

Pullens, T. J. M. (2012). *Bij wijze van schrijven* (dissertatie). Universiteit Utrecht.
 Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modeling. *Language Testing*, 22(1), 1-30.
 Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs*. Den Haag: Stichting voor Onderzoek van het Onderwijs.

RENSKE BOUWER (1984) studeerde aan de Radboud Universiteit, waar zij in 2008 de research master Behavioural Science afrondde. In 2012 is zij gestart met een promotieproject (aan de Universiteit Utrecht) naar verbetering van de schrijfvaardigheid van leerlingen in de bovenbouw van het basisonderwijs. Hierbij concentreert zij zich onder meer op de vraag hoe schrijfvaardigheid van leerlingen valide en betrouwbaar geëvalueerd kan worden, en hoe aan leerlingen adequate feedback op hun teksten gegeven kan worden.
 E-mail: <i.r.bouwer@uu>

HUUB VAN DEN BERGH is hoogleraar toetsing en didactiek van taalvaardigheid aan de Universiteit Utrecht. Hij was betrokken bij vele groot- en kleinschalige projecten naar de kwaliteit van ons (taal)onderwijs.
 E-mail: <h.vandenbergh@uu.nl>

Beoordeling van Engelse taalvaardigheid door docenten en Cambridge-examinatoren

TON KOET & HUUB VAN DEN BERGH

Kunnen docenten en opleiders met 'oude' Cambridge-examens tot oordelen over de taalvaardigheid van hun leerlingen en studenten komen die vergelijkbaar zijn met de oordelen van Cambridge-examinatoren? Gezien de nadelen van de officiële examens is het antwoord op deze vraag voor scholen en opleidingen interessant. Een pilot-onderzoek gaf echter geen aanleiding te veronderstellen dat het antwoord bevestigend zou zijn.

In de kennisbases voor tweedegraadsopleidingen voor leraar moderne vreemde talen (Duits, Engels, Frans en Spaans) zijn aan het Europees Referentiekader (ERK) gerelateerde eisen aan de taalvaardigheid geformuleerd (HBO-raad, 2009). De vier vreemde talen verschillen in de vereiste niveaus, waarbij alleen Frans een onderscheid maakt tussen de vaardigheden (zie tabel 1). Voor Engels staat daarnaast vermeld dat een voldoende voor het Certificate of Proficiency in English (CPE) is vereist.

TAAL	VAARDIGHEID	ERK-NIVEAU
Duits		C1
Engels		C; niet nader gespecificeerd
Frans	Leesvaardigheid	C1
	Kijk- en luistervaardigheid	B2 – C1
	Schrijfvaardigheid	niet gespecificeerd
	Spreekvaardigheid	B2 – C1
	Gespreksvaardigheid	B2 – C1
Spaans		hoog B2

Tabel 1. De vereiste taalvaardigheidsniveaus voor de vier moderne vreemde talen

De examens voor het CPE worden aangeboden door Cambridge ESOL (English for Speakers of Other Languages; de naam is gewijzigd in Cambridge English Language Assessment). Aan het CPE kleven een aantal financiële, organisatorische en inhoudelijke bezwaren. Het examen is relatief duur (in 2013 was het examengeld E 234 per student). De tijd tussen het afleggen van de beide delen van het examen en de uitslag is relatief lang (als het mondeling deel van het examen in mei en het schriftelijk deel in het midden van juni is afgelegd, zal de uitslag meestal niet eerder dan in het midden van juli beschikbaar zijn). De einduitslag wordt gegeven als score (een getal tussen 1 en 100) en in de letters A, B, C, die een voldoende score aangeven en D en E die een onvoldoende representeren; de score op de verschillende onderdelen wordt echter slechts bij benadering gegeven, wat bezwaarlijk is voor opleidingen die eisen dat voor bepaalde onderdelen (meestal de spreekvaardigheid) een voldoende wordt behaald.

