouwen). Een vrouwelijke beoordelaar had slechts één van de vier teksten beoordeeld en is daarom niet meegenomen in de analyse. De participanten waren tussen de 19 en 50 jaar ( $M = 25$ ,  $SD = 8.46$ ). De beoordelaarsgroep bestond uiteindelijk uit drie masterstudenten, twee vierdejaars bachelorstudenten, drie tweedejaars bachelorstudenten en twee eerstejaars bachelorstudenten.<sup>1</sup>

### 2.4 Procedure

De masterstudenten en de vierdejaars bachelorstudenten kregen vier teksten op papier mee naar huis. De beoordelaars kregen de opdracht om de teksten te beoordelen op inhoud en taalgebruik. Om de twee criteria te specificeren, kregen de beoordelaars een lijst met punten waarop ze de teksten moesten beoordelen. Bij het beoordelen op deze punten konden de beoordelaars kiezen uit onvoldoende, matig, voldoende of goed. De lijst met de punten staat in Bijlage III. In de instructie werd de beoordelingsschaal expliciet gemaakt: de schaal liep van 1 tot en met 10 waarbij een 5.5 net voldoende was. De beoordelaars mochten zelf beslissen waar en wanneer ze de teksten beoordeelden. Ze kregen hiervoor een week de tijd. Dit was het meest ecologisch valide, aangezien docenten teksten normaal gesproken ook niet op een vast tijdstip of in groepsverband nakijken. Er werd benadrukt dat de beoordelaars niet met elkaar mochten overleggen over de teksten en de beoordeling, omdat de beoordelaars zo het doel van het experiment konden ontdekken. De procedure bij de tweede- en eerstejaars bachelorstudenten was iets anders. Wegens tijdgebrek van de onderzoeker beoordeelden zij de teksten in klassensetting onder begeleiding van de onderzoeker. Op deze manier was het zeker dat de studenten de beoordeelde teksten direct zouden inleveren. Hierdoor was de respons maximaal en hoefde de onderzoeker niet te wachten op de resultaten.

### 2.5 Data-analyse

Om de resultaten te analyseren zijn een eenweg-ANOVA, een tweeweg-ANOVA en twee herhaaldemetingen-ANOVA's toegepast op de data. De keuze viel op deze statistische procedures, omdat zo meerdere gemiddelden met elkaar vergeleken konden worden. De eenweg-ANOVA voor deelvraag 1 had de *between-subjects*factor presentatiewijze met vier niveaus (mooi, gemiddeld, lelijk, getypt). De herhaaldemetingen-ANOVA voor deelvraag 2 had de *within-subjects*factor criterium (inhoud, taalgebruik) en de *between-subjects*factor presentatiewijze. De tweeweg-ANOVA voor deelvraag 3 nam de *between-subjects*factoren presentatiewijze en taal (T1, T2) mee. Ten slotte werden voor deelvraag 4 twee herhaaldemetingen-ANOVA's uitgevoerd. Voor de eerste ANOVA werden alleen de Nederlandse teksten meegenomen in de analyse, voor de tweede alleen de Engelse teksten. De *within-subjects*factor was bij beide ANOVA's criterium en de *between-subjects*factor was presentatiewijze.

---

<sup>1</sup> Ik had de resultaten van een universitaire docent geschiedenis/politicologie en van een docent filosofie tot mijn beschikking. Wegens homogeniteit van de participantengroep laat ik deze resultaten buiten beschouwing.

### 3.0 Resultaten

#### 3.1 Gemiddelden en standaarddeviaties

De gemiddelden en standaarddeviaties van de totale scores en de inhoud- en taalgebruikscores staan beschreven in Tabel 1. In de tabel zijn deze scores voor teksten in verschillende talen en presentatiewijzen uitgezet. Deze tabel bevat de waarden die van belang zijn bij deelvraag 1, 2, 3 en 4. Met behulp van superscripten wordt aangegeven welke waarden bij welke deelvraag van belang zijn.

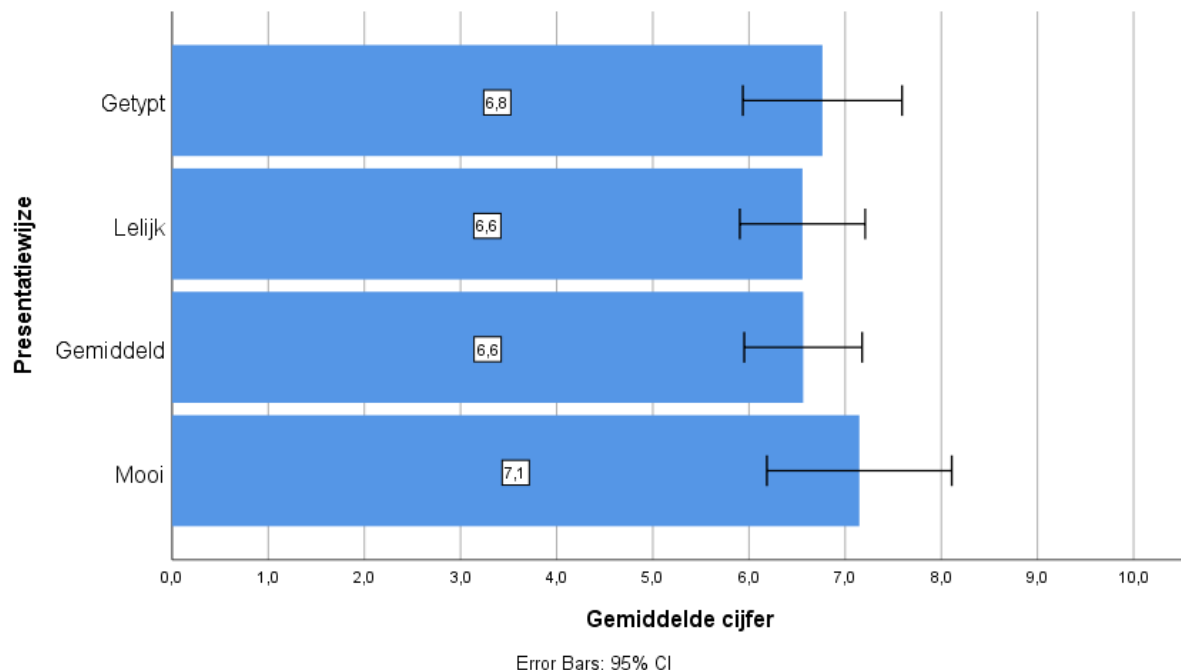
*Tabel 1: De gemiddelden en standaarddeviaties van de gemiddelde scores en de inhoud- en taalgebruikscores voor teksten in verschillende talen en presentatiewijzen*

