

Ten Geleide

CEF of cijfers bij de beoordeling van schrijfvaardigheid¹

Dit nummer opent met een artikel van Huub van den Bergh en Lianne Klein Gunnewiek. Zij hebben docenten Duitse en Engels schrijfproducten van Nederlandse leerlingen uit de derde en vijfde klas van het voortgezet onderwijs laten beoordelen op basis van de criteria uit het Common European Framework en op basis van cijfers. Nederlandse en Duitse docenten Duits hebben de Duitstalige schrijfproducten beoordeeld, en de Nederlandse en Engelse docenten Engels hebben de Engelstalige schrijfproducten beoordeeld.

Ongeacht het soort beoordeling konden verschillen in kwaliteit van teksten van de derde- en vijfdeklassers aangetoond worden. Verrassend genoeg waren de verschillen in gemiddeld oordeel tussen de Nederlandse docenten en hun buitenlandse collega's wel groter wanneer gebruik gemaakt werd van het Common European Framework: de buitenlandse docenten Duits en Engels waren een stuk coulanter in hun oordeel dan de Nederlandse docenten.

In het tweede artikel, van Geert van de Ven, passeren kenmerken de revue van leerzame en ook uitdagende schrijftaken. Om zelf zulke schrijftaken te ontwikkelen dient men de centrale vraag binnen de schrijftaak af te stemmen op het taal- en denkniveau van de leerlingen. Het Common European Framework of Reference kan volgens de auteur daarbij van pas komen. Om leerlingen te motiveren voor het schrijven van een tekst dient men bovendien geschikte pre-writingactiviteiten te organiseren. De auteur beargumenteert hoe het Knowledge Framework van Mohan daartoe aanknopingspunten biedt.

Sinds het vorige schooljaar kan een school besluiten een vakonderdeel taalkunde bij het vak Nederlands in te voeren en te toetsen in het schoolexamen. De auteur van de derde bijdrage in dit nummer, Jessica van Disseldorp, is op zoek naar een antwoord op de vraag hoe taalkunde als vakonderdeel gegeven kan worden zonder dat het weinig of geen extra tijd kost. Door taalkunde aan te bieden in combinatie met schrijfvaardigheid (leerlingen maken een werkstuk over de eigen taal) dan wel mondelinge taalvaardigheid (leerlingen houden een voordracht over de eigen taal), ontwikkelen de leerlingen hun taalvaardigheid en doen kennis op via de teksten die ze lezen en de voordrachten waar ze naar luisteren of die ze zelf houden. Bijkomend voordeel: de leerlingen vinden de eigen taal een aantrekkelijk onderwerp om zich in te verdiepen.

In de vierde bijdrage beschrijft Marianne Hermans de reacties van leerlingen uit de bovenbouw van havo en vwo op zogenoemde multiculturele verhalen (een definitie vindt de lezer in het artikel). Leerlingen lezen verhalen van Marokkaans-Nederlandse auteurs en beantwoordden vragen over hun begrip en waardering van het verhaal. Leerlingen die zelf een niet-Nederlandse achtergrond hebben, reageerden anders op de verhalen dan de autochtone leerlingen. De auteur concludeert dat verhalen met een multiculturele setting een toegevoegde waarde kunnen hebben in de literatuurles, ook voor leerlingen die niet van huis uit met deze thema's bekend zijn.

Namens de redactie,
Carla Driessen

HUUB VAN DEN BERGH & LISANNE KLEIN GUNNEWIEK

In deze studie zijn Duitse en Engels schrijfproducten van Nederlandse leerlingen uit de derde en vijfde klas beoordeeld met en zonder de criteria uit het Common European Framework (CEF). De Duitstalige schrijfproducten zijn beoordeeld door Nederlandse en Duitse docenten Duits, en de Engelstalige schrijfproducten zijn beoordeeld door Nederlandse en Engelse docenten Engels. Conform de resultaten uit eerder onderzoek bleken de oordelen met behulp van de omschrijvingen uit het CEF zeer betrouwbaar. Zowel met de oordelen op basis van de in een land gebruikelijke cijferschaal als met die op basis van het CEF kunnen verschillen in kwaliteit van teksten van derde- en vijfdeklassers aangetoond worden. Echter, de verschillen in gemiddeld oordeel tussen de Nederlandse docenten en hun buitenlandse collegae bleken groter wanneer gebruik gemaakt wordt van het CEF. Hierbij bleken de buitenlandse docenten Duits en Engels duidelijk coulanter in hun oordeel.

