

‘Nee, ik ben van de trap gevallen.’

Hoe geschikt zijn de niveaudecriptoren van het CEF voor NT2?

ANNE-MIEKE JANSSEN-VAN DIETEN

De doelstellingen van het Common European Framework of Reference (CEF) - meer samenhang en afstemming op talengebied in alle vormen van onderwijs en de vergelijkbaarheid van taalkwalificaties - zijn enthousiast omhelsd door de overheid. Het CEF vormt momenteel dan ook de basis voor diverse projecten in het vreemde-talenonderwijs van basisschool tot universiteit en in het onderwijs Nederlands als tweede taal aan volwassen migranten (Kerckhoff 2001; Liemberg 2001). Er zijn redenen te over voor aansluiting bij dit Europese initiatief (zie ook de artikelen van Stoks en de Jong in dit nummer). Juist bij zo'n brede acceptatie is het ook nodig kritisch stil te staan bij de uitgangspunten van dit model en te bezien of de geclaimde doelstellingen op basis van het voorliggende document waargemaakt kunnen worden^{*}. Hoewel het document nadrukkelijk ook bedoeld is als hulpmiddel bij besluitvorming rond de inrichting van het talenonderwijs, zullen we ons in deze bijdrage beperken tot de consequenties voor de meting van taalvaardigheid en daarbij in het bijzonder aandacht besteden aan problemen die zich kunnen voordoen bij de inschaling van laag opgeleide migranten.

Wat is taalvaardigheid?

Om taalvaardigheid te kunnen meten dient men een theoretisch concept te hebben van taalvaardigheid. De samenstellers van het CEF hebben zich bij de ontwikkeling gebaseerd op een aantal modellen voor communicatieve competentie en taalgebruik die in de afgelopen decennia ontwikkeld zijn. North en Schneider (1998, 226) vatten die als volgt samen: ‘... the scheme sees communicative language competence (linguistic, pragmatic, socio-linguistic) as a part of general human competences (including socio-cultural competence). Learners draw on these competences to the extent that they can, given conditions and constraints operating, and adopt in the process strategies appropriate to their purpose in the circumstances in order to complete the specific task in the language activity concerned.’

Het oudste van de genoemde modellen is dat van Canale en Swain (1980). Zij ontwikkelden een model voor communicatieve competentie bestaande uit drie componenten: linguïstische (taalregels), pragmatische (taalgebruiksregels) en strategische competentie. Deze drie componenten vinden we terug in de descriptoren van het CEF, die bedoeld zijn om de kwaliteit te beoordelen van de uitgevoerde communicatieve activiteiten. Om te kunnen laten zien in welke mate je over deze compe-

tenties beschikt, moeten communicatieve activiteiten uitgevoerd worden. In een testsituatie zijn dat de toetstaken. Bij de uitvoering daarvan spelen meer factoren een rol dan deze drie criteria. We kunnen hierbij denken aan kenmerken van de kandidaat als, leeftijd, sekse, vooropleiding, kennis van de wereld of persoonlijkheidskenmerken. In een toetsituatie spelen daarnaast ook de kenmerken van de beoordelaars en in sommige situaties die van de toetsafnemers een rol. Het samenspel van al deze factoren levert het uiteindelijke resultaat op. Daarop is gewezen door onder meer Bachman (1990) en McNamara (1995) die dan ook genoemd worden als inspiratiebronnen bij de ontwikkeling van het theoretisch concept. McNamara's artikel, waarnaar verwezen wordt, draagt de titel *Modelling performance: opening Pandora's box* en dat is terecht. Het is een feit dat we niet veel weten over dit soort factoren en hun interactie. Desalniettemin is het voor een juiste interpretatie van toetsresultaten noodzakelijk daarmee, voor zover dat mogelijk is, rekening te houden. North en Schneider (1998) laten op conceptueel niveau wel zien zich bewust te zijn van dat soort factoren door communicatieve competentie op te vatten als een onderdeel van algemeen menselijke competenties, die alle een rol spelen bij de uitvoering van een communicatieve activiteit. De vraag is of we daar iets van terugvinden, en zo ja hoe, in de concrete uitwerking van het model, dat gevormd wordt door beschrijvingen van niveaus voor gebruikers, ontwikkelaars en beoordelaars.