In 2009 hebben enkele tweedegraadsopleidingen Engels besloten te onderzoeken of het mogelijk is zelf een examen af te nemen dat vergelijkbaar is met het officiële CPE-examen maar waaraan de bovengenoemde bezwaren niet kleven. De prestaties van de studenten op zo'n gezamenlijk examen zouden beoordeeld worden door docenten van de eigen instelling én door docenten van een zusterinstelling. Deze beoordeling door zusterinstellingen zou gebeuren in een zogenaamde 'carrousel'. De negen deelnemende opleidingen zouden in een reeks worden gerangschikt, waarbij iedere opleiding zou samenwerken met twee opleidingen met vestigingen die niet al te ver van elkaar verwijderd waren; het werk van opleiding A zou dan ook worden beoordeeld door docenten van opleiding B, terwijl het werk van studenten van opleiding B op zijn beurt beoordeeld zou worden door docenten van opleiding C. Aldus zou een carrousel van samenwerkende opleidingen ontstaan.

De veronderstelling was dat er een samenhang zou zijn tussen de oordelen van de opleiders over de prestaties van de studenten op een 'oud' CPE-examen en de oordelen van de Cambridge-examinatoren over de prestaties van de studenten op het officiële CPE-examen. Voor wat betreft de oordelen over de prestaties op de onderdelen schrijfvaardigheid en spreekvaardigheid was er op voorhand enige twijfel. Deze twijfel is niet alleen toe te schrijven aan algemeen bekende beoordelingsproblematiek, waarbij gewezen wordt op verschillen in strengheid tussen beoordelaars, het verschijnsel dat verschillende beoordelaars andere kenmerken van een tekst beoordelen, eventuele bekendheid met de schrijver van de tekst, of instabiliteit van individuele beoordelaars (zie bv. McColly, 1970; Wesdorp, 1981; Huot, 1990; Bouwer, Beguin, Sanders, & Van den Bergh, 2015), maar wellicht ook aan de (Nederlandstalige) achtergrond van een aantal opleiders. (Het komt overigens voor dat opleiders – ook Nederlandstalige – bij de beoordeling van de onderdelen schrijfvaardigheid en spreekvaardigheid van het CPE-examen betrokken worden.)

Voor wat betreft de schrijfvaardigheid zouden deze problemen gering kunnen zijn. Weliswaar hebben bijvoorbeeld Van den Bergh en Klein Gunnewiek (2009) geconstateerd dat Engelse docenten coulanter zijn in hun oordeel over Engelstalige schrijfproducten van Nederlandse leerlingen dan Nederlandstalige docenten, maar Johnson en Lim (2009) hebben slechts minimale verschillen tussen Engelstalige en niet-Engelstalige beoordelaars van geschreven Engels aangetroffen.

Voor wat betreft de spreekvaardigheid was er op voorhand minder reden tot optimisme. Weliswaar heeft Kim (2009) geconstateerd dat Engelstalige en Koreaanstalige beoordelaars even streng zijn in hun oordeel over de mondelinge Engelse taalvaardigheid, maar Zhang en Elder (2011), die geen ver-

schil tussen de oordelen van Engelstalige en niet-Engelstalige beoordelaars over de mondelinge taalvaardigheid Engels van Chinese studenten hadden gevonden, gaven aan dat deze groepen beoordelaars verschilden in het gewicht dat zij hechtten aan aspecten van de mondelinge taalvaardigheid. Daarnaast vonden Carey, Mannel en Dunn (2011) belangrijke verschillen tussen beoordelaars die vertrouwd waren met het Engels van sprekers van het Chinees, Koreaans en Indisch, en beoordelaars die daar minder vertrouwd mee waren.

De vraag die in de pilot beantwoord moest worden is: kunnen docenten en opleiders met 'oude' Cambridge-examens tot oordelen over de taalvaardigheid van hun leerlingen en studenten komen die vergelijkbaar zijn met de oordelen van Cambridge-examinatoren?