Het Common European Framework (CEF) is een zeer veel omvattend systeem. Het CEF beoogt een richtsnoer te zijn voor curriculumontwikkeling en een basis te zijn voor de ontwikkeling van onderwijsmethoden voor

de vreemde talen. Hiervoor worden algemene omschrijvingen gegeven over wat leerders moeten leren om een vreemde taal als communicatiemiddel te kunnen gebruiken, maar ook welke kennis en vaardigheden zij moeten ontwikkelen (Council of Europe, 2001, 2003, 2004). Daarnaast is een niveauomschrijving vastgesteld, zodat de prestaties van leerders aan een referentiekader gekoppeld kunnen worden. Het belang van een dergelijk referentiekader kan moeilijk overschat worden. Het CEF biedt een gemeenschappelijke 'taal', waardoor deskundigen over ontwikkeling en niveaus kunnen praten in een zelfde perspectief. Dit bevordert een grotere eenheid tussen de leden van de EU, want het is nu mogelijk om met elkaar van gedachten te wisselen over het taalniveau. Met behulp van het CEF kan het niveau van leerlingen uit verschillende landen eenvoudig langs dezelfde meetlat gelegd worden (Council of Europe, 2001, 2003, 2004). Het CEF biedt een omschrijving van de taalvaardigheid van een gebruiker op zes verschillende niveaus, variërend van basisgebruiker tot uiterst vaardige gebruiker.

Met behulp van beschrijvingen van specifieke situaties (*can do statements*) kan een taalgebruiker zelf zijn niveau inschatten. Hierbij wordt niet alleen een onderscheid gemaakt

naar de vier taalvaardigheden, maar ook bin-
nen de taalvaardigheden wordt een onder-
scheid gemaakt naar meer specifieke situaties
waarin taal gebruikt wordt. Zo wordt voor
schrijven een onderscheid gemaakt tussen
algemene schrijfvaardigheid, creatief schrij-
ven en het schrijven van rapporten en essays.

Het CEF is niet zo maar een theoretische
exercitie, ook in ons onderwijs heeft het ingang
gevonden. Niet alleen worden de examens
gerelateerd aan de niveauomschrijvingen van
het CEF (zie: Noijons & Van Hest, 2006), maar
ook wordt het CEF gebruikt als anker voor
niveauomschrijvingen van het Nederlands
(zie bijvoorbeeld het 'Raamwerk Nederlands'
voor het (v)mbo (Bohenn, Jansen, Kuijpers,
Thijssen, Schot & Stockmann, 2007) en 'Over
de drempels met taal' van de Expertgroep
Doorlopende Leerlijnen (2008a, 2008b). In
de beide laatste rapporten wordt met het CEF
in de hand een omschrijving gegeven van
het niveau dat een leerling op een bepaald
moment in het onderwijs behaald zou moe-
ten hebben.

Voor schrijfvaardigheid kan het CEF van
grote betekenis zijn. Er is dan immers een
algemeen geldende omschrijving van de ken-
merken waaraan een (bepaald type) tekst van
een bepaald niveau moet voldoen. Een der-
gelijke omschrijving heeft een grote invloed
op de kwaliteit van oordelen over teksten.
Bijvoorbeeld Slagter (2006) vroeg in een
reeks van experimenten moedertaalsprekers
Spaans en niet-moedertaalsprekers Spaans
een oordeel te geven over schrijfproducten
van studenten uit het WO. Met de omschrij-
vingen uit het CEF als referentie blijken de
verschillen tussen beoordelaars vele malen
kleiner dan wanneer de beoordelaars hier niet
de beschikking over hadden. Dit geldt zowel
voor beoordelaars met Spaans als moedertaal
als voor beoordelaars die Spaans niet als
moedertaal hebben. Een conclusie die overigens
ook blijkt uit allerlei ander onderzoek
naar de beoordelingsproblematiek bij schrijf-

vaardigheid; het geven van een houvast aan
beoordelaars in termen van een omschrijving
van kenmerken van teksten leidt tot een gro-
tere overeenstemming in oordelen (zie onder
andere Wesdorp, 1974; Meuffels, 1994).

De vraag is echter of docenten hetzelfde
oordeel geven. Overeenstemming in oordelen
is wat dat betreft iets heel anders dan het zelfde
oordeel. Als verschillende beoordelaars tot
eenzelfde rangorde komen, dan is hun over-
eenstemming hoog, maar zegt dat niet dat zij
het product in kwestie ook gelijk waarderen.
De ene beoordelaar heeft bijvoorbeeld een A1
aan de ene tekst en A2 aan de andere tekst
toegekend, terwijl een andere beoordelaar C1
en C2 aan dezelfde teksten toegekend heeft.
De overeenstemming is dan hoog (beiden
hebben de eerste tekst lager beoordeeld dan
de tweede tekst), maar er is wel een duidelijk
verschil in oordelen over de kwaliteit van
de betreffende producten. Nu is één van de
doelstellingen van het CEF dat we (eindelijk)
een gemeenschappelijke taal hebben om over
de kwaliteit van teksten te praten, en dat dit
leidt tot gelijke oordelen. Dus: de vraag is
of in verschillende landen eenzelfde product
eenzelfde oordeel krijgt.

