

naal onderzoek, want we weten niet hoe goed de leerlingen in deze leerwegen spelden toen zij in de achtste groep van het primair onderwijs zaten. Misschien gaan juist veel zwakke spellers wel naar deze beide leerwegen van het vmbo. Hoewel dit waarschijnlijk is, kunnen we op grond van deze gegevens daarover geen enkele uitspraak doen.

LITERATUUR

- Bonset, H. (2010). Spelling in het onderwijs: hoe staat het ermee, en hoe kan het beter? *Levende Talen Tijdschrift*, 11, 3-17.
- Expertgroep Doorlopende leerlijnen (2008). *Over de drempels met taal. De niveaus van taalvaardigheid*. Enschede: SLO.
- Gein, J. van de (2010a). Komd een kind van de basisschool. Onderwijscommissie onderschat spelvaardigheden basischolieren. *Onze Taal*, 79, 228-230.
- Gein, J. van de (2010b). Wenselijke, haalbare en feitelijke niveaus voor de werkwoordspelling. Kanttekeningen bij de empirische onderbouwing van een doorlopende leerlijn. *Levende Talen Tijdschrift*, 11, 25-29.
- Sijstra, J., Schoot, F. van der & Hemker, B. (2002). Balans van het taalonderwijs aan het einde van de basisschool 3. *Uitkomsten van de derde peiling 1998*. Arnhem: Citogroep.
- Kuhlemeier, H. & Bergh, H. van den (1989). *De proefpeiling PPO: Een onderzoek naar het niveau van het onderwijs in het Nederlands in de derde klas van het voortgezet onderwijs*. Arnhem: Cito.

ANNEMIEKE VAN ES heeft Nederlandse taal en cultuur gestudeerd aan de Universiteit Utrecht en is momenteel werkzaam als docente op een middelbare school in Barneveld. E-mail: <Annemike.van.es.@gmail.com>

SANNE SPIJKER (1986) heeft Nederlands gestudeerd aan de Universiteit Utrecht en daar haar onderwijsbevoegdheid gehaald aan het IVLOS. Momenteel is ze werkzaam als docent Nederlands op een middelbare school in Gouda. E-mail: <A.J.C.Spijker@students.nl>

HUUB VAN DEN BERGH (1957) is als bijzonder hoogleraar vanwege 'Levende Talen' verbonden aan het departement Nederlands en het Utrechts Instituut voor Linguïstiek (UIL-OTS). E-mail: <h.vandenbergh@uu.nl>

De beoordeling van Engelstalige essays: het ERK in het universitair schrijfvaardigheidsonderwijs

PETRA JANSMA, SAKE JAGER, WANDER LOWIE & KEVIN HAINES

Hoe kunnen taalvaardigheidsdocenten aan de universiteit ondersteund worden in eenduidig gebruik van het ERK? In het EMBED-project hebben we geprobeerd deze vraag te beantwoorden, en een hulpmiddel te ontwikkelen dat docenten van dienst kan zijn. In de academische omgeving van de Rijksuniversiteit Groningen wordt het onderwijs verzorgd door een docententeam met veel verschillende achtergronden en afkomstig uit uiteenlopende toetsculturen. Binnen het taalvaardigheidsonderwijs wordt gestreefd naar het eenduidig beoordelen van geschreven werk geproduceerd door tweede-taalleerders. Onder dergelijke omstandigheden is het een groot voordeel dat het ERK één gemeenschappelijke structuur biedt waarop de interpretaties van de taalbeheersing van studenten kunnen worden gebaseerd. In dit artikel doen wij verslag van het EMBED-project, waarin is getracht de beoordeling van schrijfvaardigheid Engels door meerdere docenten op één lijn te brengen door hulpmiddelen te bieden bij het gebruik van het ERK. Het project had ten doel werkstukken te verzamelen die gebruikt kunnen worden tijdens docenttrainingen en die als representatief kunnen worden beschouwd voor de verschillende ERK-niveaus binnen het schrijfvaardigheidsonderwijs Engels aan de universiteit. Wij

gaan in op de achtergrond en opzet van het project, de bereikte resultaten en de lessen die we hieruit kunnen trekken voor de praktische toepasbaarheid van de aanbevolen standaardiseringsprocedures.