Methode

Van de opleidingen Engels zegden er acht toe deel te zullen nemen aan deze pilot. (Een opleiding was nog niet betrokken bij het overleg over de examinering; een andere opleiding gaf aan dat zij nog onvoldoende ervaring met het CPE had.) Iedere opleiding zou vier studenten laten deelnemen aan het onderzoek, bij voorkeur twee vrouwelijke en twee mannelijke; in totaal zouden dus 32 studenten deelnemen. Deze studenten zouden dankzij

een subsidie van de HBO-raad gratis kunnen deelnemen aan het CPE-examen. Voor hun deelname aan de zogenaamde carrouseltoets zouden zij geen vergoeding ontvangen.

De studenten namen deel aan het examen voor het CPE dat door de British Council in mei en juni 2010 werd aangeboden; het examen werd op een aantal plaatsen aangeboden en de studenten kozen de plaats die voor hen het handigst was. Het mondelinge gedeelte vond plaats in mei en het schriftelijke gedeelte in juni; het schriftelijk gedeelte werd op alle plaatsen op dezelfde tijd afgelegd. De studenten namen ook deel aan het examen dat door de opleidingen was gekozen; zij deden dit op hun eigen hogeschool en op een tijd die voor hen handig was.

Certificate of Proficiency in English

Het CPE-examen bestond tot 2013 uit vijf onderdelen: Lezen, Schrijven, Use of English, Luisteren en Spreken (voor een uitgebreide beschrijving van het examen zie UCLES, 2008). De onderdelen Lezen en Luisteren werden digitaal nagekeken. Bij de onderdelen Schrijven en Use of English werden sommige stukken digitaal en andere stukken door Cambridge-examinatoren nagekeken. In tabel 2 vindt men de onderdelen van het 'oude' CPE-examen, met de afnameduur in minuten en de wijze van beoordelen.

ONDERDEEL	MINUTEN	BEOORDELING
Lezen	90	Digitaal
Schrijven	120	digitaal / Cambridge-examinatoren
Use of English	90	digitaal / Cambridge-examinatoren
Luisteren	40	Digitaal
Spreken	19	Cambridge-examinatoren

Tabel 2. De onderdelen van het CPE-examen tot 2013

ONDERDEEL	MINUTEN	SAMENWERKEN
Interview	3	Nee
Collaborative task	4	Ja
Long turn and discussion	12	Ja

Tabel 3. De onderdelen van het mondelinge examen

De beoordeling van de productieve mondelinge taalvaardigheid binnen het CPE-examen is gecompliceerd doordat er twee beoordelaars bij betrokken zijn; de beoordelaar die het gesprek voert (de interlocutor) geeft een globaal oordeel, terwijl de tweede beoordelaar (de examiner) oordelen geeft op vijf verschillende schalen: *grammatical resource*, *lexical resource*, *discourse management*, *pronunciation* en *interactive communication*. Bij het mondelinge examen worden twee kandidaten beoordeeld, die bij twee van de drie onderdelen moeten samenwerken. (Voor een kritische bespreking van het mondelinge examen, zie Macquoen & Harding, 2009.) Voor een overzicht van de drie onderdelen van het mondelinge examen, zie tabel 3.

Zoals eerder vermeld wordt de score op het CPE-examen zowel in letters (A, B, C, D of E) als in een cijfer tussen 1 en 100 gegeven. Sinds 2013 worden kandidaten die een A, B of C ontvangen geacht het niveau C2 van het Europees Referentiekader bereikt te hebben. Kandidaten die geen A, B of C