Onderzoeksoptzet

Om de bruikbaarheid van oordelen op basis
van het CEF te onderzoeken, hebben derde- en
vijfdeklassers (N = 129) van het vwo één of
twee schrijfoopdrachten in het Duits dan wel
Engels gemaakt. Afhankelijk van de opdracht
schreven zij een brief en/of een betoog in het
Duits dan wel het Engels.

Voor de brief werden de leerlingen gecon-
fronteerd met een actie van een fabrikant van
yummy-yummy smikkelrepen. Door tien actie-
wikkels op te sturen naar de fabrikant kunnen
zij, volgens de opdracht, een i-pod winnen.
Helaas, hoewel de actie nog niet afgelopen
is en de leerling maar acht actiewikkels heeft

verzameld, zijn smikkelrepen met actiewik-
kels in geen enkele winkel meer te vinden.
Daarom schrijft de leerling een brief aan de
fabrikant waarin duidelijk gemaakt wordt
dat er weliswaar tien wikkels meegestuurd
zijn, maar dat het slechts in acht gevallen
een actiewikkel betreft. De leerling heeft dus
wel tien repen gegeten en wil toch graag
meedoen aan de wedstrijd, want een i-pod is
nooit weg.

Voor de tweede opdracht moest de leer-
ling een betoog schrijven over de stelling 'het
spelen van computerspelletjes is verslavend en maakt
mensen eenzaam'. In de opdracht is benadrukt
dat argumenten vóór of tegen deze stelling
gegeven moesten worden. De argumenten
moesten met voorbeelden ondersteund wor-
den en uitgewerkt tot een conclusie over dit
thema.

Deze teksten zijn door de onderzoekers
geclassificeerd als gemiddeld, beneden
gemiddeld en beter dan gemiddeld. Van de
Duitse schrijfproducten zijn per opdracht per
klas tien teksten geselecteerd (waarover het
oordeel van de onderzoekers unaniem was);
twee gemiddelde teksten, twee bovengemid-
delde, twee heel goede teksten, twee beneden
gemiddelde en twee slechte teksten. In
totaal zijn er dus twintig Duitse teksten van
derde- en twintig Duitse teksten van vijfde-
klassers geselecteerd. Voor het Engels zijn

per opdracht per klas vijf teksten geselec-
teerd (waarover het onafhankelijke oordeel
van de onderzoekers wederom unaniem was):
één gemiddelde, één boven gemiddelde, één
heel goede, één beneden gemiddelde en één
slechte tekst. In totaal zijn er twintig Engelse
schrijfproducten geselecteerd (twee opdrach-
ten x twee klassen x vijf teksten).

De Duitstalige schrijfproducten zijn vervol-
gens beoordeeld door Nederlandse docenten
Duits (N=10) en Duitse docenten Duits in (en
uit) Duitsland (N=10). De Engelstalige schrijf-
producten zijn beoordeeld door Nederlandse
docenten Engels (N=20) en Engelse docen-
ten Engels in (en uit) Engeland (N=10). De
docenten hebben een oordeel gegeven vol-
gens een onderzoeksontwerp zoals weerge-
geven in Tabel 1.

Zoals uit de tabel blijkt, geeft een docent
cijfers voor brieven en oordelen op basis
van het CEF over de betogen, dan wel cijfers
voor de betogende teksten en op het CEF-
gebaseerde oordelen over de brieven. Op deze
manier beoordeelt een docent alle producten
slechts één keer. De helft van de docenten
heeft als eerste cijfers gegeven, en daarna de
schrijfproducten beoordeeld met de richtlij-
nen uit het CEF, en voor de andere helft was
de volgorde precies omgekeerd. Hierdoor
zijn contaminatie-effecten, indien aanwezig,
gelijkelijk verdeeld over de beoordelingscon-

| DOCENT | LAND | CIJFERS | | CEF | |
|--------|--------------------|---------|--------|-------|--------|
| | | BRIEF | BETOOG | BRIEF | BETOOG |
| 1 | Nederland | X | | | X |
| 2 | Duitsland/Engeland | X | | | X |
| 3 | Nederland | | X | X | |
| 4 | Duitsland/Engeland | | X | X | |
| ... | | | | | |

Tabel 1: Onderzoeksoptzet voor het beoordelen van schrijfproducten op basis van een gebruikelijke cijferschaal en op basis van het CEF

dities, zonder dat ze invloed hebben op de (overall) resultaten.