Presentatiewijze	Taal	Criterium	Gemiddelde	Standaarddeviatie
Mooi	T1	Inhoud	7.6 <sup>4</sup>	1.50 <sup>4</sup>
		Taalgebruik	7.5 <sup>4</sup>	1.34 <sup>4</sup>
		Beide	7.6 <sup>3</sup>	1.35 <sup>3</sup>
	T2	Inhoud	6.7 <sup>4</sup>	1.30 <sup>4</sup>
		Taalgebruik	6.5 <sup>4</sup>	1.80 <sup>4</sup>
		Beide	6.6 <sup>3</sup>	1.46 <sup>3</sup>
	Beide	Inhoud	7.2 <sup>2</sup>	1.43 <sup>2</sup>
		Taalgebruik	7.0 <sup>2</sup>	1.57 <sup>2</sup>
		Beide	7.1 <sup>1</sup>	1.43 <sup>1</sup>
Gemiddeld	T1	Inhoud	6.3 <sup>4</sup>	1.06 <sup>4</sup>
		Taalgebruik	7.0 <sup>4</sup>	1.41 <sup>4</sup>
		Beide	6.6 <sup>3</sup>	1.24 <sup>3</sup>
	T2	Inhoud	6.6 <sup>4</sup>	1.11 <sup>4</sup>
		Taalgebruik	6.5 <sup>4</sup>	1.07 <sup>4</sup>
		Beide	6.6 <sup>3</sup>	.92 <sup>3</sup>
	Beide	Inhoud	6.5 <sup>2</sup>	1.06 <sup>2</sup>
		Taalgebruik	6.6 <sup>2</sup>	1.07 <sup>2</sup>
		Beide	6.6 <sup>1</sup>	.91 <sup>1</sup>
Lelijk	T1	Inhoud	6.5 <sup>4</sup>	1.12 <sup>4</sup>
		Taalgebruik	5.7 <sup>4</sup>	.97 <sup>4</sup>
		Beide	6.1 <sup>3</sup>	.96 <sup>3</sup>
	T2	Inhoud	7.1 <sup>4</sup>	1.05 <sup>4</sup>
		Taalgebruik	6.8 <sup>4</sup>	.76 <sup>4</sup>
		Beide	6.9 <sup>3</sup>	.87 <sup>3</sup>
	Beide	Inhoud	6.8 <sup>2</sup>	1.07 <sup>2</sup>
		Taalgebruik	6.3 <sup>2</sup>	.98 <sup>2</sup>
		Beide	6.6 <sup>1</sup>	.97 <sup>1</sup>
Getypt	T1	Inhoud	7.1 <sup>4</sup>	1.14 <sup>4</sup>
		Taalgebruik	6.3 <sup>4</sup>	1.62 <sup>4</sup>
		Beide	6.7 <sup>3</sup>	1.35 <sup>3</sup>
	T2	Inhoud	6.5 <sup>4</sup>	.71 <sup>4</sup>
		Taalgebruik	7.5 <sup>4</sup>	.71 <sup>4</sup>
		Beide	7.0 <sup>3</sup>	.71 <sup>3</sup>
	Beide	Inhoud	7.0 <sup>2</sup>	1.07 <sup>2</sup>
		Taalgebruik	6.5 <sup>2</sup>	1.54 <sup>2</sup>
		Beide	6.8 <sup>1</sup>	1.23 <sup>1</sup>

### 3.2 Het effect van presentatie

Mijn eerste deelvraag was: ‘Verschillen de beoordelingen van een mooi handgeschreven, een gemiddeld handgeschreven, een lelijk handgeschreven en een getypte tekst van elkaar?’ Het gemiddelde cijfer voor een tekst in een mooi handschrift was 7.1 ( $SD = 1.43$ ), voor een getypte tekst 6.8 ( $SD = 1.23$ ), voor een tekst in een gemiddeld handschrift 6.6 ( $SD = .91$ ) net als voor een tekst in een lelijk handschrift 6.6 ( $SD = .97$ ). Om te testen of deze verschillen significant zijn, is een eenweg-ANOVA uitgevoerd. Aan de assumpties van homogeniteit van varianties en normaliteit was voldaan. De assumptie van onafhankelijkheid was geschonden. Een beoordelaar scoorde telkens teksten in alle presentatiewijzen, waardoor deze scores niet onafhankelijk van elkaar zijn.

In Figuur 1 zijn de gemiddelde cijfers voor de verschillende presentatiewijzen uitgezet in een staafdiagram. De overlappende betrouwbaarheidsintervallen wijzen erop dat de gemiddelde cijfers niet significant van elkaar verschillen. Dit wordt bevestigd door een eenweg-ANOVA: de gemiddelde scores voor de vier presentatiewijzen bleken niet significant van elkaar te verschillen: de presentatiewijze had geen significant effect op de beoordeling ( $F(3,43) = .63, p = .599$ ).