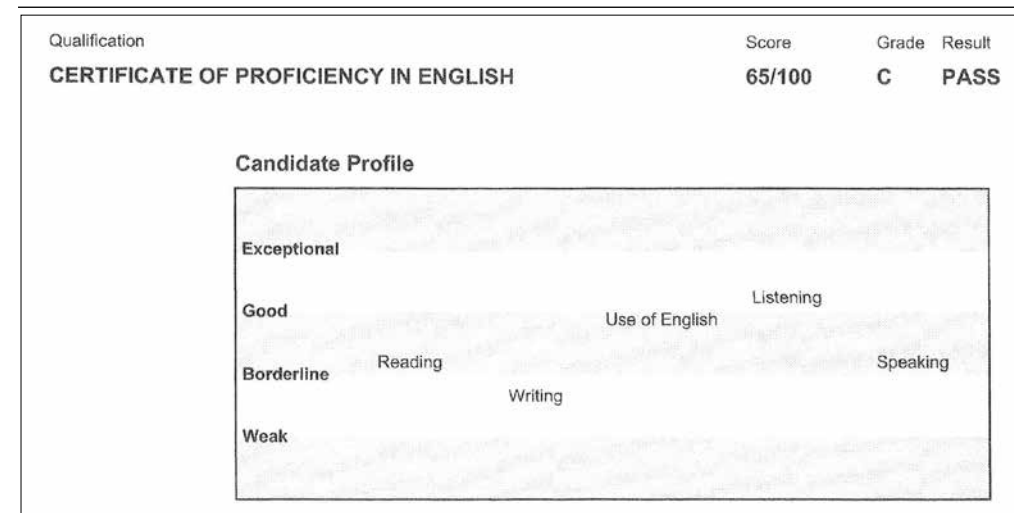
halen, kunnen geacht worden niveau C1 van het ERK bereikt te hebben, mits hun score niet te laag is. Van kandidaten die een onvoldoende behalen, wordt geen ERK-niveau aangegeven. De scores en hun betekenis worden weergegeven in tabel 4.

Uit bovenstaande tabel blijkt dat het continuüm van A tot C zeer ongelijk verdeeld is. Waar men een verdeling in drie ongeveer gelijke delen zou verwachten, is slechts een zeer klein deel toegewezen aan een oordeel B. Zoals boven aangegeven worden de scores op de individuele onderdelen niet bekendgemaakt. Deze scores zijn wel op te maken uit het profiel van de kandidaat dat bij de resultatenlijst grafisch wordt weergegeven. Voor een voorbeeld van een profiel van een kandidaat, zie figuur 1.

In de figuur vindt men vijf banden, waarvan de grenzen worden omschreven als 'uitzonderlijk, goed, net voldoende' en 'zwak'. De grenzen van de onderste en bovenste band hebben geen verdere omschrijving, maar aangenomen mag worden dat een

SCORE	RESULTAAT	BETEKENIS	ERK-NIVEAU
80 – 100	A	Zeer sterke prestatie	C 2
75 – 79	B	Goede prestatie	C 2
60 – 74	C	Krappe voldoende tot goede prestatie	C 2
45 – 59	D	Zwak tot net voldoende	C 1
0 – 44	E	Onvoldoende	Geen ERK-niveau aangegeven

Tabel 4. Scores en resultaten op het CPE-examen, hun betekenis en hun relatie tot de ERK-niveaus



Figuur 1. Voorbeeld van een profiel van een kandidaat

score in de onderste band een onvoldoende zou betekenen. De namen van de verschillende onderdelen van het CPE-examen worden in de grafiek geplaatst, zodat een beeld van het profiel van de kandidaat ontstaat. Uit de grafiek kan men opmaken dat Luisteren zich bij de kandidaat in het onderste gedeelte van de 'goed'-band bevindt en Use of English in het bovenste gedeelte van de 'net voldoende tot goed'-band, terwijl Spreken en Lezen in het onderste deel van de 'net voldoende tot goed'-band liggen en Schrijven in het bovenste deel van de 'zwak tot net voldoende'-band, waarmee deze vaardigheid niet meer op C2-niveau is maar op C1-niveau. (In theorie is het mogelijk dat een kandidaat een slechte prestatie voor een onderdeel compenseert door betere prestaties voor andere onderdelen.)