Bij het verzamelen van de oordelen is de Nederlandse docenten Duits en Engels gevraagd de producten te beoordelen als waren het de producten van hun leerlingen uit de vierde klas van het vwo. De Duitse en Engelse docenten is gevraagd om de producten te beoordelen als waren het de producten van leerlingen die net uit Nederland bij hen in de (vierde) klas kwamen.

Voor de beoordeling op basis van het CEF hadden de docenten de beschikking over een omschrijving van de zes niveaus (zie www.europeestaalportfolio.nl). Vanzelfsprekend is dit beoordelingsschema met een toelichting beschikbaar gesteld in de moedertaal van de beoordelaar.

CIJFERSCHAAL

De cijferschaal die in de drie landen gebruikelijk is, verschilt van land tot land. Duitse docenten zijn gewend cijfers te geven op een schaal van 1 tot 6, waarbij het cijfer 1 staat voor een excellente prestatie. Engelse docenten gebruiken een letterschaal, waarbij de letter -a- staat voor een excellente prestatie, en de letter -e- voor een zware onvoldoende. In Nederland is een tienpuntsschaal gebrui-

kelijk, waarbij een 10 een excellente prestatie en een 1 een minimale prestatie aangeeft. Voor een vergelijking van de oordelen van de docenten uit de verschillende landen, én een vergelijking met de oordelen uit het CEF zijn alle oordelen getransformeerd naar een zespuntsschaal, waarbij een 1 staat voor een minimale en een 6 voor een excellente prestatie. Dit maakt een vergelijking van de oordelen in de verschillende landen gemakkelijker².

Resultaten

De presentatie van de resultaten valt uiteen in vier delen. Eerst gaan we in op de betrouwbaarheid van de beoordeling. Vervolgens wordt nagegaan of de gegevens voldoen aan een minimale eis van validiteit: de 'known group validity'. Daarna komen verschillen tussen gemiddelde cijfers en CEF-oordelen aan de orde, alsmede correlaties tussen de beide typen oordelen. Tot slot wordt kort ingegaan op de verschillen tussen beide opdrachten.

BETROUWBAARHEID

Als eerste is nagegaan in hoeverre de oordelen van de verschillende groepen docenten per beoordelingsmethode overeenstemmen.

| | KLAS | DUITSE SCHRIJFPRODUCTEN | | ENGELSE SCHRIJFPRODUCTEN | |
|---------|------|-------------------------|-----------|--------------------------|-----------|
| | | DOCENT | | DOCENT | |
| | | Nederlands | Duits | Nederlands | Engels |
| Cijfers | 3 | 2.9 (0.8) | 3.1 (0.6) | 3.0 (0.8) | 3.5 (0.8) |
| | 5 | 3.4 (0.7) | 3.6 (0.9) | 3.9 (0.8) | 4.2 (1.0) |
| CEF | 3 | 2.4 (0.7) | 3.0 (0.7) | 2.4 (0.7) | 3.3 (0.6) |
| | 5 | 3.0 (0.8) | 3.9 (0.8) | 3.3 (0.9) | 3.9 (1.0) |

Noot. Een 1 weerspiegelt een minimaal en een 6 een excellent product.

Tabel 2: Gemiddelde oordelen en standaarddeviaties (tussen haakjes) per klas voor de oordelen op de gebruikelijke cijferschaal en de oordelen met behulp van het CEF.

Zoals verwacht, bestaat tussen docenten een grote mate van overeenstemming over de kwaliteit van de schrijfproducten. Voor de oordelen van de Duitse teksten varieert de betrouwbaarheid van 0.87 tot 0.88. Voor de oordelen over de Engelse teksten is de betrouwbaarheid van de oordelen zelfs nog hoger ($0.91 < \alpha < 0.92$). Dit is natuurlijk niet zo vreemd als we bedenken dat het hier om een selectie van duidelijk in kwaliteit verschillende schrijfproducten gaat.

KNOWN GROUP VALIDITY

Een meetinstrument is adequaat als het bestaande verschillen kan blootleggen. Verwacht mag worden dat de schrijfproducten van vijfdeklassers (gemiddeld) van hogere kwaliteit zijn dan die van derdeklassers. Dit is de zogenaamde 'known group' validiteit. In Tabel 2 staan de gemiddelde oordelen per beoordelingsmethode, uitgesplitst naar herkomst van de docent.

Voor schrijfvaardigheid Duits varieert het gemiddelde cijfer van 2.9-3.1 voor derdeklassers tot 3.4-3.6 voor vijfdeklassers (voor respectievelijk Nederlandse en Duitse docenten Duits), op een schaal van 1 (voor een minimale prestatie) tot 6 (voor een excellente prestatie). Dit verschil is significant ($F(1, 37) = 2.18; p = 0.03$). Op de cijferschaal hebben de schrijfproducten van vijfdeklassers een positiever oordeel ontvangen dan die van de derdeklassers.