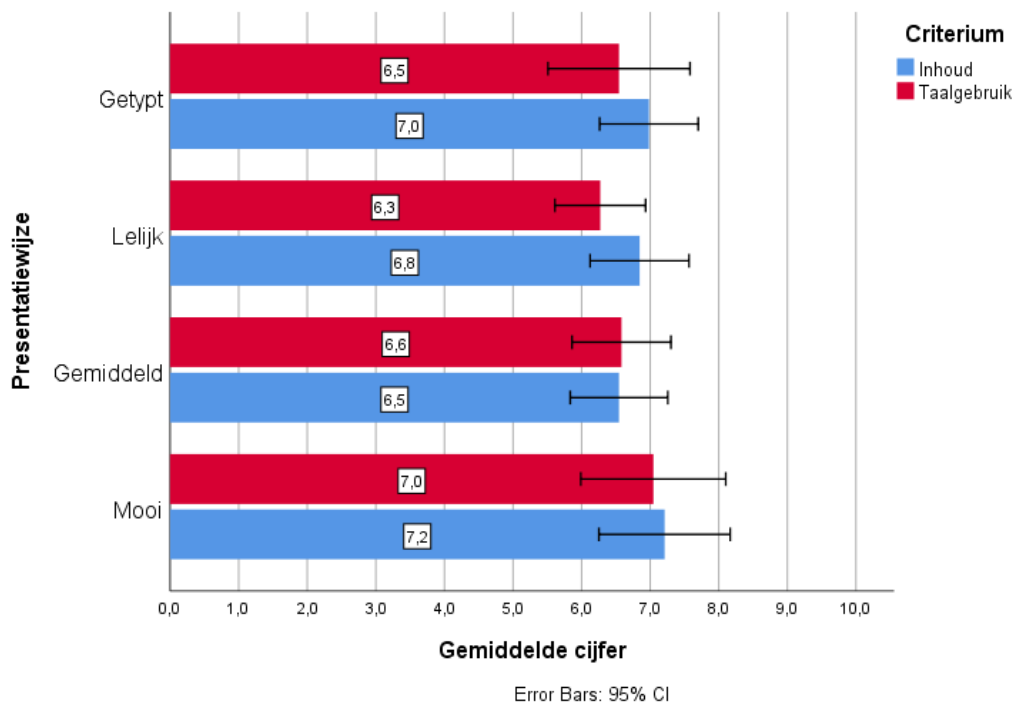
Figuur 1: Een staafdiagram van het gemiddelde cijfer voor een tekst in een bepaalde presentatiewijze met 95%-betrouwbaarheidsintervallen

### 3.3 Het interactie-effect van presentatie en criterium

Mijn tweede deelvraag was: ‘Zijn de verschillen tussen de beoordelingen van de teksten in de vier presentatiewijzen anders voor inhoud en taalgebruik?’ Het gemiddelde cijfer voor inhoud in een mooi handschrift was 7.2 ( $SD = 1.43$ ), voor een tekst in een gemiddeld handschrift 6.5 ( $SD = 1.06$ ), voor een tekst in een lelijk handschrift 6.8 ( $SD = 1.07$ ) en voor een getypte tekst 7.0 ( $SD = 1.07$ ). Het gemiddelde cijfer voor het taalgebruik in een tekst in een mooi handschrift was een 7.0 ( $SD = 1.57$ ), voor een tekst in een gemiddeld handschrift een 6.6 ( $SD = 1.07$ ) en voor een tekst in een lelijk handschrift een 6.3 ( $SD = .98$ ). Teksten in een getypt format kregen gemiddeld een 6.5 ( $SD = 1.54$ ). Om de deelvraag te testen, is een herhaaldemetingen-ANOVA uitgevoerd. Aan de assumptie van homogeniteit van

varianties was voldaan. De assumptie van sfericiteit was niet van toepassing. De scores op inhoud waren echter niet normaal verdeeld ( $W(44) = .94, p < .05$ ). De groepsgroottes zijn gelijk, waardoor de F-statistiek redelijk robuust is tegen de schending van deze assumptie.

In Figuur 2 zijn de gemiddelde cijfers voor taalgebruik en inhoud per presentatiewijze uitgezet. De overlappende betrouwbaarheidsintervallen indiceren dat de gemiddelde cijfers niet significant verschillen. Uit een herhaaldemetingen-ANOVA bleek dat er geen hoofdeffect van criterium was ( $F(1,40) = 3.85, p = .057$ ). Daarnaast had presentatiewijze geen significant effect op de beoordeling van inhoud en taalgebruik ( $F(3,43) = .90, p = .451$ ).



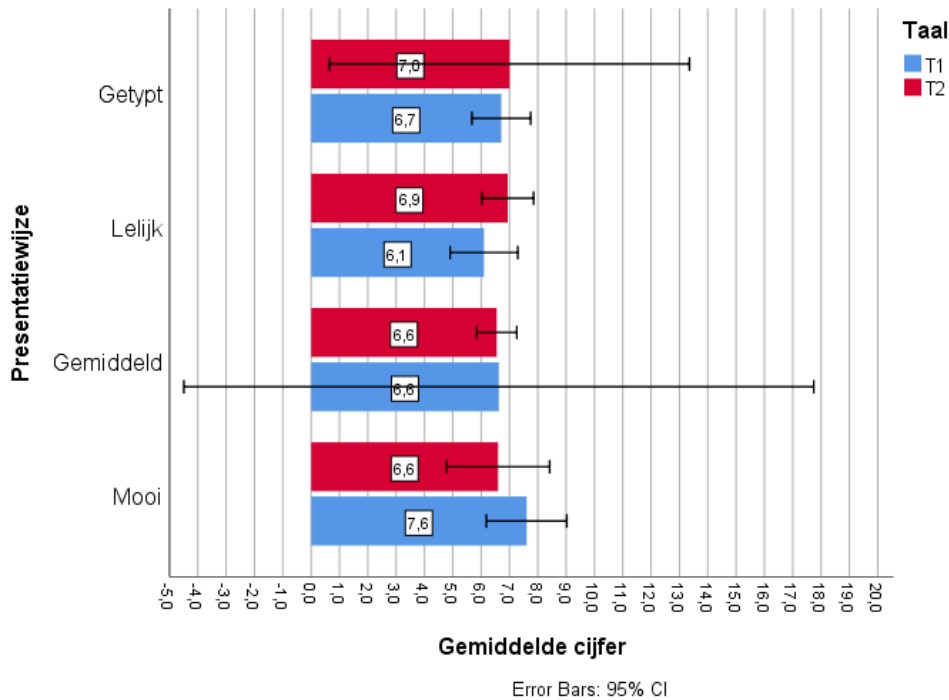
*Figuur 2: Een geclusterde staafdiagram van het gemiddelde cijfer voor inhoud en taalgebruik van teksten in een bepaalde presentatiewijze met 95%-betrouwbaarheidsintervallen*

### 3.4 Het interactie-effect van presentatie en taal

Mijn derde deelvraag was: 'Zijn de verschillen tussen de gemiddelde beoordelingen van de teksten in de vier presentatiewijzen anders voor teksten in de T1 en T2 van de beoordelaar?' De Nederlandse teksten in een mooi handschrift werden gemiddeld beoordeeld met een 7.6 ( $SD = 1.35$ ), die in een gemiddeld handschrift met een 6.6 ( $SD = 1.24$ ) en die in een lelijk handschrift met een 6.1 ( $SD = .96$ ). De Nederlandse teksten in een getypt format werden gemiddeld met een 6.7 ( $SD = 1.35$ ) beoordeeld. De Engelse teksten in een mooi handschrift werden gemiddeld met een 6.6 ( $SD = 1.46$ ) beoordeeld. De teksten in een gemiddeld handschrift werden ook gemiddeld met een 6.6 ( $SD = .92$ ) beoordeeld. De Engelse teksten in een lelijk handschrift werden gemiddeld met een 6.9 ( $SD = .87$ ) beoordeeld en de teksten in de getypte conditie met een 7.0 ( $SD = .71$ ). Aan de assumptie van homogeniteit van varianties en normaliteit was voldaan. De assumptie van onafhankelijkheid was geschonden. Een beoordelaar scoorde telkens teksten in alle presentatiewijzen en in beide talen, waardoor deze scores niet onafhankelijk van elkaar zijn.