De pilot

Voor de pilot is gebruik gemaakt van een oud CPE-examen uit Mansfield en Nuttall (2007). (Oude examens worden voor Cambridge English in de serie CPE Practice

Tests gepubliceerd door Cambridge University Press). De onderdelen die in het CPE-examen digitaal nagekeken worden, zijn in de pilot door de eigen docent nagekeken. De onderdelen die in het CPE-examen door een Cambridge-examinator worden beoordeeld (zie tabel 2), zijn zowel door twee docenten van de eigen instelling als door twee docenten van een zusterinstelling beoordeeld, zodat over deze onderdelen telkens vier oordelen beschikbaar zijn. Van het mondelinge deel van het examen zijn video-opnamen gemaakt, die ter beoordeling naar de docenten van de zusterinstelling zijn gestuurd; ook over het mondelinge examen zouden telkens vier oordelen beschikbaar moeten zijn, twee van docenten van de eigen opleiding en twee van de docenten van de zusteropleiding.

Bij de uitvoering van de pilot deden zich veel problemen voor. Niet op alle opleidingen bleek het mogelijk vier studenten (twee mannelijke en twee vrouwelijke) te vinden die zowel aan het CPE-examen als aan het carousel-examen zouden deelnemen; op één opleiding kon geen enkele student gevonden worden. Sommige beoordelaars

konden geen zogenaamde bandscore geven, zoals gebruikelijk bij de CPE-examens; zij gaven cijfers op een schaal van 1 tot 10. In sommige gevallen waren de cijfers van de twee beoordelaars niet bekend, maar was slechts het eindresultaat gegeven. Door deze problemen moesten de gegevens van een aantal van de 32 deelnemers terzijde gelegd worden.

Uiteindelijk bleek dat van slechts 17 studenten die aan beide examens hadden deelgenomen de gegevens compleet waren. De studenten studeerden aan de lerarenopleidingen Engels van vier hogescholen; het betrof 4 studenten (2 mannelijke en 2 vrouwelijke) van de Fontys Lerarenopleiding Sittard (FLOS), 5 studenten (alle vrouwelijk) van de Hogeschool Arnhem Nijmegen (HAN), 3 studenten (2 vrouwelijke en 1 mannelijke) van de Hogeschool Rotterdam (HR) en 4 studenten (3 vrouwelijke en 1 mannelijke) van de Noordelijke Hogeschool Leeuwarden (NHL). De resultaten van deze studenten worden hieronder besproken.

Procedure

Toen zowel de oordelen van de carousel-beoordelaars als de resultaten van het CPE-examen bekend waren, werden deze oordelen gestandaardiseerd als scores op een schaal van 0 tot 5. Voor de scores op een schaal van

1 tot 10 was dit niet moeilijk, maar voor de band-scores was het niet eenvoudig te bepalen hoe dicht een vaardigheid bij de grens van de band lag. Van de scores van de verschillende carousel-beoordelaars werd het gemiddelde berekend.

Resultaten en discussie

De correlatie tussen de oordelen van de carousel-beoordelaars over elk onderdeel en die van de Cambridge-examinatoren over hetzelfde onderdeel werd berekend. De correlatiecoëfficiënten worden weergegeven in tabel 5.

Uit tabel 5 blijkt dat de correlatie tussen de oordelen van de carousel-beoordelaars en de oordelen van de Cambridge-examinatoren niet significant was wat betreft de onderdelen Lezen, Luisteren, Use of English en Spreken; de correlatie was zelfs negatief wat het onderdeel Use of English betreft. Het blijkt ook dat de correlatie tussen de oordelen van de carousel-beoordelaars en de oordelen van de Cambridge-examinatoren significant was, zij het zwak, wat het onderdeel Schrijven betreft. Dit resultaat is teleurstellend; het is ook bevreemdend. Men had verwacht dat er, in ieder geval voor wat betreft de onderdelen die objectief respectievelijk grotendeels objectief beoordeeld konden worden (Luisteren, Lezen resp. Use of English), sprake zou zijn van een sterke samenhang. Dat deze samen-