Bij beoordeling van dezelfde schrijfproducten volgens normen van het CEF komen Nederlandse docenten Duits tot een gemiddeld oordeel tussen A2 en B1 in (2.4) voor derdeklassers, en voor vijfdeklassers komen zij tot B1 (3.0) als gemiddeld niveau. Duitse docenten Duits zijn van mening dat het gemiddelde schrijfproduct van derdeklassers al niveau B1 (3.0) bereikt heeft, en dat vijfdeklassers gemiddeld (bijna) tot B2 (3.9) komen. Ook op grond van deze verschillen kan geconcludeerd worden dat de schrijf-

producten van vijfdeklassers beter zijn dan die van derdeklassers ($F(1, 37) = 10.89; p < .002$). Bij beide beoordelingsmethoden kan voor Duits het verwachte verschil tussen de prestaties van derde- en vijfdeklassers aangetoond worden.

Voor het schrijven van Engelse teksten is het gemiddelde oordeel van Nederlandse docenten Engels 3.0 voor derdeklassers en 3.9 voor vijfdeklassers. Engelse docenten Engels hebben voor beide groepen een iets hoger oordeel (3.5 voor derdeklassers en 4.2 voor vijfdeklassers). Gemiddeld over beide groepen docenten is het verschil in (gemiddelde) oordelen van derde- en vijfdeklassers duidelijk significant ($F(1, 18) = 5.09; p < 0.04$).

Volgens de CEF-oordelen zit de schrijfvaardigheid van derdeklassers tussen A2 en B1 (2.4) in, terwijl vijfdeklassers ruim boven B1-niveau scoren (3.3). Engelse docenten Engels hebben gemiddeld een positiever oordeel dan hun Nederlandse collegae; de derdeklasser komt gemiddeld al boven B1 niveau (3.3) uit, en de vijfdeklasser heeft gemiddeld (bijna) niveau B2 (3.9) bereikt. Het verschil tussen de gemiddelde prestaties van derde- en vijfdeklassers is wederom significant ($F(1, 18) = 4.50; p = 0.04$). Ook op basis van de oordelen volgens de omschrijving in het CEF kan een onderscheid gemaakt worden tussen de (gemiddelde) prestatie van derde- en vijfdeklassers.

VERSCHILLEN IN GEMIDDELDE OORDELEN OP DE CIJFERSCHAAL EN DE CEF-SCHAAL

Voor de oordelen over de Duitse schrijfproducten blijkt dat de oordelen van Duitse docenten Duits gemiddeld hoger zijn dan die van Nederlandse docenten Duits ($F(1, 37) = 14.21; p < 0.01$). Maar bij het CEF zijn de verschillen tussen Nederlandse docenten en Duitse docenten groter dan bij de traditionele wijze van beoordelen ($F(1, 37) = 7.19; p = .01$). Dit geldt voor zowel derde- als vijfdeklassers ($F(1, 37) = 0.99; p = 0.33$). Duitse docenten

Duits zijn, bij zowel de cijferschaal als bij de oordelen op basis van het CEF, coulanter dan hun Nederlandse collegae (zie ook Figuur 1). Opvallend is dat vijfdeklassers volgens Nederlandse docenten Duits gelijk zijn aan derdeklassers volgens Duitse docenten Duits. In beide gevallen wordt het oordeel B1 (Threshold) gegeven.

Voor het Engels blijkt er op zowel de cijferschaal als met behulp van het CEF een significant verschil tussen de Nederlandse en Engelse docenten Engels ($F(1, 18) = 20.9$; $p < 0.01$, respectievelijk $F(1, 18) = 108.5$; $p < 0.01$). Op de cijferschaal is het verschil tussen de oordelen van Nederlandse docenten Engels en Engelse docenten Engels kleiner dan op de CEF-schaal ($F(1, 18) = 5.79$; $p = 0.03$, maar het verschil tussen cijfers en CEF verschilt niet tussen derde- en vijfdeklassers ($F(1, 18) = 0.82$; $p = .78$). Dit is in Figuur 1 grafisch weergegeven. Volgens Engelse docenten Engels kunnen derdeklassers al wat zij volgens Nederlandse docenten Engels pas in de vijfde realiseren.