In Figuur 3 zijn de gemiddelde cijfers per presentatie uitgezet voor de twee talen. De betrouwbaarheidsintervallen voor de Nederlandse teksten in een gemiddeld handschrift en de Engelse teksten in een getypt format zijn zo groot, omdat het aantal beoordelingen in die categorieën slechts 2 is. Met een tweeweg-ANOVA werd geen significant hoofdeffect van

taal gevonden ( $F(1, 44), p = .977$ ). Daarnaast werd er geen significant interactie-effect tussen de presentatiewijze en de taal van de tekst gevonden ( $F(7, 36) = 1.19, p = .328$ ).



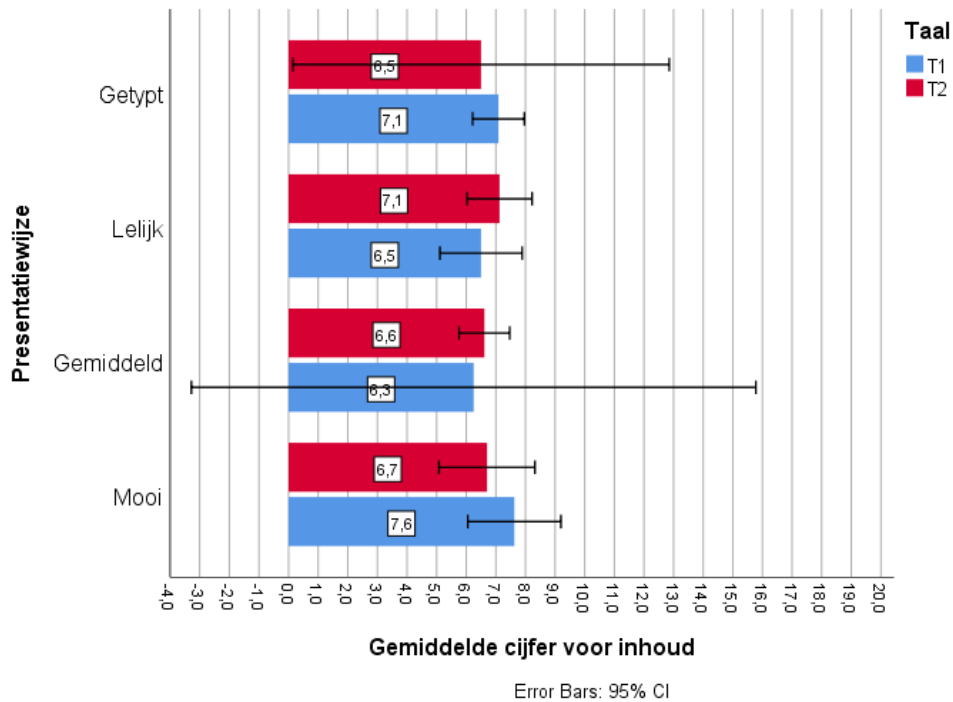
*Figuur 3: Een geclusterde staafdiagram van het gemiddelde cijfer voor teksten in een bepaalde presentatiewijze, uitgesplitst naar de taal van de tekst met 95%-betrouwbaarheidsintervallen*

### 3.5 Het interactie-effect van presentatie en criterium in de T1 en T2

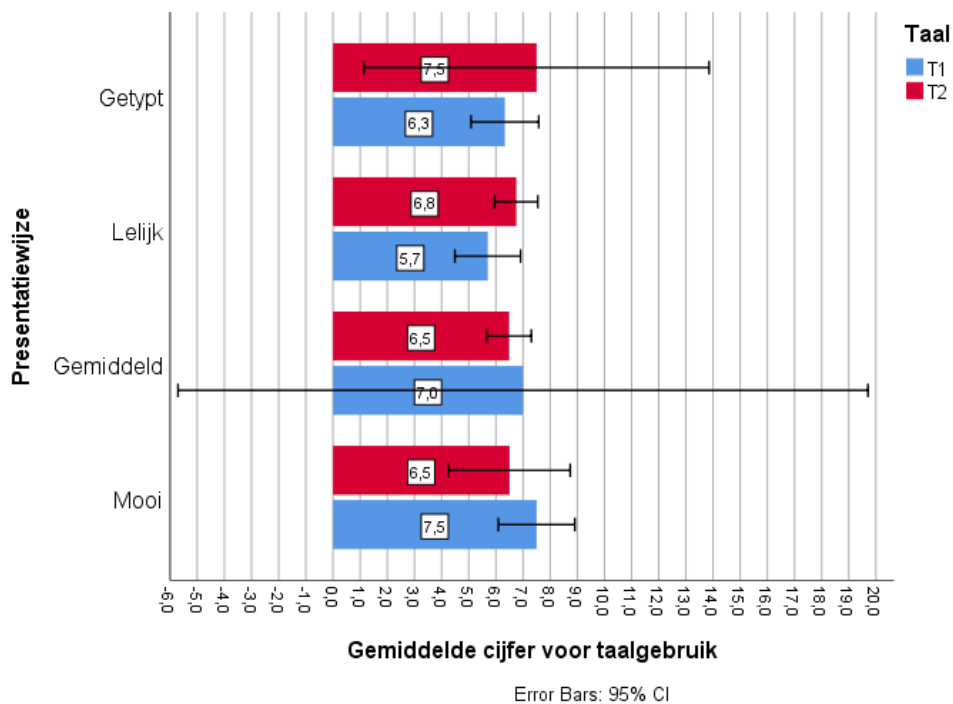
Mijn laatste deelvraag was: ‘Zijn de verschillen tussen de beoordelingen van de teksten in de vier presentatiewijzen anders voor inhoud en taalgebruik afhankelijk van de taal waarin de tekst is geschreven?’ Om deze vraag te kunnen beantwoorden, zijn twee herhaaldemetingen-ANOVA’s uitgevoerd. Aan de assumptie van homogeniteit van varianties was voldaan. De assumptie van sfericiteit was niet van toepassing. De scores voor inhoud bij de Engelse teksten waren niet normaal verdeeld ( $W(22) = .90, p < .05$ ).