Achtergrond

Het ERK is in de eerste plaats een hulpmiddel voor zelfbeoordeling. Om het mogelijk te maken de verschillende aspecten van de eigen taalvaardigheid te beoordelen, onderscheidt het ERK 34 schalen voor luisteren, lezen, spreken en schrijven, waarin zowel algemene taalcompetenties als communicatieve taalcompetenties aan bod komen. Deze zijn gedefinieerd voor verschillende domeinen en onderscheiden zes verschillende niveaus (Little, 2007). De niveaus zijn door zowel kwantitatieve als kwalitatieve studies gevalideerd (Alderson, 2002; Hasselgreen, 2003). In zijn functie als instrument voor zelfbeoordeling heeft het ERK ertoe bijgedragen dat taalvaardigheid niet meer in de eerste plaats wordt gezien als een door de docent bepaald product, maar als een door de student bepaald proces.

Het ERK heeft tevens in toenemende mate ingang gevonden als leidraad voor de didac-

tiek. Niet alleen binnen bve, vo en hbo, maar ook in het universitair onderwijs heeft het zijn invloed doen gelden. Ongeveer 10 jaar geleden hebben de Nederlandse universiteiten zich geëngageerd aan het ERK bij het vaststellen van de beoogde taalvaardigheidsniveaus in de moderne vreemde talen in de bachelorfase (De Bot e.a., 2001). Sindsdien wordt het kader steeds meer gebruikt voor beoordeling binnen het universitair vreemdetalenonderwijs. Bij de Rijksuniversiteit Groningen is het gebruik van het ERK de afgelopen jaren met name binnen de faculteiten Letteren, Economie en Bedrijfskunde en Rechten sterk toegenomen. Ook bij het trainen van (nieuwe) docenten is het ERK de afgelopen jaren steeds meer van belang gebleken. Door de sterk groeiende internationalisering is het aantal taalvaardigheidsdocenten Engels snel toegenomen. Hierdoor is de behoefte aan meer coördinatie en docenttraining gestegen.

Om vertrouwd te raken met de niveaus en de criteria voor beoordeling die in het ERK zijn opgenomen, worden regelmatig docenttrainingen (standaardiseringsbijeenkomsten, zie Council of Europe, 2009, 37-55) georganiseerd. Hierbij wordt gezamenlijk gekeken naar het ERK en de verschillende descriptoren en niveaus, waarna werkstukken van taalleerders worden beoordeeld en besproken. Deze standaardiseringsbijeenkomsten maken docenten beter bekend met het ERK en bevorderen de eenduidige interpretatie van het ERK. Ook worden docenten aangemoedigd om, naast Dialang, gebruik te maken van Cefrain en Webcef om bekend te raken met het ERK.

Het EMBED-project is gericht op schrijfvaardigheid op niveaus B2, C1 en C2, omdat dit overeenkomt met het niveau van de meerderheid van onze studenten. Nederlandse studenten komen over het algemeen op de universiteit met ten minste een B2 niveau voor Engels. Het vereiste niveau voor studenten om Engels te halen in bijvoorbeeld het

programma van de faculteit Rechten is C1, wat inhoudt dat het kwalitatieve onderscheid tussen B2 en C1 bijzonder relevant is in onze context.

Het project

Het EMBED-project heeft als doel het beoordelen van geschreven werk van studenten met behulp van het ERK te vergemakkelijken. Hierdoor wordt de inhoud en kwaliteit van het examen of de toets transparanter voor zowel docent, toetsers als getoetste, zoals geadviseerd in de *Manual* (Council of Europe, 2009, 27). Tevens zijn in het EMBED-project verschillende activiteiten op het gebied van docenttraining samengebracht, zoals het verzamelen en beoordelen van relevante voorbeelden voor training en het online beschikbaar stellen van hulpmiddelen waarmee het standaardiseringsproces kan worden ondersteund. Op deze manier wordt het ERK beter ingebed in het curriculum.

Het uiteindelijke doel van EMBED is het wegnemen van verschillende belemmeringen om tot een consistentere beoordeling van schrijfwerkstukken te komen en het ERK betekenisvol te maken binnen de lokale context. Hiertoe wordt een gemakkelijk toegankelijke informatiesite ingericht, als een soort ERK-toolkit voor coördinatoren en docenten, die gebaseerd is op de procedures voor het vertrouwd maken met het ERK en standaardisering, zoals deze worden beschreven in de *Manual*. In deze toolkit worden verschillende lokaal geproduceerde academische schrijfvaardigheidswerkstukken samengebracht voor referentie- en trainingsdoeleinden. Op deze manier kunnen docenten een deel van het standaardiseringsproces online doorlopen en hebben coördinatoren eenvoudig toegang tot informatie en materialen die ze kunnen gebruiken tijdens standaardiseringsbijeenkomsten. De werkstukken worden geru-

briceerd volgens de indeling van het ERK en doorzoekbaar gemaakt. Naarmate het aantal representatieve werkstukken toeneemt, zal deze databank steeds interessanter worden voor coördinatoren, docenten en wellicht in de toekomst ook voor studenten.