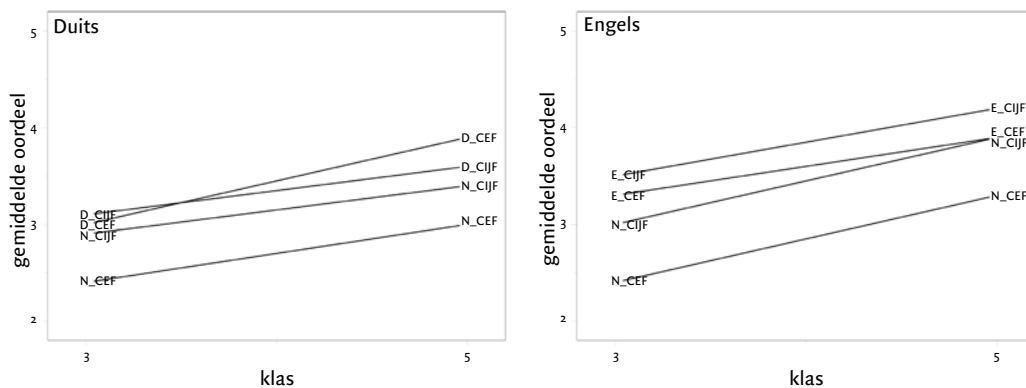
CORRELATIES TUSSEN OORDELEN

Omdat alle docenten Duits en alle docenten Engels dezelfde (twintig respectievelijk tien) schrijfproducten beoordeeld hebben, kan de

correlatie tussen de (gemiddelde) oordelen per beoordelingsmethode berekend worden. Deze staan in Tabel 3.

Voor Nederlandse docenten is de correlatie tussen cijfers en CEF-oordelen 0.6, en voor Duitse docenten is deze correlatie 0.7. Ook gecorrigeerd voor de (on)betrouwbaarheid van de juryoordelen wijken de correlaties duidelijk af van 1.0. Met behulp van het CEF wordt door Duitse en Nederlandse docenten Duits dus niet exact hetzelfde beoordeeld als op de gewone cijferschaal, want dan zou de (gecorrigeerde) correlatie niet afwijken van 1.0.

De correlatie tussen de oordelen van Nederlandse en Duitse docenten Duits op de cijferschaal is 0.4 (gecorrigeerd voor onbetrouwbaarheid 0.5). We moeten dus de conclusie trekken dat Nederlandse docenten Duits en Duitse docenten Duits op andere kenmerken van een schrijfproduct letten wanneer zij deze op de voor hen gebruikelijke wijze beoordelen; de ene groep docenten vindt andere kenmerken van de teksten relevant dan de andere groep docenten. De correlatie tussen de CEF-oordelen van beide groepen docenten Duits is veel hoger (0.8). Wanneer beide groepen met het behulp van de omschrijvingen uit het CEF een oordeel geven, dan spelen meer dezelfde kenmerken een rol.



Figuur 1: Gemiddelde oordelen voor Nederlandse (NL) docenten Duits en Duitse docenten Duits (links) en Nederlandse docenten Engels en Engelse docenten Engels (rechts) op de cijferschaal (C) als met het CEF.

| Taal tekst | Herkomst docent | CORRELATIES TUSSEN: | | | | | |
|------------|-----------------|---------------------|-------|-----------------|-------|-----------|-------|
| | | CIJFERS-CEF | | CIJFERS-CIJFERS | | CEF - CEF | |
| Duits | Nederland | 0.6 | (0.7) | 0.4 | (0.5) | 0.8 | (0.9) |
| | Duitsland | 0.7 | (0.8) | | | | |
| Engels | Nederland | 0.7 | (0.8) | 0.5 | (0.6) | 0.8 | (0.9) |
| | Engeland | 0.6 | (0.7) | | | | |

Tabel 3: Correlaties tussen verschillende oordelen (tussen haakjes de voor jury-onbetrouwbaarheid gecorrigeerde betrouwbaarheid van de oordelen).

Voor het Engels geldt hetzelfde als voor het Duits. Ook daar is de samenhang tussen de oordelen op de cijferschaal en de oordelen op basis van het CEF kleiner dan 1.0. Dus: ook bij het beoordelen van Engelse teksten wordt op de traditionele manier net iets anders beoordeeld dan wanneer het CEF een rol speelt bij de beoordeling van de teksten. Ook de correlaties tussen de cijfers (0.5) van Engelse en Nederlandse docenten Engels is veel lager dan de correlatie tussen de CEF-oordelen (0.8) van beide groepen. Ook voor het beoordelen van Engelse teksten door Nederlandse en Engelse docenten Engels blijkt de duidelijk uniformerende werking van het CEF.