In Figuur 4 en 5 zijn de gemiddelde cijfers voor inhoud en taalgebruik uitgezet voor de twee talen. Wederom zijn de betrouwbaarheidsintervallen van de Nederlandse teksten in het gemiddelde handschrift en de getypte Engelse teksten groot, omdat die categorieën een klein aantal beoordelingen bevatten. Uit de herhaaldemetingen-ANOVA bleek dat er binnen de Nederlandse teksten geen significant interactie-effect was tussen presentatiewijze en criterium ( $F(3, 18) = 2.5, p = .092$ ). Ook binnen de Engelse teksten was er geen significant interactie-effect tussen presentatiewijze en criterium ( $F(3, 18) = 1.0, p = .417$ ).





Figuur 4: Een geclusterde staafdiagram van het gemiddelde cijfer voor de inhoud van teksten in een bepaalde presentatiewijze, uitgesplitst naar de taal van de tekst met 95%-betrouwbaarheidsintervallen



Figuur 5: Een geclusterde staafdiagram van het gemiddelde cijfer voor het taalgebruik in teksten in een bepaalde presentatiewijze, uitgesplitst naar de taal van de tekst met 95%-betrouwbaarheidsintervallen

#### *4.0 Discussie*

Het doel van dit onderzoek was om te ontdekken of teksten anders beoordeeld worden afhankelijk van de manier waarop de tekst gepresenteerd is. Daarbij heb ik onderzocht of het presentatie-effect verschilt naar gelang het criterium dat beoordeeld wordt en de taal waarin de tekst geschreven is. Hiermee hoop ik de kennis op dit gebied uit te breiden, aangezien er weinig recent onderzoek naar dit fenomeen is en de factoren criterium en taal nog niet eerder meegenomen waren.

#### *4.1 Bespreking resultaten*

Uit dit onderzoek blijkt dat er geen significant hoofdeffect van presentatiewijze, taal of criterium is. Tevens is er geen significant verschil in het presentatie-effect tussen de verschillende niveaus van criterium en taal gevonden. Ook wanneer per taal wordt bekeken of het presentatie-effect groter is bij een bepaald criterium, blijkt dit niet zo te zijn. De conclusie op basis van de huidige studie is dus dat de beoordelingen van een mooi handgeschreven, een gemiddeld handgeschreven, een lelijk handgeschreven en een getypte tekst niet aantoonbaar van elkaar verschillen. Het presentatie-effect is niet significant anders wanneer de scores op inhoud en taalgebruik worden vergeleken of wanneer de tekst in de T1 of T2 geschreven is. Ten slotte is het presentatie-effect op de scores op inhoud en taalgebruik niet significant anders wanneer T1 en T2 apart worden genomen.

De resultaten van het huidige onderzoek laten zien dat er geen presentatie-effect is. Een gevolg van deze resultaten zou kunnen zijn dat het overbodig blijkt om docenten te trainen op het presentatie-effect. Dit spreekt echter meerdere onderzoeken tegen. Uit de onderzoeken van Powers en collega's (1994) en Russell en Tao (2004b) bleek namelijk dat er wél een presentatie-effect was en dat training het effect verminderde of deed verdwijnen. De resultaten van het huidige onderzoek spreken ook verschillende andere studies naar dit fenomeen tegen. Onder andere Klein en Taub (2005) vonden verschillen in de beoordeling van teksten op basis van de leesbaarheid van het essay en het medium waarin het essay gepresenteerd werd. Daarnaast vond onder andere Markham (1976) een significant verschil tussen teksten in een mooi en een lelijk handschrift. De resultaten in de huidige studie komen overeen met de resultaten van één ander onderzoek: in het onderzoek van Harrington en collega's (2000) werden namelijk ook geen significante verschillen gevonden. Hun onderzoek had een iets andere opzet dan het huidige onderzoek: zij vergeleken de beoordelingen van handgeschreven, getypte en omgezette (van handgeschreven naar getypt) teksten met elkaar.

De resultaten in de huidige studie spreken de eerder geformuleerde verwachtingen tegen. Ik verwachtte verschillen in de beoordelingen van de teksten in de vier presentatiewijzen te vinden. Ik verwachtte dat teksten in een mooi handschrift en getypte teksten beter beoordeeld zouden worden dan teksten in een lelijk handschrift. Daarnaast vermoedde ik dat de beoordelingen van een tekst in een mooi handschrift en een getypte tekst onderling zouden verschillen. Een andere verwachting was dat het effect van presentatie anders zou zijn bij de beoordelingen van taalgebruik en inhoud. Daarnaast verwachtte ik dat het presentatie-effect groter zou zijn wanneer de tekst in de T2 van de beoordelaar geschreven was. Ten slotte vermoedde ik dat het presentatie-effect in de T2 groter zou zijn bij inhoud dan bij taalgebruik. Het omgekeerde verwachtte ik in de T1: ik dacht dat het presentatie-effect daar groter zou zijn bij taalgebruik dan bij inhoud.

Wegens de conflicterende resultaten met vele andere studies naar het presentatie-effect en mijn verwachtingen ben ik terughoudend met het generaliseren van de resultaten van de huidige studie naar de populatie van docenten Engels.

## 4.2 Mogelijke verklaringen resultaten

De achterliggende verklaring van de nulresultaten kan allereerst zijn dat de beoordelaars in de huidige studie goede beoordelaars zijn die zich niet laten beïnvloeden door de presentatiewijze van een tekst. De uitvoering van het experiment in de huidige studie kent echter een aantal tekortkomingen. Dit roept de vraag op of de resultaten geldig zijn. Ik zal een aantal tekortkomingen vanuit een kritisch oogpunt bespreken.