Om deze databank mogelijk te maken, is het EMBED-project begonnen met het verzamelen van door studenten geschreven werk, de zogenaamde 'lokale voorbeeldwerkstukken', met als doel deze geschikt te maken voor gebruik in docenttrainingen. Onze procedures zijn een uitwerking van een kernvraag uit hoofdstuk 5 van de *Manual*: hoe ervoor te zorgen dat dergelijk 'lokaal' materiaal van goede kwaliteit is om te gebruiken als ijkpunt (en toekomstige instructie) (2009: 54).

Procedures

Onze eerste stap in het proces van het verzamelen van werkstukken was het specificeren van de schrijfpdracht. We hebben hiervoor de Content Analysis Grid voor het Schrijven van Taken (deel B2 van aanhangsel B in de *Manual*) gebruikt. De volgende stap was het screenen van een eerste verzameling 'authentieke' essays, die zouden kunnen dienen als voorbeeldwerkstukken. De *Manual* beveelt aan dat de eerste essays die ter illustratie gebruikt worden, een relatief 'vlak profiel' hebben voor de descriptoren Vocabulary, Coherence, Accuracy en Argumentation, zoals beschreven in tabel B4. Dit houdt in dat een taalleerder op deze descriptoren ongeveer hetzelfde niveau heeft (2009: 45).

Vooraf stond vast dat het een uitdaging zou zijn om dergelijke 'vlakke' werkstukken te vinden. Onze ervaring met taalleerders is namelijk dat bepaalde elementen van taalontwikkeling pieken zullen vertonen tengevolge van natuurlijke progressie en ongelijke ontwikkeling van deze elementen. Daar komt bij dat de descriptoren van het ERK de kwanti-

tatieve en kwalitatieve kenmerken van taalvaardigheid niet voldoende onderscheiden, terwijl dat onderscheid er in de praktijk wel is. Het lokaliseren van 'authentieke' werkstukken met een vlak profiel bleef echter een belangrijk vertrekpunt in het project.

We hebben 120 door Rechtenstudenten geschreven Engelstalige essays gescreend en hieruit die werkstukken geselecteerd voor elk van de relevante niveaus (B2, C1 en C2) die door de docenten waren beoordeeld als vlak van profiel. Zeven werkstukken voldeden aan deze criteria.

De volgende stap was het opzetten van een panel van deskundigen om de werkstukken nogmaals te beoordelen. Een 'lokaal' panel werd samengesteld bestaande uit vijf docenten Engels, werkzaam binnen de vakgroep Engelse Taal en Cultuur en het Talencentrum. Tevens werd een lokale consultant ingezet, om de interpretaties van het panel te controleren. De rol van de lokale consultant was om als klankbord te fungeren bij de ontwikkeling van onze procedures en om een *second opinion* te geven over de niveaus zoals vastgesteld door het lokale panel.

De rol van de lokale panelleden was om individueel het niveau van elk van de zeven werkstukken te beoordelen. Voor dit doel hebben we een projectsite ontwikkeld in Blackboard die later ook toegankelijk is gemaakt voor externe beoordelaars. De taakbeschrijving werd in een gemakkelijk toegankelijke vorm verstrekt aan het panel ter verduidelijking van de context en de voorwaarden van de schrijfpdracht (zie tabel 1).

Vervolgens bepaalden de panelleden het niveau van alle werkstukken volgens de descriptoren Vocabulary, Coherence, Accuracy en Argumentation. Ook werd hen gevraagd een algemeen niveau (Overall) aan elk werkstuk toe te wijzen. De beoordelaars werden aangemoedigd om hun bevindingen te commentariseren, wat ons in de gelegenheid stelde de