ONDERDEEL	CORRELATIE	SIGNIFICANTIE P <0,005
Lezen	0,37	
Luisteren	0,12	
Use of English	-0,28	
Schrijven	0,66	*
Spreken	0,16	

Tabel 5. Correlatie tussen de oordelen van de carousel-beoordelaars en de oordelen van de Cambridge-examinatoren

Opleiding	DOCENTEN			CAMBRIDGE		
	n	gemiddeld	Sd	n	gemiddeld	sd
FLOS	4	3,25	0,68	4	4,00	0,91
HAN	5	2,90	0,42	5	2,50	1,17
HR	3	2,67	0,52	3	2,67	0,29
NHL	5	3,20	0,27	5	3,80	0,97
Totaal	17	3,03	0,48	17	3,27	1,11

Tabel 6. Gemiddelde oordelen van docenten en Cambridge-examinatoren over Schrijven (FLOS = Fontys Lerarenopleiding Sittard; HAN = Hogeschool Arnhem Nijmegen; HR = Hogeschool Rotterdam; NHL = Noordelijke Hogeschool Leeuwarden)

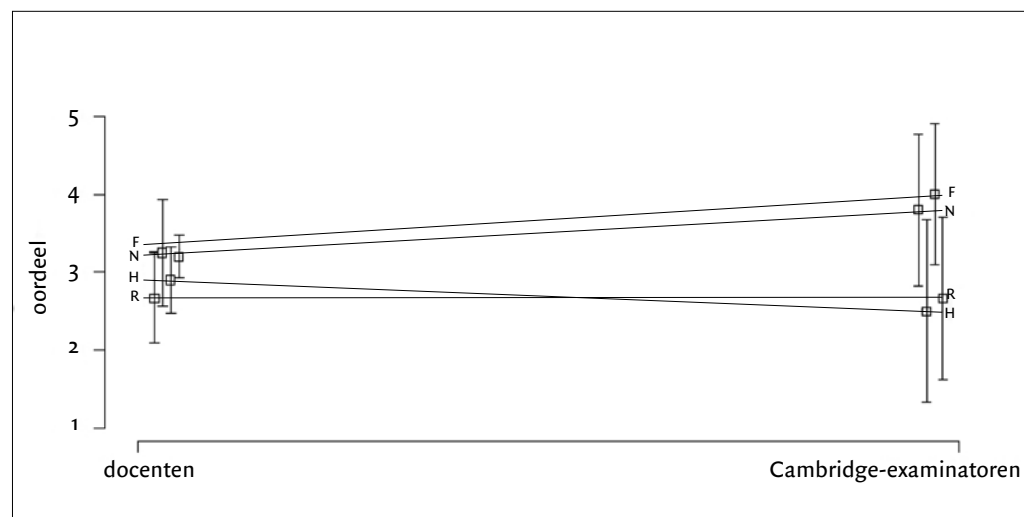
hang niet aangetoond kon worden, zou er op kunnen duiden dat de verschillen tussen de examens groot zijn. Dat er bij het onderdeel Spreken geen significante samenhang was, wekt in het licht van de bevindingen van Zhang en Elder (2011) en Carey, Mannel en Dunn (2011) geen verbazing. Dat er bij het onderdeel Schrijven wel een samenhang is geconstateerd, is evenmin verbazend in het licht van de bevindingen van Johnson en Lim (2009). De veronderstelling dat opleiders en Cambridge-examinatoren tot vergelijkbare oordelen kunnen komen over de schrijfvaardigheid lijkt ondersteund te worden. Om hier meer zekerheid over te verkrijgen zijn de gemiddelde oordelen van de docenten over Schrijven vergeleken met de oordelen van de Cambridge-examinatoren. De gemiddelde oordelen evenals als de standaarddeviaties zijn weergegeven in tabel 6.