### 4.2.1. Steekproefgrootte

De steekproefgrootte in de huidige studie was klein; slechts 11 beoordelaars participeerden. Een te kleine steekproef representeert de populatie niet goed, waardoor de mogelijkheid ontstaat dat het onderzoek en de getrokken conclusies niet betrouwbaar zijn. Om dit te ondersteunen, is de *power* van de toetsen berekend. De toetsen die gebruikt zijn bij deelvraag 1 en 2 hadden een lage *power* van 0.23. De toets die gebruikt is bij deelvraag 3 had een hoge *power* van 0.95. De toetsen die gebruikt zijn bij deelvraag 4 hadden een *power* van 0.80 (voor de teksten in de T1) en 0.48 (voor de teksten in de T2). De *power* van de toets bij deelvraag 3 en van de toets van de T1-teksten bij deelvraag 4 was hoog: de kans dat een eventueel effect met effectgrootte 0.25 werd opgepikt was groot. De kleine steekproefgrootte was alleen bij deelvraag 1, 2 en deels bij 4 een probleem. De *power* van die toetsen was laag, wat betekent dat de kans dat een effect gedetecteerd werd, erg klein was, gegeven dat er een effect met een middelmatige effectgrootte van 0.25 zou zijn.

### 4.2.2. Bewustzijn

Het is mogelijk dat sommige beoordelaars zich bewust waren van het doel van het onderzoek. De docente van de vierdejaars bachelorstudenten had bij mijn introductie voor het wervingspraatje genoemd dat mijn onderzoek met handschriften te maken had. De kans is groot dat de vierdejaars bachelorstudenten hierdoor beïnvloed zijn. Daarnaast kan het zo zijn dat het de andere beoordelaars was opgevallen dat zij teksten in vier verschillende presentatiewijzen moesten beoordelen, waardoor ze zich bewust waren van dit verschil. Uit onderzoeken van Powers et al. (1994) en Russell en Tao (2004b) bleek dat het presentatie-effect verminderde of verdween wanneer de beoordelaars getraind waren om te letten op het presentatie-effect. Wanneer de beoordelaars in de huidige studie op de hoogte waren van het doel van het onderzoek, kan dat de resultaten hebben beïnvloed.

### 4.2.3. Homogeniteit beoordelaars

In de huidige studie is getracht om de homogeniteit van de beoordelaarsgroep te waarborgen. De beoordelaars waren echter niet op alle punten vergelijkbaar. Het niveau van de Engelse taal kan bijvoorbeeld niet als homogeen verondersteld worden. De beoordelaarsgroep varieert van eerstejaars bachelorstudenten tot masterstudenten. De masterstudenten zijn naar alle waarschijnlijkheid vaardiger in het Engels en hebben meer kennis over het Engels en lesgeven, omdat zij verder zijn in hun studie. Dit kan ervoor hebben gezorgd dat de verschillen in beoordelingsgedrag bij het beoordelen van een tekst in de T1 of de T2 niet even groot waren voor de verschillende beoordelaars.

### 4.2.4 Analyse resultaten

Bij de analyses is gebruikt gemaakt van een eenweg-ANOVA, een tweeweg-ANOVA en twee herhaaldemetingen-ANOVA's. In de resultatensectie werd al genoemd dat niet aan alle assumpties van de ANOVA's voldaan was. De reden dat de ANOVA's toch zijn uitgevoerd, is omdat deze manier van toetsen het dichtst in de buurt komt bij de juiste statistische toetsing. Een *multilevel model* kan in sommige gevallen de schending van onafhankelijkheid oplossen. Deze statistische toetsing valt echter buiten mijn kennis van

statistiek. De schending van de assumptie van onafhankelijkheid leidt ertoe dat de kans op een Type I-fout wordt vergroot. Hierdoor kunnen fout-positieven worden gevonden. De resultaten in de huidige studie waren niet significant, waardoor Type I-fouten uitgesloten kunnen worden. Ook is de assumptie van normaalverdeeldheid geschonden waar de groepsgroottes niet gelijk waren. Een schending van normaliteit kan de *power* van F in dit geval onvoorspelbaar beïnvloeden. Bij het interpreteren van de resultaten moet rekening worden gehouden met de schendingen van de assumpties: de resultaten zijn niet volledig betrouwbaar.

#### *4.3 Conclusie*

Op basis van dit onderzoek kan geconcludeerd worden dat in de huidige studie de beoordelingen van een mooi handgeschreven, een gemiddeld handgeschreven, een lelijk handgeschreven en een getypte tekst niet van elkaar verschillen. Ook verschillen de beoordelingen van de teksten in de vier presentatiewijzen niet significant anders op het gebied van inhoud en taalgebruik of op basis van de taal waarin de tekst is geschreven. Ik ben terughoudend met het generaliseren van deze conclusie naar de populatie wegens de kleine steekproefgrootte en de conflicterende resultaten met andere studies. Het zou interessant zijn om dit onderzoek te herhalen met een aantal verbeteringen. Bij een replicatieonderzoek zouden bovenstaande tekortkomingen in gedachten moeten worden gehouden. Ik zou aanraden om een grotere steekproef te nemen, alle betrokkenen zo naïef mogelijk te houden en de homogeniteit van de beoordelaarsgroep te waarborgen. Dit komt de betrouwbaarheid van het onderzoek en de analyse ten goede.