TARGET LEVEL FOR TASK COMPLETION	C ₁
Task Purpose	To measure student ability to write an argumentative essay based on current world issues.
Background to task	Second semester writing exam. Students had to write an argumentative essay at B2.2 level under exam conditions to pass semester one. They have been writing essays in class and receiving peer and teacher feedback on them.
Structure of task	A single writing task consisting of one 500-word essay, based on one of two preselected questions.
Specific information	The candidates are asked to write an essay to demonstrate their ability to formulate an argumentative thesis, support that thesis with appropriate arguments, refute an opposing argument and round off with an effective conclusion. The students will be assessed on their use of academic and legal language, use of formal register, their ability to link their ideas together in clear, smooth flowing text, and to maintain a high degree of grammatical correctness. The students must achieve an average of a C ₁ on accuracy, range, coherence and argument and specifically at least a C ₁ on accuracy in order to pass the course.
Number of tasks	One
Time for task (component)	75 minutes
Integration of skills	Integration of information from multiple sources (preparatory coursework)
Channel	Handwritten
Number of words expected	500
Genre	Academic essay
Imagined audience	Academics within the field of international law
Rhetorical function	Arguing, persuading and describing
Register	Formal academic
Domain	Educational/academic
Assessment criteria	Range, coherence, accuracy, argument

Tabel 1. Taakbeschrijving op basis van de ERK-tabel voor Schrijfp opdrachten (Manual 2009, p. 161-164)

gedachtegang achter de verschillende interpretaties te achterhalen.

Daarna hebben we de interpretaties van de lokale panelleden geanalyseerd en de drie werkstukken verwijderd waarvoor de evaluaties het minst consistent bleken. De beoordeling door het lokale panel resulteerde hierdoor in vier Rechtenessays, elk met een relatief vlak profiel (zie Resultaten hieronder). De beoordelingen van de panelleden met betrekking tot de ERK-niveaus voor elk van de standaardcriteria kwamen goed met elkaar overeen, en de opmerkingen van de lokale consultant ondersteunden deze lokale interpretaties in grote lijnen.

We hebben het gebruik van de vier werkstukken getoetst in een standaardiseringssessie voor de docenten van de cursus Engels voor Rechten. In navolging van de beschreven procedure in de *Manual* (2009: 36-55) hebben we eerst een externe standaardiseringsprocedure gevolgd door de docenten de ERK-procedures op de Cefrain-site te laten doornemen. De docenten werd daarna gevraagd om de vier EMBED-werkstukken te beoordelen en met elkaar te vergelijken, waarbij zij hun interpretaties dienden te onderbouwen door middel van de ERK-descriptoren. Vervolgens werden de belangrijkste moeilijkheden en interessantste bevindingen plenair besproken.

De commentaren van het lokale panel bij twee van de werkstukken hebben we aan de docenten voorgelegd om te vergelijken met hun eigen opmerkingen. Aan de ene kant had dit ten doel het bevorderen van de standaardisering door het bijstellen van individuele afwijkingen van de algemene interpretaties, terwijl het ons aan de andere kant hielp de commentaren te verfijnen. We hebben bijvoorbeeld naar aanleiding van opmerkingen van docenten het commentaar voor één van de werkstukken verduidelijkt, zodat dit beter bij de ERK-descriptor voor B2 past. We waren het ook eens met de mening van de docenten, dat het oorspronkelijke B2-niveau

voor Coherence in een ander werkstuk iets te streng was.

Om na te gaan of de conclusies van ons panel in overeenstemming waren met interpretaties elders in Europa, hebben we vervolgens een extern panel van deskundigen opgezet. Dit panel bestond uit beoordelaars van partner-universiteiten in Italië (Padua), Finland (Helsinki) en België (Leuven). Ons criterium voor de selectie van deze externe beoordelaars was dat ze ervaring moesten hebben in het gebruik van het ERK in een context vergelijkbaar met de onze. Deze deskundigen beoordeelden de vier werkstukken voortgekomen uit de eerdere selectieprocedures, volgens dezelfde procedure in Blackboard.

Resultaten

Om een volledig beeld te geven van de gevolgte procedures en de uitkomsten zullen hier achtereenvolgens een voorbeeldwerkstuk en de interpretaties door de verschillende panelleden besproken worden.

Should there be pressure for the abolition of power of veto held by the permanent members of the Security Council?

[...] A second argument that supports the view that the power of veto has no place in our modern world is that the power of veto is detrimental to the balance of politics. A draft text needs to be approved by all permanent members before it can become a draft resolution. The power of veto makes this nearly impossible. The power of veto is also rumored to being used for the purpose of the “diplomatic cover” of P5 allies. For example, the USA has used its veto power more than any other power since 1972, often concerning draft resolu-

tions condemning the policies and the actions of the State of Israel.

While the P5 agrees that the power of veto should remain their last resort they use their hidden veto often. Hidden veto is the quite threat of using the power of veto. Hidden veto's can have terrible consequences. For example, the UN failed to act in the 1994 Rwandan Genocide because of hidden veto by France and the US. 800,000 people died because the permanent members saw UN intervention as contrary to their personal agenda's. [...]