Het CEF-gedeelte dat het meest bekend is, bestaat uit een aantal *can do statements* (Modern Languages 1998, 133) betreffende luisteren, lezen, mondelinge interactie, spreken en schrijven, op zes niveaus. Dit gedeelte is bedoeld voor de gebruikers; leerders die hun eigen niveau met behulp daarvan kunnen inschatten of bijvoorbeeld afnemers, als personeelsfunctionarissen, die kunnen zien wat een persoon met een bepaald niveau kan.

In de bijlage van het document vinden we lijsten met descriptoren. Die zijn onderverdeeld in drie hoofdcategorieën: communicatieve activiteiten, strategieën en communicatieve linguïstische competentie. De communicatieve activiteiten dienen vooral als bron voor leergang- en toetsontwikkelaars. Dat zijn de taken die iemand op een bepaald niveau geacht wordt te kunnen uitvoeren. De descriptoren op het terrein van pragmatische en linguïstische competentie zijn beschrijvingen van de kwaliteit van de taakuitvoering en derhalve vooral bestemd voor beoordelaars. Merkwaardig is dat naar deze bijlage steeds verwezen wordt als zijnde een set van illustratieve voorbeelden, terwijl ze de basis vormen waarop het hele model berust.

Hoe is het model tot stand gekomen?

Bij aanvang van het project is men begonnen met het verzamelen van een bestand descriptoren, ontleend aan een groot aantal bestaande schalen voor de beschrijving van niveaus van taalvaardigheid. Bij de herformulering, categorisering en selectie daarvan is gebruikt gemaakt van de ervaring en deskundigheid van vreemde-taaldocenten uit vier verschillende sectoren van het onderwijs: de onder- en bovenbouw van het voortgezet onderwijs, beroepsopleidingen en het onderwijs aan volwassenen. Vervolgens hebben docenten tien van hun leerlingen met behulp van deze descriptoren beoordeeld en twee weken later nog eens (uitsluitend voor mondeling taalgebruik Engels) een aantal videopnames van geselecteerde leerlingen beoordeeld. De resultaten van die beoordelingen zijn onderworpen aan een statistische analyse die het mogelijk maakt de descriptoren te schalen. Dat wil in dit geval zeggen te ordenen naar wat kennelijk hoger of lager op de niveauladder geplaatst moet worden. Tijdens deze procedure zijn nog een aantal descriptoren weggevallen die niet bleken

te passen in het model. Deze bleken zich te concentreren rond een paar gebieden. Het is niet verwonderlijk dat strategieën een van die gebieden was. Dat zijn vaak moeilijk waarneembare en zeker moeilijk schaalbare kenmerken. Wat ook problematisch was, waren descriptoren uit de categorie 'socio-culturele competentie' én - voor de verdere discussie niet onbelangrijk - docenten bleken moeite te hebben met het inschatten van activiteiten die in hun klassen zelden aan bod kwamen. Dit betrof bijvoorbeeld descriptoren die veelal aan werk gerelateerd waren zoals telefoneren, deelnemen aan vergaderingen of rapporten schrijven (North & Schneider 1998). Dit bouwwerk van niveaubeschrijvingen berust dus op de consensus van een groot aantal vreemde-taaldocenten (292) die het kennelijk moeilijk vonden het niveau in te schatten van zaken waarmee ze niet dagelijks gewend zijn te werken. Dat is geen verwijt aan die docenten. Dat is iets wat we in ons achterhoofd moeten houden bij de interpretatie en generalisatie van het model.

Een tweede punt dat bij de generalisatie van het model van belang is, is de aanname dat beheersing van een vaardigheid op een hoger niveau impliceert dat iemand de activiteiten die lager op de niveauschaal staan met de daaraan gekoppelde kwaliteitsbeschrijvingen, kan uitvoeren (Modern Languages 1998, 166). Dat brengt ons tenslotte bij een derde relevant punt: de koppeling van activiteiten en kwaliteitscriteria.

Het CEF en toetsing

Een van de doelstellingen van het CEF is de bevordering van de vergelijkbaarheid van taalqualificaties binnen de Europese gemeenschap. Daartoe is de aard van de communicatieve activiteiten die iemand op een bepaald niveau kan uitvoeren beschreven, en daartoe zijn criteria voor de kwaliteit van die uitvoering

op diezelfde niveaus geschaald. Eenvoudiger en duidelijker kan het niet. Helaas is de realiteit niet zo simpel.