## Literatuurlijst

- Arnold, V., Legas, J., Obler, S., Pacheco M.A., Russell, C., & Umbdenstock, L. (1990). *Do Students Get Higher Scores on Their Word-Processed Papers? A Study of Bias in Scoring Hand-Written vs. Word-Processed Papers*. Ongepubliceerd manuscript, Rio Hondo College, Whittier, CA.
- Árva, V., & Medgyes, P. (2000). Native and non-native teachers in the classroom. *System*, 28(3), 355-372.
- Braine, G. (2010). Non-native-speaker English teachers: Research, Pedagogy, and Professional Growth. *The Encyclopedia of Applied Linguistics*, 17-29.
- Bridgeman, B., & Cooper, P. (1998). *Comparability of scores on word-processed and handwritten essays on the Graduate Management Admission Test*. Gepresenteerd op de jaarlijkse bijeenkomst van de American Educational Research Association, San Diego, CA.
- Briggs, D. (1970). The influence of handwriting on assessment. *Educational Research*, 13(1), 50-55.
- Graham, S., Weintraub, N., & Berninger, V. W. (1998). The relationship between handwriting style and speed and legibility. *The Journal of Educational Research*, 91(5), 290-297.
- Greifeneder, R., Alt, A., Bottenberg, K., Seele, T., Zelt, S., & Wagener, D. (2010). On writing legibly: Processing fluency systematically biases evaluations of handwritten material. *Social Psychological and Personality Science*, 1(3), 230-237.
- Harrington, S., Shermis, M. D., & Rollins, A. L. (2000). The influence of word processing on English placement test results. *Computers and Composition*, 17(2), 197-210.
- Hecht, S. A., & Greenfield, D. B. (2002). Explaining the predictive accuracy of teacher judgments of their students' reading achievement: The role of gender, classroom behavior, and emergent literacy skills in a longitudinal sample of children exposed to poverty. *Reading and Writing*, 15(7-8), 789-809.
- Johnson, J. S., & Lim, G. S. (2009). The influence of rater language background on writing performance assessment. *Language Testing*, 26(4), 485-505.
- Kim, A.Y., & Di Gennaro, K. (2012). Scoring Behavior of Native vs. Non-native Speaker Raters of Writing Exams. *Language Research* 48.2, 319-342.
- Klein, J., & Taub, D. (2005). The effect of variations in handwriting and print on evaluation of student essays. *Assessing writing*, 10(2), 134-148.
- Kobrin, J. L., Deng, H., & Shaw, E. J. (2007). Does Quantity Equal Quality? The Relationship Between Length of Response and Scores on the SAT Essay. *Journal of Applied Testing Technology*, 8(1).
- Kobrin, J. L., Deng, H., & Shaw, E. J. (2011). The association between SAT prompt characteristics, response features, and essay scores. *Assessing Writing*, 16(3), 154-169.
- Kohler, B. (2015). Based or Computer-Based Essay Writing: Differences in Performance and Perception. *Linguistic Portfolios*, 4(1), 13.
- Lee, H. K. (2009). Native and nonnative rater behavior in grading Korean students' English essays. *Asia Pacific Education Review*, 10(3), 387-397.
- MacCann, R., Eastment, B., & Pickering, S. (2002). Responding to free response examination questions: Computer versus pen and paper. *British Journal of Educational Technology*, 33(2), 173-188.
- Markham, L. R. (1976). Influences of handwriting quality on teacher evaluation of written work. *American Educational Research Journal*, 13(4), 277-283.
- Marshall, J. C., & Powers, J. M. (1969). Writing neatness, composition errors, and essay grades. *Journal of Educational Measurement*, 6(2), 97-101.
- Mattern, K., Camara, W., & Kobrin, J. L. (2007). *SAT writing: An overview of research and psychometrics to date*. (College Board Research Note No. RN-32). New York: College Entrance Examination Board.
- Medgyes, P. (1994). *The non-native teacher*. London: Macmillan.
- Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48-60.

- Park, S. K. (2015). The Interplay of Task, Rating Scale, and Rater Background in the Assessment of Korean EFL Students' Writing. *English Teaching, 70*(2).
- Peacock, M. (1988). Handwriting versus wordprocessed print: an investigation into teachers' grading of English Language and Literature essay work at 16+. *Journal of Computer Assisted Learning, 4*(3), 162-172.
- Perelman, L. (2005, 29 mei). New SAT: Write Long, Badly and Prosper. *Los Angeles Times*. Geraadpleegd van <https://www.latimes.com/archives/la-xpm-2005-may-29-oe-perelman29-story.html>
- Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1994). They Think Less of My Handwritten Essay If Others Word Process Theirs? Effects on Essay Scores of Intermingling Handwritten and Word-Processed Essays. *Journal of Educational Measurement, 31*(3), 220-233.
- Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal, 48*(2), 335-360.
- Russell, M., & Tao, W. (2004a). Effects of handwriting and computer-print on composition scores: A follow-up to Powers, Fowles, Farnum, & Ramsey. *Practical Assessment, Research and Evaluation, 9*(1).
- Russell, M., & Tao, W. (2004b). The influence of computer-print on rater scores. *Practical Assessment, Research and Evaluation, 9*(10), 1-14.
- Schoonen, R., van Gelderen, A., Stoel, R. D., Hulstijn, J., & de Glopper, K. (2011). Modeling the development of T1 and EFL writing proficiency of secondary school students. *Language learning, 61*(1), 31-79.
- Sweedler-Brown, C. O. (1985). The influence of training and experience on holistic essay evaluations. *The English Journal, 74*(5), 49-55.
- Sweedler-Brown, C. O. (1991). Computers and assessment: The effect of typing versus handwriting on the holistic scoring of essays. *Research and Teaching in Developmental Education, 5*-14.

Bijlage I: Teksten uit het onderzoek

*Tekst 1 (144 woorden, score van 115.5)*

Dear readers,

Everybody knows that the students in England have to wear uniforms. It's different in Holland. You make your own decision about what you're going to wear at school. You look different every day. I think that wearing uniforms at school is a bad rule.

Of course there are good and bad sides about wearing uniforms at school. When everybody is wearing the same clothes, nobody will bully you because you are always wearing the same clothes. And you are not wasting your time by making so many decisions every day. But when everybody looks the same, it will be boring at school just like school already is. I want that school is a place that everybody like it. School is already boring and nobody wants to go there, so don't make it worse by obliging the students to wear uniforms!!

K. R.

---

*Tekst 2 (200 woorden, score van 113.5)*

Hi Keith, it's me, F. Thanks a lot for your letter. I'm flattered that you asked me to write an article about English school uniforms.

I think that wearing school uniforms is better for everyone. I know that a lot of people disagree with me on that, but I have my reasons for thinking this.

First of all, almost everyone is being judged by their appearance, and then of course, by their clothes, or the way they wear certain clothes. And I think that if everyone would be wearing the same outfit, that there would be a lot less teasing and people being prejudiced, because everybody would be the same, in some ways.

And, yes, I know that it's important for everyone to be or look individually, so I suggest that the uniform-wearing shouldn't be too strict. By that I mean that girls for example, could wear certain things in their hair that they like, or that boys could decide whenever they want to wear long, short or no sleeves. Those are just simple examples, but that's just to show that everybody

can be different, and still look the same.

That's why I think that it's best that students should wear uniforms.

Your friend, F.

*Tekst 3 (262 woorden, score van 114.0)*

Er is de laatste tijd veel gediscussieerd over 'skaters'.(Mensen op skeelers.)  
Waar kunnen die mensen nou het beste "rijden"? Ik vind het fietspad het beste. Want zeg nou zelf, de stoep is superonhandig voor skaters! Ik bedoel, het is niet glad, je struikelt heel makkelijk dankzij die richeltjes tussen de tegels in, en de slome voetgangers zitten ook nog eens in de weg als je net lekker aan het skaten bent!