Het bovenstaande fragment werd door een student Rechten geschreven tijdens een tentamen taalvaardigheid Engels, waarbij het doel was het schrijven van een argumentatief essay van 500 woorden.

Onafhankelijk van elkaar beoordeelde het lokale panel werkstuk als een goede vertegenwoordiger van niveau C1 (zie tabel 2). Eén panellid vond de Coherence van dit werk echter niet overtuigend C1, en schaalde deze daarom in op niveau B2. Het panellid was namelijk van mening dat, hoewel het werkstuk gestructureerd was, organisatorische structuren en verbindingswoorden (noodzakelijk voor het C1-niveau) ontbraken. Deze beoordelaar stelde dat vooral het volgende voorbeeld geen goede samenhang vertoonde:

A second argument that supports the view that the power of veto has no place in our modern world is that the power of veto is detrimental to the balance of politics. A draft text needs to be approved by all permanent members before it can become a draft resolution. The power of veto makes this nearly impossible.

WERKSTUK	RANGE	COHERENCE	ACCURARY	ARGUMENT	OVERALL
Lokaal panel	C1	B2/C1	C1	C1	C1
Lokale consultant	C1	C1	C1	C1	C1
Extern Helsinki 1	C2	C2	C1	C2	C2
Extern Helsinki 2	C1	C1	B2	C1	C1
Extern Padua	C1	C2	C2	C2	C2
Extern Leuven	C2	C2	C2	C2	C2

Tabel 2. ERK-interpretaties voor Rechten werkstuk 2

De beoordelaar gaf aan dat bij elk van deze zinnen een verbindingswoord ontbrak:

A second argument that supports the view that the power of veto has no place in our modern world is that the power of veto is detrimental to the balance of politics. [For example, a] draft text needs to be approved by all permanent members before it can become a draft resolution. [However, the] power of veto makes this nearly impossible.

De meerderheid van de panelleden beoordeelde de Coherence van dit werkstuk echter als zijnde C1, aangevend dat de tekst als geheel makkelijk te volgen was en er goede verbindingswoorden gebruikt waren. Ook de lokale adviseur verklaarde dat B2 een te lage beoordeling voor Coherence zou zijn.

Interessant is dat de beoordelingen door de externe panelleden variëren tussen C1 en C2 voor alle descriptoren, met een uitzondering voor Accuracy. De meerderheid van de interpretaties voor de descriptor Overall door de externen lag op C2, met in de opmerkingen nadruk op de relevantie en de verscheidenheid. Echter, Helsinki 2, waarvan de beoordeling grotendeels samenvalt met die van het interne panel, wijst erop dat de tekst niet consequent is op Accuracy.

De verscheidenheid in de beoordelingen is deels te verklaren door de aanwezige inconsistentie in de tekst, waarbij de tekst niet op alle descriptoren van hetzelfde niveau is. Dit kan van invloed zijn op de beoordeling van de andere descriptoren. Het kan namelijk zo zijn dat een werkstuk dat consequent bijvoorbeeld B1 scoort op Argument, Coherence en Accuracy, door de beoordelaar ook een B1 toegewezen krijgt op Range, zonder dat het daadwerkelijke niveau hiervan B1 is. Het toekennen van bepaalde niveaus aan bepaalde descriptoren kan verwachtingen scheppen voor het niveau van de andere descriptoren.

Ook kan het zo zijn dat een beoordelaar meer gewicht legt op slechts één van de descriptoren, waardoor de beoordeling van die descriptor van invloed is op de beoordeling van de andere descriptoren.

Het is belangrijk in overweging te nemen wat de gevolgen zouden zijn van het gebruik van een dergelijke inconsistente tekst tijdens standaardiseringbijeenkomsten. Het toont ook aan hoe moeilijk het is een door een taalleerder geproduceerd werkstuk te vinden dat hetzelfde niveau heeft op elk van de descriptoren.

Discussie

Tijdens het EMBED-project is een aantal zaken ten aanzien van het ERK en standaardisering aan het licht gekomen. Het voordeel van het ERK is dat het één gemeenschappelijke structuur biedt waarop de interpretaties van de taalbeheersing van studenten kunnen worden gebaseerd. Deze structuur heeft ook als voordeel dat zij empirisch is ontwikkeld door deskundigen op het gebied van taaltoetsing (Council of Europe, 2009, 4) en dat ze is gebaseerd op eerdere taxonomieën van taalniveaus (Council of Europe, 2001, 23). Het kader wordt verder ondersteund door eerdergenoemde hulpmiddelen zoals Cefrain, die *good practice* bevorderen door de docenten aan te zetten tot het doorlopen van strikte procedures bij de beoordeling van taalvaardigheid. Deze hulpmiddelen zijn tot stand gekomen doordat deskundigen uit verschillende vooraanstaande Europese universiteiten hun krachten hebben gebundeld in verschillende projecten om het gemeenschappelijk kader te kunnen toepassen in de dagelijkse praktijk. Ook de *Manual* zelf is ontwikkeld met inbreng en feedback van experts en mensen uit de praktijk (Council of Europe, 2009, ix). Hierdoor ontstaat draagvlak voor toepassing van het ERK in de beoordeling van curricu-