Uit tabel 6 blijkt dat het gemiddelde oordeel van de carousel-beoordelaars over Schrijven minder positief is dan het gemiddelde oordeel van de Cambridge-examinatoren. Ook blijkt dat de gemiddelde oordelen van de carousel-beoordelaars per opleiding bij twee opleidingen minder positief zijn dan de gemiddelde oordelen van de Cambridge-examinatoren, terwijl deze bij één opleiding gelijk en bij één opleiding positiever zijn. Hoewel deze verschillen in geen van de gevallen significant

zijn, kan in figuur 2 toch een verschil tussen de oordelen van de carousel-beoordelaars en die van de Cambridge-examinatoren gesignaleerd worden.

In figuur 2 is duidelijk zichtbaar dat de gemiddelde oordelen van de carousel-beoordelaars per opleiding veel dichterbij elkaar liggen dan de gemiddelde oordelen van de Cambridge-examinatoren. Klaarblijkelijk is er bij de carousel-beoordelaars sprake van minder onderscheid tussen de prestaties van de verschillende instellingen dan de Cambridge-examinatoren. Ook is zichtbaar dat de oordelen van de carousel-beoordelaars bij twee opleidingen negatiever en de oordelen van de carousel-beoordelaars bij één opleiding positiever zijn dan die van de Cambridge-examinatoren, terwijl de oordelen van de carousel-beoordelaars bij één opleiding gelijk zijn aan die van de Cambridge-examinatoren. Uit de weergegeven betrouwbaarheidsintervallen wordt duidelijk dat de oordelen van de Cambridge-examinatoren verder uiteen liggen dan die van de carousel-beoordelaars; de laatsten blijken minder geneigd een prestatie als heel goed of heel slecht te beoordelen.

In het licht van de bevindingen van Van den Bergh en Klein Gunnewiek (2009) zou men kunnen speculeren dat wellicht op één opleiding relatief veel Engelstalige en op de



Figuur 2. Gemiddelde oordelen van docenten en Cambridge-examinatoren (BC) over Schrijven per opleiding (F = Fontys Lerarenopleiding Sittard; H = Hogeschool Arnhem Nijmegen; R = Hogeschool Rotterdam; N = Noordelijke Hogeschool Leeuwarden; verticale lijnen indiceren betrouwbaarheidsinterval)

andere relatief veel Nederlandstalige carousel-beoordelaars bij de beoordeling betrokken waren, en dat dat een mogelijke verklaring is voor de gevonden verschillen. De gevonden verschillen zouden erop kunnen duiden dat het voor het onderdeel schrijven moeilijk zal zijn voor docenten en opleiders om tot oordelen te komen die qua strengheid vergelijkbaar zijn met de oordelen van Cambridge-examinatoren.

Conclusie en aanbevelingen

Bij de beschreven pilot waren minder deelnemers en minder opleidingen betrokken dan beoogd. Men zou daarom enerzijds kunnen stellen dat de resultaten niet representatief zijn voor alle opleidingen. Anderzijds is het een feit dat dankzij de carousel opleiders van meer dan vier opleidingen bij de beoordeling betrokken waren. Hoe dit ook zij, op

basis van de hier gepresenteerde resultaten kan men niet concluderen dat opleiders en docenten voor alle onderdelen, zelfs niet voor het onderdeel Schrijven, tot oordelen kunnen komen die vergelijkbaar zijn met die van Cambridge-examinatoren.

Op basis van de hier gepresenteerde resultaten lijkt het aanbevelenswaardig voor docenten en opleiders om grote terughoudendheid te betrachten bij pogingen de taalvaardigheid aan de hand van oude Cambridge-examens te beoordelen. De opleidingen Engels hebben dan ook besloten de beoordeling van het eindniveau van de taalvaardigheid van hun studenten in handen van Cambridge English Language Assessment te leggen.