Daarom stel ik voor dat alle skaters verplaatst moeten worden naar het fietspad! Want het is gladder dan de stoep, fietsers zijn niet zo langzaam dus minder hinderlijk en voetgangers hebben ook geen last meer van die skaters die zo snel voorbij sjeesen. Volgens mij gebeuren ook de meeste ongelukken door het feit dat skaters tegen voetgangers opbotsen of dat skaters struikelen op het fietspad door het onregelmatige 'wegdek' genaamd de stoep.

Ik vind het vanzelfsprekend dat skaters niet de rijweg op moeten, want auto's zijn natuurlijk veel te snel en minder wendbaar dan fietsers of voetgangers. Ik wil niet weten wat een bloedbad er worden aangericht als skaters opeens die gevaarlijke rijweg op moeten!

(ook al moet ik toegeven de rijweg wel lekker glad is!)

En zeg nou zelf, wat voor nadelen zijn er als skaters het fietspad op moeten? Bijna geen een. Misschien dat de (brom-)fietsers gaan klagen dat het te druk wordt ofzo, maar dat weegt niet op tegen de argumenten van de anderen.

Daarom vind ik dat skaters het fietspad op moeten. ( Ze zouden het eigenlijk geen 'fietspad' meer moeten noemen, maar 'wielpad' ofzo...)

S. H.

---

*Tekst 4 (203 woorden, score van 115)*

Beste lezers,

Er is de laatste tijd veel discussie over het verkeer: 'wie mag of moet waar rijden?'. De auto's moeten op de rijweg, de mensen lopen op de stoep, de fietsers moeten op het fietspad, brommers en dergelijke moeten ook op de rijweg, maar waar mogen of moeten de skeelers rijden? Er is geen speciale weg aangelegd dat bestemd is voor de skeelers. Ik vind dat de skeelers op het fietspad moeten rijden. Dat is de beste oplossing.

De skeelers hebben een snelheid dat groter is dan de voetgangers, maar weer kleiner is dan de fietsers (of ongeveer gelijk). Het is gevaarlijk voor de voetgangers als de skeelers op de stoep moeten rijden. Er lopen veel kleine kinderen rond die geen rekening kunnen houden met snelle skeelers die plotseling opduiken. Er kunnen gemakkelijk ongelukken gebeuren. Ook lopen er oude mensen op de stoep, zij lopen meestal nogal langzaam. De skeelers kunnen dan geen vaart maken, dat is dan wel een beetje zonde, ze kunnen dan net zo goed gaan lopen. Op de rijweg is het weer te gevaarlijk voor de skeelers, ze kunnen makkelijk aangereden worden. Bovendien hebben ze ook weinig bescherming. Dus op het fietspad is de beste oplossing!!

M. L.



## Bijlage II: Voorbeelden van de drie handschriften

### Mooi handschrift

Er is de laatste tijd veel gediscussieerd over 'skaters' (Mensen op scoelers.)  
Waar kunnen die mensen nou het beste "rijden"? Ik vind het fietspad het beste. Want zeg nou zelf, de stoep is superonhandig voor skaters! Ik bedoel, het is niet glad, je struikelt heel makkelijk dankzij die richeltjes tussen de tegels in, en de slome voetgangers zitten ook nog eens in de weg als je net lekker aan het skaten bent!

### Gemiddeld handschrift

Er is de laatste tijd veel gediscussieerd over 'skaters' (Mensen op scoelers.)  
Waar kunnen die mensen nou het beste "rijden"?  
Ik vind het fietspad het beste. Want zeg nou zelf, de stoep is superonhandig voor skaters!  
Ik bedoel, het is niet glad, je struikelt heel makkelijk dankzij die richeltjes tussen de tegels in, en die slome voetgangers zitten ook nog eens in de weg als je net lekker aan het skaten bent!

### Lelijk handschrift

Er is de laatste tijd veel gediscussieerd over 'skaters' (Mensen op scoelers.)  
Waar kunnen die mensen nou het beste "rijden"? Ik vind het fietspad het beste. Want zeg nou zelf, de stoep is superonhandig voor skaters! Ik bedoel, het is niet glad, je struikelt heel makkelijk dankzij die richeltjes tussen de tegels in, en die slome voetgangers zitten ook nog eens in de weg als je net lekker aan het skaten bent!

### Bijlage III: Instructie voor het experiment

Bedankt dat je deel wilt nemen aan dit onderzoekje! Je hebt vier teksten meegekregen om te beoordelen. Deze teksten zijn willekeurig gekozen uit een set van teksten van middelbare scholieren. De middelbare scholieren zaten op dat moment in havo 4. De scholieren hebben korte essays geschreven over het onderwerp skaten of schooluniformen, zowel in het Engels als in het Nederlands. De essays moeten beoordeeld worden op inhoud en taalgebruik aan de hand van het bijgevoegde beoordelingsformulier. Uiteindelijk geef je elke tekst apart een cijfer voor de inhoud en het taalgebruik. Ga uit van een beoordelingsschaal van 1 tot en met 10 waarbij een 5,5 net voldoende is.

Bijlage IV: Beoordelingsformulier voor het experiment

<b>Inhoud</b>	<b>Criterium</b>	<b>Beoordeling</b>	<b>Cijfer voor inhoud</b>
	Is het onderwerp goed afgebakend?	onvoldoende / matig / voldoende / goed	
	Zijn de tekstrelaties duidelijk?	onvoldoende / matig / voldoende / goed	
	Is de argumentatie juist?	onvoldoende / matig / voldoende / goed	
	Wordt de centrale vraag beantwoord?	onvoldoende / matig / voldoende / goed	
	Is de tekst logisch opgebouwd?	onvoldoende / matig / voldoende / goed	
<b>Taalgebruik</b>	<b>Criterium</b>	<b>Beoordeling</b>	<b>Cijfer voor taalgebruik</b>
	Is het woordgebruik gevarieerd en is de woordkeuze correct?	onvoldoende / matig / voldoende / goed	
	Zijn de spelling en grammatica correct?	onvoldoende / matig / voldoende / goed	
	Is de tekst zorgvuldig geformuleerd?	onvoldoende / matig / voldoende / goed	
	Is het taalgebruik goed afgestemd op de lezer?	onvoldoende / matig / voldoende / goed	
	Is de tekst in een aansprekende stijl geschreven?	onvoldoende / matig / voldoende / goed	