lumtaken. Door het opvolgen van de aanwijzingen voor taakspecificatie, het vergroten van de bekendheid en standaardisering kan de beoordeling in overeenstemming worden gebracht met de principes en uitgangspunten van het ERK.

Ondanks bovengenoemde voordelen kent het ERK ook een aantal beperkingen. In combinatie met praktische belemmeringen binnen de universitaire omgeving kan dit leiden tot een minder goede toepassing van het kader in de praktijk. Ten eerste zijn de ERK-niveaus omschreven in algemene termen, waarbij bewoordingen worden gebruikt die veel ruimte laten voor persoonlijke interpretatie door de betrokken beoordelaars. Daarnaast zijn de niveaus zeer globaal waardoor academische schrijfvaardigheid aanzienlijke verschillen kan vertonen, zelfs binnen het omschreven niveau. Er kan een aanzienlijk verschil zijn tussen wat wordt beschouwd als een 'hoge B2' en een 'lage B2' (Council of Europe, 2001, 35). Ook is er relatief weinig vergelijkingsmateriaal op het niveau van universitaire studenten Engels, binnen het genre en het domein van hun academische werk (Bedrijfskunde, Rechten, etc.), waardoor het 'benchmarken' aan elders geproduceerde producten, zoals aanbevolen door de *Manual*, wordt bemoeilijkt.

Om deze beperkingen op te lossen is in het verleden binnen het Talencentrum en de Afdeling Engelse Taal en Cultuur geprobeerd de ERK-descriptoren te herzien en toe te spitsen op het lokale academische domein. Dit was gebaseerd op een aanbeveling in de *ERK Manual* uit 2003 waarin de mogelijkheid werd opengelaten om de beoordelingstabel (in het manual tabel 5.5) aan te passen, indien deze niet geschikt (b) leek voor beoordeling van lokale werkstukken (Council of Europe, 2003, 88).

Uiteindelijk bleek echter dat de oplossing niet kon liggen in het herzien van de descriptoren zelf. Het resultaat van een der-

gelijke aanpak was namelijk een aangepaste set descriptoren die wel betekenis had voor de bedenkers ervan, maar niet altijd voor andere docenten. Daarnaast stonden ook de aangepaste descriptoren weer open voor interpretatie, wat van invloed kan zijn op de validiteit ervan als beoordelingsinstrument. Het uiteindelijke gevolg van de herziening was dat het moeilijker was geworden onze beoordeling te staven aan de hand van het ERK.

De publicatie van de 2009 *Manual* gaf meer houvast, omdat deze aanzienlijk duidelijker bleek dan de versie van 2003, vooral wat betreft de uitleg van het belang van taakspecificatie. In de 2009 *Manual* is een bijlage 'CEFR Content Analysis Grids for Writing and Speaking Tasks' opgenomen. Niet alleen heeft deze appendix onze aandacht gevestigd op de noodzaak taken duidelijk te specificeren, maar ze heeft ons ook een relatief gebruiksvriendelijke methode aangereikt om taakspecificaties in directe samenhang te kunnen brengen met de ERK-niveaus. Dit is van onschatbare waarde als uitgangspunt voor het verzamelen van geschikte 'lokale' voorbeeldwerkstukken.

Dit neemt niet weg dat er terechte kritiek bestaat op de descriptoren die worden gebruikt om de verschillende niveaus aan te duiden. Hulstijn (2007) trekt in twijfel of de descriptoren de aard van het taalleerproces correct kunnen weergeven. Hij vraagt zich af of kwantitatieve kenmerken (wat een leerder kan doen) en kwalitatieve kenmerken (hoe goed een leerder dit kan doen) op een objectieve manier in één set descriptoren kunnen worden verenigd. Hij wijst erop dat de kwantitatieve en kwalitatieve indicatoren voornamelijk zijn gebaseerd op oordelen door docenten en andere deskundigen. Ze worden onvoldoende ondersteund door empirische studies onder taalleerders. Hierdoor blijft onduidelijk in hoeverre een leerder die een algeheel niveau B2 op taal-