VERSCHILLEN TUSSEN OPDRACHTEN

Tot dusver is geen verschil gemaakt tussen de beide opdrachten: brief en betoog. Er blijkt echter een duidelijk verschil in beoordeling tussen de teksten van beide opdrachten. Voor Duits is er voor de brief een groter verschil tussen de traditionele beoordelingsmethode en de CEF-oordelen dan bij het betoog ($F(1, 37) = 14.32$; $p = 0.01$). Hetzelfde geldt bij de beoordeling van Engelse teksten ($F(1, 18) = 4.51$; $p = 0.48$); ook bij het Engels zijn de verschillen tussen de beide beoordelingsmethoden groter bij de brief dan bij het betoog. Kennelijk is het verschil tussen

de voor de docent gebruikelijke cijferschaal en het CEF afhankelijk van kenmerken van de specifieke opdracht. Wellicht leent het betoog, met zijn argumentatieve structuur, zich beter voor een beoordeling met behulp van het CEF vanwege het argumentatieve karakter van de opdracht.

Conclusies en discussie

De resultaten voor de beoordeling van Duitse teksten door Nederlandse en Duitse docenten Duits, én de beoordeling van Engelse teksten door Nederlandse en Engelse docenten Engels vertonen een opmerkelijke convergentie. We hebben gedemonstreerd dat met de gebruikelijke cijferschalen zowel voor de oordelen over Duitse teksten als de oordelen over Engelse teksten een onderscheid gemaakt kan worden tussen de kwaliteit van de teksten van derde- en vijfdeklassers van het vwo. Ook met oordelen die gebaseerd zijn op omschrijvingen uit het CEF kan voor beide talen een onderscheid gemaakt worden tussen de prestaties van derde- en vijfdeklassers.

De resultaten laten ook zien dat de verschillen tussen oordelen over de schrijfproducten afhankelijk zijn van de moedertaal van de beoordelaar. Duitse docenten Duits zijn

meer coulant in hun oordelen over de Duitse teksten dan Nederlandse docenten Duits. En hetzelfde geldt voor het Engels: Engelse docenten Engels geven hogere oordelen dan Nederlandse docenten Engels. Het is niet zo vreemd dat dit geldt voor oordelen op een voor de docent gebruikelijke cijferschaal, maar wel voor de oordelen op basis van de omschrijvingen uit het CEF. Immers, daar is een vaste omschrijving van kenmerken waar een schrijfproduct van een bepaald niveau aan moet voldoen. Een dergelijke omschrijving heeft echter niet voldoende effect om tot een uniform oordeel over schrijfproducten te komen. Het is dan ook opmerkelijk dat de verschillen tussen de oordelen op de cijferschaal kleiner zijn dan die op de schaal volgens het CEF.

Ook blijkt de correlatie tussen de oordelen van verschillende groepen docenten hoger wanneer zij gebruik gemaakt hebben van het CEF, dan wanneer zij op de voor hen vertrouwde cijferschaal geoordeeld hebben. Dat betekent dat zij door het CEF 'gestuurd' worden in hun oordeel; Duitse docenten Duits en Nederlandse docenten Duits letten meer op dezelfde productkenmerken wanneer zij hun oordelen baseren op de omschrijvingen uit het CEF, dan wanneer zij op de voor hen gebruikelijke wijze een oordeel geven. Eenzelfde patroon geldt voor de oordelen over de Engelse teksten, ook daar kan niet aangevoeld worden dat Engelse en Nederlandse docenten Engels andere kenmerken van de schrijfproducten relevant vinden wanneer zij gebruik maken van het CEF om tot een oordeel te komen.

Een waarschuwing is hier echter wel op zijn plaats. Immers, in dit onderzoek hebben wij (relatief) extreme producten uitgezocht (zowel relatief erg goede als relatief erg slechte teksten). Dit is natuurlijk van invloed op de hoogte van de samenhang; in feite wordt deze namelijk enigszins overschat omdat een lage samenhang niet veroorzaakt

wordt door een gebrek aan overeenstemming tussen producten die zich aan het einde van de schaal bevinden, maar door gebrek aan overeenstemming tussen producten op het midden van de schaal.

Tot slot blijkt ook de opdracht van invloed te zijn op de verschillen tussen cijfers en CEF-oordelen. Opmerkelijk genoeg blijken oordelen op de cijferschaal minder opdrachtafhankelijk dan CEF-oordelen. Van globale oordelen is bekend dat zij verschillen per opdracht (vergelijk, Godshalk, Swineford & Coffman, 1966; Schoonen, 1991). Het CEF pretendeert juist opdrachtafhankelijk te zijn, maar lijkt daar vooralsnog niet volledig in te slagen.