In een voorstudie (North 1993) en in North & Schneider (1998) wordt terecht gewezen op feiten waaraan bestaande niveauschalen mank kunnen gaan. Een van de vermelde problemen betreft het feit dat veel schalen een combinatie van criteria omvatten op één bepaald niveau. Dat maakt het moeilijk voor beoordelaars om één niveau toe te kennen, omdat in werkelijkheid kandidaten op bepaalde criteria onder en op andere criteria weer boven een bepaald niveau presteren. Een andere tekortkoming betreft het feit dat criteria vaak relatief geformuleerd zijn, dat wil zeggen afhankelijk van de formulering van criteria op een ander niveau. Om die reden hebben zij gekozen voor afzonderlijke, onafhankelijke descriptoren die met ja of nee te beantwoorden zijn. Zij hebben daarmee inderdaad winst geboekt voor wat betreft die onderlinge afhankelijkheid, maar dat betekent nog niet dat daarmee het probleem van de verschillen in prestatieniveau op afzonderlijke criteria bij de toekenning van een niveau is opgelost.

Als men alleen maar hoeft te weten of iemand voldoet aan de eisen gesteld aan spreken op niveau A2, zou men kunnen kiezen voor de simpele methode om representatieve toetstaken op dat niveau te selecteren en te stellen dat iemand moet voldoen aan bepaalde voor die taken relevante criteria. Als die persoon honderd procent ja scoort kan met enige stelligheid gezegd worden dat het een beheerser is van spreken A2. Als die score niet gehaald wordt, en dat zal in werkelijkheid vaker voorkomen, moeten er toch besluiten genomen worden over de niveautoekenning. Als daarover geen afspraken gemaakt worden, wordt getornd aan de vergelijkbaarheid. Iemand die op deze toets wel een score van honderd procent haalt, kan niet laten zien of hij meer kan. Dat was misschien wel een

beheersers van B1. In een onderwijscontext wordt een dergelijke vraag niet vaak gesteld. Daar willen we een preciezer beeld van iemands taalvaardigheid hebben. Het model biedt daartoe de mogelijkheid door taken en criteria rond het gewenste of verwachte niveau te combineren. In dat geval krijgt iemand bijvoorbeeld een taak op A2-niveau, maar daarbij worden criteria gehanteerd op A1-, A2- en B1-niveau. Door aan die criteria een verschillend gewicht toe te kennen, kunnen we nauwkeuriger laten zien met welke kwaliteit een taak overwegend wordt uitgevoerd. Criteria waarvoor A1 behaald wordt, zouden dan gecompenseerd kunnen worden met B1-presaties. De uiteindelijke niveautoekenning is echter in wezen hetzelfde probleem als dat wat geconstateerd werd in niveautoekenningen met gecombineerde criteria op een niveauschaal. Het verschil is dat we nu preciezer weten waar iemand tekortschiet of overpresteert.

Ook met de taken kan flexibel omgegaan worden. De communicatieve taken zijn onderverdeeld in diverse subcategorieën. Daarbinnen zie je dat die taken vaak in essentie hetzelfde zijn, maar dat er met het opklimmen van het niveau elementen aan toegevoegd worden. Zo kan iemand op het terrein van mondelinge interactie in de subcategorie 'doelgerichte samenwerking' op niveau A2 betreffende niet al te ingewikkelde problemen praten over mogelijke oplossingen. Op niveau B1 kan hij daar argumenten aan toevoegen. Als je een dergelijke taak op B1-niveau aanbiedt en de beoordeling splitst in het doen van de suggestie en het geven van de argumenten, dan heb je een taak op twee niveaus en kan de kandidaat beter laten zien wat hij waard is.

Tot dusver hebben we gesuggereerd dat het volstrekt helder zou zijn voor toetsconstructeurs - en daarmee bedoelen we niet alleen professionele - wat communicatieve taken of kwaliteitscriteria op een bepaald niveau inhouden. Die vormen op hun beurt

echter opnieuw een bron voor subjectiviteit. Een descriptor voor vocabulaire beheersing op niveau B1 als *'Shows good control of elementary vocabulary but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations'* (Modern Languages 1998, 220) zal niet onmiddellijk door iedere beoordelaar op dezelfde manier geïnterpreteerd worden. Deze formulering laat overigens ook zien dat verondersteld wordt dat er in de taak complexe gedachten geuit moeten worden of onbekende situaties aan de orde zijn. Dat impliceert eveneens dat de toetsconstructeur en de beoordelaar zouden moeten weten wat voor de kandidaat complex of onbekend is.