productie heeft bereikt dit niveau ook moet hebben bereikt binnen de onderliggende, specifieke linguïstische competentieschalen. Het ERK is namelijk gebaseerd op eendimensionale taalvaardigheidsschalen (zie ook North & Schneider 1998), terwijl taalvaardigheid eerder multidimensionaal is. Hoewel het ERK zonder meer toelaat dat een leerder een B1- spreekvaardigheidsniveau heeft en een C1-schrijfvaardigheidsniveau, zijn de kwalitatieve en kwantitatieve aspecten binnen deze niveaus vaak onvoldoende gespecificeerd. Deze specificatie is van groot belang bij de beoordeling van taalvaardigheid in de academische context, waar de kwalitatieve aspecten net zo belangrijk zijn als de kwantitatieve.

Ook tijdens de verschillende fasen van het EMBED-project kwamen dergelijke punten naar voren. Het meest opvallend was dat het bijzonder moeilijk bleek een authentiek werkstuk te vinden voor zowel de taalkundige kenmerken (Range, Accuracy, Coherence) als de structurele (Description, Argumentation). Echter, de verschillen in interpretatie van werkstukken, zelfs tussen beoordelaars van dezelfde afdeling, zijn deels inherent aan beoordelingsprocessen door individuen met verschillende achtergronden en deels het resultaat van het eendimensionale karakter van de descriptoren. Een beoordelaar neemt letterlijk honderden kleine beslissingen bij de beoordeling van een essay over zowel kwalitatieve als kwantitatieve dimensies van taalvaardigheid. Het is bijzonder moeilijk al deze beslissingen om te zetten in een eenduidige beoordeling. Dit benadrukt de waarde van standaardisering bij het bevorderen van overeenstemming tussen de beoordelaars, waarbij we de inherente verschillen in interpretaties moeten accepteren, en zeker niet negeren. De ERK standaardiseringbijeenkomsten zijn derhalve onmisbaar voor het eenduidig kunnen beoordelen van het niveau van de talige producten van taalleerders.

Conclusie

De vraag die we onszelf aan het begin van het project hadden gesteld, was hoe we docenten kunnen ondersteunen in het eenduidige gebruik van het ERK. Er was behoefte aan authentieke werkstukken die gebruikt kunnen worden als hulpmiddelen tijdens docenttrainingen, omdat ze als representatief kunnen worden beschouwd voor de ERK-niveaus binnen het universitaire schrijfvaardigheids-onderwijs. We hebben onderzocht of het mogelijk is om dergelijke lokale, domeinspecifieke werkstukken te identificeren die consequent de verschillende niveaus van vaardigheid van onze studenten vertegenwoordigen.

Deze praktische toepassing van het ERK als beoordelingsinstrument bleek verre van eenvoudig. De stapsgewijze procedure in het EMBED-project heeft aangetoond dat het moeilijk is 'vlakke' werkstukken te vinden, waarin alle descriptoren consequent hetzelfde ERK-niveau hebben. Zelfs na een zorgvuldige voorselectie, gevolgd door een verdere selectie door een ervaren lokaal panel, was de beoordeling van de overgebleven werkstukken niet volledig consequent op alle aspecten. Gezien het multidimensionale en dynamische karakter van taalvaardigheid, is dit niet erg verrassend.

Dit betekent echter niet dat we standaardisering binnen het ERK moeten opgeven. Het project heeft aangetoond dat algemene overeenstemming kan worden bereikt binnen de beoordeling van de verschillende componenten van het schrijven. De belangrijkste conclusie is dat we de verschillende dimensies van taalvaardigheid moeten erkennen. Voor standaardisering betekent dit dat we ons moeten richten op de beoordeling van de specifieke aspecten van taalvaardigheid, het geven van expliciete feedback hierover en het illustreren van de beweegredenen van de keuzes van de 'experts' met behulp van tekstuele voorbeelden. Vanuit het oogpunt van de taalleerder is

dit volkomen logisch: verschillende beoordelingsniveaus voor verschillende aspecten van taalvaardigheid kunnen de leerder helpen bij het vaststellen van sterke en zwakke punten.

Een belangrijke opbrengst van het EMBED-project is dat er een op docent- en expertpanels gebaseerde, kwalitatieve procedure voor de selectie van taakspecifieke lokale werkstukken is ontwikkeld, die vervolgens in de praktijk is getoetst. De geselecteerde werkstukken kunnen worden ingebed in het lokale curriculum door standaardiseringsprocedures zoals die beschreven zijn in de *Manual* (Council of Europe, 2009).