Ondanks de teleurstellende resultaten van de pilot zou men zich kunnen voorstellen dat een herhalingsonderzoek tot andere conclusies zou kunnen leiden. Wellicht is er een 'oud' examen te vinden dat voldoende overeenkomst vertoont met het vernieuwde

CPE-examen. Wellicht is het mogelijk docenten en opleiders zo te trainen dat ze zelfs voor het onderdeel Spreken in staat zijn oordelen te geven die overeenkomen met die van de Cambridge-examinatoren. Cambridge English Language Assessment maakt vanaf 2015 afzonderlijke scores voor de onderdelen van het CPE-examen bekend, waarmee aan één van de bezwaren tegen het CPE-examen zou zijn tegemoetgekomen (UCLES, 2014). Een herhalingsonderzoek zou daardoor ook makkelijker zijn uit te voeren.

LITERATUUR

- Bergh, H. van den, & Klein Gunnewiek, L. (2009). CEF of cijfers bij de beoordeling van schrijfvaardigheid. *Levende Talen Tijdschrift*, 10(1), 3-11.
- Bouwer, R., Béguin, A., Sanders, T., & Bergh, H. van den (2015). Effect of genre on the generalizability of writing scores. *Language Testing*, 32, 83-100.
- Carey, M. D., Mannell, R. H., & Dunn, P. K. (2011). Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing*, 28, 201-219.
- HBO-raad (2009a). *Kennisbasis Nederlands Engels Duits Frans Spaans voor de Lerarenopleiding Voortgezet Onderwijs*. Den Haag: HBO-raad.
- Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60(2), 237-263.
- Johnson, J. S., & Lim, G. S. (2009). The influence of rater language background on writing performance assessment. *Language Testing*, 26, 485-505.
- Kim, Y. H. (2009). An investigation into native and non-native teachers' judgments of oral English performance: A mixed methods approach. *Language Testing*, 26, 187-217.
- Macqueen, S., & Harding, L. (2009) Review of the Certificate of Proficiency in English (CPE) Speaking Test. *Language Testing*, 26, 467-475.

- Mansfield, F., & Nuttall, K. (2007). *Proficiency. practice tests with key*. London: Thomson.
- McColly, W. (1970). What does educational research say about the judging of writing ability? *Journal of Educational Research*, 64(4), 148-56.
- UCLES. (2008). *Certificate of Proficiency in English: Handbook for teachers*. Cambridge: University of Cambridge ESOL Examinations.
- UCLES. (2012). *Cambridge English Proficiency: Handbook for teachers*. Cambridge: University of Cambridge ESOL Examinations.
- UCLES. (2014). From Cambridge English to you. May 2014. Geraadpleegd op <http://esolcentresupport.cmail1.com/t/ViewEmail/r/0FDF7CD7682A6B082540EF-23F30FEDED/1971119F6F89DB1138A555EB6E97B45B>.
- Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs*. 's-Gravenhage: SVO.
- Zhang, Y., & Elder, C. (2011). Judgments of oral proficiency by non-native and native English speaking teacher raters: Competing or complementary constructs? *Language Testing*, 28, 31-50.

TON KOET is docent Engels en lerarenopleider. Hij was verbonden aan het Domein Onderwijs en Opvoeding van de Hogeschool van Amsterdam en is verbonden aan de Interfacultaire Lerarenopleidingen van de Universiteit van Amsterdam. Hij houdt zich o.a. bezig met het onderwijs in en de beoordeling van de Engelse taalvaardigheid. E-mail: <A.G.M.Koet@uva.nl>

HUUB VAN DEN BERGH is hoogleraar toetsing en didactiek van taalvaardigheid aan de Universiteit Utrecht. Hij was betrokken bij vele groot- en kleinschalige projecten naar de kwaliteit van ons (taal)onderwijs. E-mail: <h.vandenbergh@uu.nl>