Een ander punt dat afbreuk doet aan de vergelijkbaarheid, is de keuzevrijheid bij het koppelen van criteria aan toetstaken. Voor de beoordeling van productieve taken vinden we maar liefst tien subschalen met mogelijke criteria. De samenstellers geven terecht aan dat het onmogelijk is voor beoordelaars zoveel criteria te hanteren. Er moeten dus keuzes gemaakt worden. Het zou voor een taak op een hoger niveau wel eens een heleboel uit kunnen maken voor de score, wanneer een criterium als grammaticale controle al dan niet gehanteerd wordt.

Of de vergelijkbaarheid van taalkwalificaties met dit model bevorderd wordt, hangt af van wat je onder vergelijkbaarheid verstaat. Als daarmee bedoeld wordt dat het een stap vooruit is dat alle neuzen in die zin dezelfde kant opstaan, dat er gerapporteerd wordt in termen van communicatieve activiteiten, dan is het terecht. Als echter de interpretatie is dat we nu exact kunnen vertellen dat de resultaten voor Frans van een leerling uit Lutjebroek die op B1-niveau is ingeschaald, vergelijkbaar zijn met die van een leerling uit Landgraaf, Oberhausen of Florence met dezelfde kwalificatie, dan kunnen we daar vraagtekens bij zetten. Het zou echter ook al een stap vooruit betekenen als met behulp van het CEF bewerkstelligd zou kunnen worden dat er binnen

scholen, regionaal of landelijk tussen scholen, en wie weet zelfs internationaal, taken op bepaalde niveaus uitgewisseld zouden kunnen worden, trainingssessies zouden plaatsvinden voor beoordelaars/docenten en afspraken gemaakt zouden worden over omgaan met criteria en cesuren.

Het CEF en verschillende vormen van taalleren/taalonderwijs

Een van de eisen waaraan het CEF moet voldoen is dat alle vormen van taalleren binnen en buiten het onderwijs met behulp van het framework beschreven moeten kunnen worden. Dat betekent dat er rekening gehouden moet zijn met kenmerken en vermogens van alle soorten leerders. Leerders die een taal leren in een omgeving waarin die taal voertaal is (immigranten) en die dus tegelijkertijd taal-leerders en taalgebruikers zijn, worden expliciet genoemd als doelgroep.

In North & Schneider (1998, 220) wordt samenstellers van andere niveauschalen onder meer verweten dat ze weinig oog hebben voor ‘... *avoiding the dangers of lifting rank-ordered scale content from one context and then using it inappropriately in another*’. Als we bezien hoe de niveauschalen tot stand gekomen zijn, gebaseerd op de oordelen van uitsluitend vreemdetalldocenten in een onderwijssetting, dan lijkt ons dit een typisch geval van ‘de pot verwijt de ketel’. Datzelfde zouden we kunnen zeggen over een ander gevaar waar in North (1993, 5) in navolging van Mareschal (1977) op gewezen wordt: ‘...*confusing cognitive complexity with linguistic complexity in the hierarchy of tasks*’ (zie ook: Janssen-van Dielen 1997, 2000, 2001a, b).

Dat het CEF in de context van het vreemdetalenonderwijs zo breed geaccepteerd wordt, is niet verbazingwekkend. In die context is het een heel bruikbaar instrument. De opklimming in complexiteit van de commu-