Concluderend: de verschillen tussen de oordelen van docenten uit verschillende landen worden niet kleiner bij gebruik van de omschrijvingen van het CEF. Hiermee kan het CEF vooralsnog niet voldoen aan de doelstelling van 'één gemeenschappelijke taal om over de kwaliteit van schrijfproducten te spreken'. Opmerkelijk genoeg betreft het hier geen random variatie, maar zijn de moedertaal docenten (Duitse docenten Duits en Engelse docenten Engels) coulanter in hun oordeel dan hun Nederlandse collegae. Dit lijkt niet alleen te gelden voor oordelen over schrijfvaardigheid, maar ook voor oordelen over spreekvaardigheid (Koet, 2007). Nader onderzoek zal nog moeten uitwijzen waardoor deze systematische verschillen veroorzaakt worden. De verschillen in oordelen op basis van het CEF lijken zowel taal- als opdrachtafhankelijk. Het tweede is lastig, maar geen principieel punt; leerlingen kunnen met de ene opdracht wellicht beter uit de voeten dan met een andere opdracht, maar dit beperkt wellicht de reikwijdte van de omschrijvingen uit het CEF. Het eerste is meer principieel. Het CEF is immers ontwikkeld om mensen meer houvast te geven. Als oordelen echter afhankelijk zijn van de taal van de beoordelaar, dan schiet het CEF zijn doel (deels) voorbij.

NOTEN

1. Met dank aan Elske Mol en Petra Ubels.
2. Let wel, geheel onproblematisch is deze aanname niet. Immers, er wordt vanuit gegaan dat de oordelen gerepresenteerd zijn op interval niveau. Echter, het is voorspands niet zeker dat de afstand tussen de schaalpunten altijd gelijk is. Wel moet hierbij bedacht worden dat bij vijf schaalpunten of meer de fout die hierdoor wellicht gemaakt wordt in het algemeen verwaarloosbaar is (Jöreskog & Sörbom, 1989). Overigens is een vergelijking zonder deze transformatie ook mogelijk, maar die is dan veel ingewikkelder, omdat het dan alleen over interactie-effecten gaat (opgemerkt zij dat zo'n analyse tot exact dezelfde conclusies leidt).

LITERATUUR

- Bohnenn, E., Jansen, F., Kuijpers, C., Thijssen, R., Schot, I., & Stockmann, W. (2007). *Het raamwerk Nederlands. Nederlands in (v)mbo-opleiding, beroep en maatschappij*. Cinop: 's-Hertogenbosch.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Language Policy Division.
- Council of Europe (2003). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEF)*. Strasbourg: Language Policy Division.
- Council of Europe (2004). *Reference Supplement to the Preliminary Pilot version of the Manual for Relating Language examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. December 2004, DGIV/EDU/LANG (2004) 13. Strasbourg: Council of Europe: Language Policy Division.
- Expertgroep 'Doorlopende leerlijnen' (2008a). *Over de drempels met taal en rekenen*. SLO: Enschede.
- Expertgroep 'Doorlopende leerlijnen'

(2008b). *Over de drempels met taal*. SLO: Enschede.

Jöreskog, K.G., & Sörbom, D. (1989). *Lisrel 7 User's reference guide*. Mooresville: Scientific software, Inc.

Godshalk, F.I., Swineford, F., & Coffman, W.E. (1966). *The measurement of writing ability*. New York: College Entrance Examination Board.

Koet, T. (2007). *Polder English in Dutch ears*. ILO: UvA, Amsterdam

Meuffels, H.L.M. (1994). *De verguisde beoordelaar. Opstellen over opstelbeoordeling*. Thesis: Amsterdam.

Noijons, J., & Hest, E. van (2006). Centrale examens mvt gekoppeld aan het ERK. *Levende Talen Magazine*, 7, 5-42.

Slagter, P.J. (2006). *How do I know that my Level B1 is your Intermediate Mid? The Common European Framework of Reference for Languages, the ACTFL Guidelines for Writing, and Assessment of Student Writing*. Ongepubliceerd manuscript.

Schoonen, R. (1991). *De evaluatie van schrijfvaardigheidsmetingen: Een empirische studie naar betrouwbaarheid, validiteit en bruikbaarheid van schrijfvaardigheidsmetingen in de achtste groep van het basisonderwijs*. Amsterdam: Universiteit van Amsterdam.

Wesdorp, H. (1974). *Het meten van productief-schriftelijke taalvaardigheid. Directe en indirecte methoden: 'opstelbeoordeling' versus schrijfvaardigheidstoetsing*. Muusses: Purmerend.

HUUB VAN DEN BERGH is verbonden aan de Universiteit Utrecht en de Universiteit van Amsterdam. Correspondentieadres: Universiteit Utrecht. Uil-OTS, Trans 10, 3512 JK Utrecht of Universiteit van Amsterdam. ILO, Spinozastraat 55, 1018 HJ Amsterdam. E-mail: <Huub.vandenBergh@let.uu.nl>.

LISANNE KLEIN GUNNEWIEK is verbonden aan de Hogeschool Arnhem en Nijmegen. Correspondentieadres: HAN, Onderwijs en Opleiden. Postbus 30011, 6503 HN Nijmegen.