De volgende stap voor dit project is om deze procedure toe te passen op essays geproduceerd door studenten binnen andere programma's aan de Rijksuniversiteit Groningen. Bovendien zal de Blackboard-omgeving waarmee het beoordelingsproces wordt ondersteund verder worden uitgebreid en verbeterd. Ook wordt er momenteel gewerkt aan een vervolg van het project waarbij voorbeelden van spreekvaardigheid worden verzameld, zodat docenten ook ondersteuning krijgen bij het beoordelen van spreekvaardigheid. Tenslotte streven we ernaar om de voorbeeldwerkstukken en onze beoordelingen ervan te delen met belangstellenden uit andere instellingen op websites zoals www.erk.nl. Wij zijn ervan overtuigd dat de EMBED-procedures kunnen worden uitgebreid naar andere taaltaken en binnen andere instellingen kunnen worden toegepast. Hierdoor kunnen de tools ontwikkeld binnen EMBED ook docenten van onderwijsinstellingen op andere niveaus, zoals die binnen het voortgezet onderwijs, van dienst zijn. Indien het project kan worden uitgebreid naar andere universiteiten in ons netwerk, heeft het de potentie om te functioneren als de basis van een databank om de consistente toepassing van het ERK te ondersteunen binnen instellingen met soortgelijke en verwante domeinen.

LITERATUUR

- Alderson, J. C. (Ed.) (2002). *Case studies in the use of the Common European Framework*. Strasbourg, France: Council of Europe.
- Bot, K.de, Bogaards, P., Haan de, P., Hannay, M., Jonge de, B., Coopmans, P., e.a. (2001). Eindtermen taalvaardigheidsonderwijs bachelorfase. In *Invoering Bachelor-Masterstructuur: Vervolgadvies DLG-commissie Bachelor-Masters*. Interne publicatie Faculteiten Letteren en Geesteswetenschappen.
- CEFTTrain Project. (n.d.). Bezocht op 1 december 2010, <http://www.helsinki.fi/project/ceftrain/index.html>
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe (2003). *Relating language examinations to the Common European Framework of Reference for languages: learning, teaching, assessment (CEFR): A manual*. Strasbourg: Language Policy Division
- Council of Europe (2009). *Relating language examinations to the Common European Framework of Reference for languages: learning, teaching, assessment (CEFR): A manual*. Strasbourg: Language Policy Division.
- Dialang. (n.d.). Bezocht op 1 December 2010, de Lancaster University website: <http://www.lancs.ac.uk/researchenterprise/dialang/about>
- Hasselgreen, A. (2003). *Bergen "Can do" Project*. Strasbourg, France: Council of Europe.
- Het ERK verplicht? (n.d.). Bezocht op 1 December 2010, de Europees Referentiekader Talen website: <http://www.erk.nl/schoolleider/erkverplicht/>
- Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language Proficiency. *Modern Language Journal*, 91(4), 663-667.
- Little, D. (2007). The Common European Framework of Reference for Languages:

- Perspectives on the making of supranational language education policy. *Modern Language Journal*, 91(4), 645-655.
- Lowie, W.M., Haines, K.B.J. & Jansma, P.N. (2010). Embedding the CEFR in the academic domain: Assessment of language tasks. *Procedia Social and Behavioral Sciences*, 3, 152-161.
- North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15, 217-262.
- WebCEF. (n.d.). Bezocht op 1 December 2010, <http://www.webcef.eu>.

PETRA JANSMA is als coördinator en docente Taalvaardigheid Engels verbonden aan de vakgroep Engelse Taal en Cultuur van de Rijksuniversiteit Groningen.
E-mail: <p.n.jansma@rug.nl>

KEVIN HAINES is als coördinator en docent Engels verbonden aan het Talencentrum van de Rijksuniversiteit Groningen.
E-mail: <k.b.j.haines@rug.nl>

SAKE JAGER is project manager ICT en Onderwijs en docent Toegepaste Taalwetenschap aan de Faculteit Letteren van de Rijksuniversiteit Groningen.
E-mail: <s.jager@rug.nl>

WANDER LOWIE is als docent en onderzoeker verbonden aan de afdelingen Engels en Toegepaste Taalwetenschap van de Rijksuniversiteit Groningen. Zijn onderzoek richt zich met name op tweede-taalontwikkeling en tweede-taalonderwijs.
E-mail: <w.m.lowie@rug.nl>