nicatieve activiteiten sluit naadloos aan bij de doelstellingenhiërarchie van ons onderwijsstelsel in secundair en tertiair onderwijs. Omdat in dat stelsel met het opklimmen van het aantal uren vreemde-talenonderwijs ook de aard van de taken die verricht moeten worden in dat type onderwijs complexer wordt, kunnen eindtermen voor verschillende onderwijstypen zonder al te veel moeite gebaseerd worden op het CEF. Stoks en de Jong (1999) weten de goede vwo-leerling voor Engels en Frans dan ook probleemloos te plaatsen op respectievelijk niveau B2 en B1. Daarbij worden twee argumenten gebruikt: het aantal onderwijsuren en de linguïstische afstand tot het Nederlands. Het argument dat niet gebruikt wordt, maar dat wel degelijk een rol speelt, is dat die vwo-leerling B2 alleen maar kan halen omdat het vwo tevens opleidt tot het lezen van artikelen en verslagen die betrekking hebben op eigentijdse problemen, waarbij de schrijvers een bijzondere houding of standpunt innemen en tot het begrijpen van eigentijds literair proza. Dat is de omschrijving van de leesvaardigheid op niveau B2. Dat de gemiddelde vmbo-leerling dit niet haalt, ligt niet alleen aan het aantal uren Engels, maar vooral aan het feit dat hij dat in zijn moedertaal ook niet kan. Dat hoeft geen probleem te zijn, ware het niet dat met het opklimmen van de complexiteit van de taken ook de daaraan gekoppelde eisen gesteld aan de kwaliteit, hoger worden. Zoals hiervoor beschreven, kun je daar uiteraard flexibel mee omgaan. Je kunt taken op een lager niveau combineren met kwaliteitseisen op een hoger niveau. Maar de vraag blijft dan hoe vervolgens de betrokken kandidaat zo in te schalen, dat hij een duidelijke, geautoriseerde aanduiding van zijn taalvaardigheid in zijn portfolio kan laten zien, die recht doet aan zijn prestaties. Laten we een voorbeeld uit de praktijk nemen. In een spreektoets werd de kandidaten, in dit geval jongvolwassen tweede-taallearners, gevraagd te reageren op de volgende tot hen

gerichte uiting: ‘Goh, ik zie iets aan je. Ben je naar de kapper geweest?’ Waarop een Turkse jongeman reageerde zonder een seconde na te denken met: ‘Nee, ik ben van de trap gevallen.’ Een taak die valt in te schatten op A2-niveau en het verwachte antwoord was iets in de trant van ‘Ja, gisteren’. Het gegeven antwoord kan op de descriptor ‘*vocabulary range*’ niet anders ingeschaald worden dan op C2. Idiomatiche uitdrukkingen en *colloquialisms* zijn echter in het model enkel voorbehouden aan hen die een hoog opleidingsniveau hebben.

Het probleem van die koppeling komt boven tafel als je tweede-taalleerders met een laag vooropleidingsniveau probeert in te schalen, of als je pogingen doet de doelstellingen voor het tweede-taalonderwijs aan deze doelgroep op basis van het CEF te formuleren. Dat ligt aan de opvatting over wat taalvaardigheid of liever gezegd communicatieve vaardigheid is. Hoe complexer de zaken zijn die je begrijpt en hoe deskundiger je daarover mondeling of schriftelijk kunt redeneren, hoe vaardiger je bent. De samenstellers verantwoorden dit door te stellen dat een individu een scala aan algemene competenties tot zijn beschikking heeft waarop hij een beroep doet bij de uitvoering van een communicatieve taak. Een van die competenties is zijn declaratieve kennis en onderdeel daarvan vormt de kennis opgedaan in formele educatie en dus gaan cognitieve complexiteit, academische ontwikkeling en taalvaardigheid hand in hand in dit model. Het duidelijkst wordt dit gedemonstreerd op de grens tussen niveau B1 en B2. Tot en met B1 wordt steeds gerefereerd aan vertrouwde onderwerpen en persoonlijke zaken. Daarna gaat het veelal over zaken buiten de onmiddellijke belevingswereld en valt regelmatig het woord complex. Waar op de eerste drie niveaus de kandidaat duidelijk het referentiepunt is (vertrouwde onderwerpen) wordt dat principe verlaten op de hogere niveaus. Daar staat niet ‘voor de kandidaat’ of

‘voor mij’ complexe zaken, maar daar worden omschrijvingen toegevoegd die duidelijk maken wat we daaronder dienen te verstaan (literair proza, specialistische artikelen). Voor een hoog opgeleide leerder van het Nederlands, laten we zeggen een Engelse hoogleraar letterkunde, hoeven dat helemaal geen complexe taken te zijn en een laag opgeleide migrant kan niet laten zien hoe goed hij in het Nederlands functioneert, omdat die taken op dat niveau te complex voor hem zijn. Kwaliteitscriteria zijn immers gekoppeld aan activiteiten.

Iets anders wat opvalt in het model is dat het veel gemakkelijker is daarmee flexibel om te gaan op het terrein van de productie en mondelinge interactie dan op het receptieve vlak. Dat komt omdat er wel afzonderlijke kwaliteitsdescriptoren zijn voor productieve uitingen, maar kenmerken van receptieve taken verwerkt zijn in de beschrijvingen van de communicatieve activiteiten zelf. Er zijn bijvoorbeeld geen aparte schalen voor spreektempo of mate van regionaal of sociaal gekleurd taalgebruik. Deze zijn enkel opgenomen in de beschrijving van de communicatieve activiteiten en in de *can do statements*. Accenten verstaan en een normaal spreektempo kunnen volgen zijn kenmerken die uitsluitend figureren op de hoogste niveaus. Daardoor kunnen doelstellingen voor een laagopgeleide migrant die ongeschoold werk gaat doen in de bouw en zowel op het werk als in de schaft zijn collega’s moet kunnen begrijpen, moeilijk geformuleerd worden op basis van het CEF.

Conclusie

Het CEF lijkt een bruikbaar instrument voor het vreemde-talenonderwijs, zij het dat er veel geïnvesteerd zal moeten worden in samenwerking, implementatie en deskundigheidsbevordering, willen de beoogde doelen bereikt worden.

Voor tweede-taalleerders en met name voor de minder hoog opgeleiden onder hen lijkt het minder geschikt. Het is deze doelgroep die het meest behoefte heeft aan valide kwalificaties. Hun wordt veel vaker gevraagd naar de beheersing van een taal die niet hun moedertaal is, dan de autochtone bevolking. Het zou derhalve zeer de moeite waard zijn een validatieonderzoek voor deze groep uit te voeren. Daarbij verdienen de punten ont koppeling van cognitieve complexiteit en taalvaardigheid en toevoeging van afzonderlijke descriptors voor de kenmerken van gesproken en geschreven teksten die begrepen moeten worden, speciale aandacht. Het is aan te bevelen daarbij niet alleen gebruik te maken van de deskundigheid van tweede-taaldocenten, maar ook van die van de leerders zelf. Zij opereren dagelijks in de praktijk in de te leren taal en zijn best in staat die activiteiten in te schalen.

* In verband met het feit dat de definitieve versie van het CEF niet verkrijgbaar was, is dit artikel gebaseerd op de tweede conceptversie *Modern languages: Learning, Teaching, Assessment. A common European Framework of reference*. (1998). Strasbourg: Council for Cultural Co-operation.

LITERATUUR

- Bachman, L. (1990). *Fundamental considerations in language testing*, Oxford: Oxford University Press.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics*, 1, 1-47.
- Janssen-van Dieten, A. (1997). Adult Second-Language Policy in the Netherlands. Some considerations. In Bongaerts, Th. & Bot, K. de (red.) *Perspectives on Foreign-Language Policy* (pp. 201-217). Amsterdam/Philadelphia: John Benjamins.
- Janssen-van Dieten, A. (2000). Hoe 'gemeenschappelijk' is het gemeenschappelijk Europees referentiekader? *Toegepaste taalwetenschap in artikelen*, 64, 59-69.
- Janssen-van Dieten, A. (2001a). Toetsen met behulp van het Europees Referentiekader (CEFR). Een kader voor assessment? *Les*, 110, 25-26.
- Janssen-van Dieten, A. (2001b). Voordelen en beperkingen van gestandaardiseerde toetsen, vooral voor laagopgeleiden. *Les*, 110, 32-33.
- Kerkhoff, A. (2001). Naast het Staatsexamen NT2 een instellingsexamen NT2. *Les*, 110, 5-9.
- Liemberg, E. (2001). Het Europees Referentiekader voor de moderne Talen. Een aanwinst of reden tot ongerustheid? *Les*, 110, 22-24.
- Mareschal, R. (1977). Normes Linguistiques. Détermination, Description, Contenu, Utilité. *Canadian Modern Language Review*, 33, 620-631.
- McNamara, T. (1995). Modelling performance. Opening Pandora's box. *Applied Linguistics*, 16, 159-179.
- Modern languages: Learning, Teaching, Assessment. A common European Framework of reference*. (1998). Strasbourg: Council for Cultural Co-operation.
- North, B. (1993). The development of descriptors of language proficiency. Perspectives, problems and a possible methodology based on a theory of measurement. *NFLC Occasional Papers*. Washington DC: The National Foreign Language Center.
- North, B. & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15, 2, 217-262.
- Stoks, G. & Jong, J. de (1999). Twee nieuwe Europese instrumenten voor het moderne vreemde talenonderwijs. Europees Referentiekader en Taalportfolio. *Levende Talen*, 540, 325-331.