
TOETSING VAN SPREEKVAARDIGHEID

Marleen Colpin & Sara Gysen

Tot het systematisch vastloggen van het niveau van spreekvaardigheid van lagere-schoolleerlingen wil het Steunpunt NT2 van de KU Leuven een deel van de instrumenten verschaffen in het kader van VLOT, het *Volgsysteem Lager Onderwijs Taalvaardigheid* (Colpin e.a. 1997). Dat is een pakket toetsen, aangevuld met richtlijnen voor diagnose en remediering, voor de vier basistaalvaardigheden. Spreken is daar één van. Er zijn spreekvaardigheidstoetsen gepland voor het tweede tot en met het zesde leerjaar.

Hieronder vatten we kort samen wat de basisprincipes voor onze toetsen zijn en hoe we die concreet hebben gerealiseerd in een spreektoets voor het vierde leerjaar. Vervolgens schetsen we hoe we te werk zijn gegaan voor het beoordelen van deze spreektaak.

1 WAT EN HOE TOETSEN?

Om te weten welke inhouden we in de spreektoetsen aan bod moeten laten komen, bekijken we de Vlaamse eindtermen basisonderwijs voor spreken (Departement Onderwijs, 1998). Daar lezen we dat de leerlingen 'het gepaste taalregister moeten kunnen hanteren' als ze de verschillende taken, gespecificeerd per verwerkingsniveau, tekstsoort en publiek, uitvoeren. De lijst van deze taken is kort, dus de eindtermen zijn eerder vaag en beknopt, maar bevatten anderzijds wel de nodige criteria om spreektaken op te stellen voor de hele lagere school.

Om (spreek)vaardigheid te evalueren beschikken we over twee grote middelen: observatie-instrumenten en toetsen. Wij hebben gekozen voor toetsen, naar analogie met de andere vaardigheden binnen VLOT, waarvoor de metingen op vaste tijdstippen gebeuren. Volgens de didactiek van het *taakgericht taalonderwijs* (Colpin, Linsen & Van Gorp 1997) worden die toetsen geïntegreerde taken, d.w.z. spreektaken waarin de leerlingen hun totale spreekvaardigheid moeten aanwenden om ze tot een goed einde te brengen. Het zijn bovendien taken die communicatief relevant zijn voor leerlingen van een bepaald leerjaar. Uiteraard gaat het altijd om een simulatie, maar dat is moeilijk uit te sluiten. Toch kunnen we spreken van directe toetsen: we toetsen precies die dingen waarvan we willen weten of de leerlingen ze kunnen.

De praktische problemen bij het evalueren van spreekvaardigheid zijn bekend. Zo neemt het individueel toetsen van leerlingen – wat niet te vermijden is – veel tijd in beslag. Je moet als toetsafnemer/scoorder ook zeer alert zijn en duidelijk op voorhand bepaald hebben welke criteria je bij de beoordeling hanteert.

2 EEN SPREEKTOETS VOOR HET VIERDE LEERJAAR

Hieronder stellen we een voorbeeld van de bovengenoemde taak/toetssoort voor. Deze toets werd in juni 1998 uitgetoetst bij een 100-tal leerlingen. De toets bestaat uit vier onderdelen:

(1) *Verjaardagsfeestje*

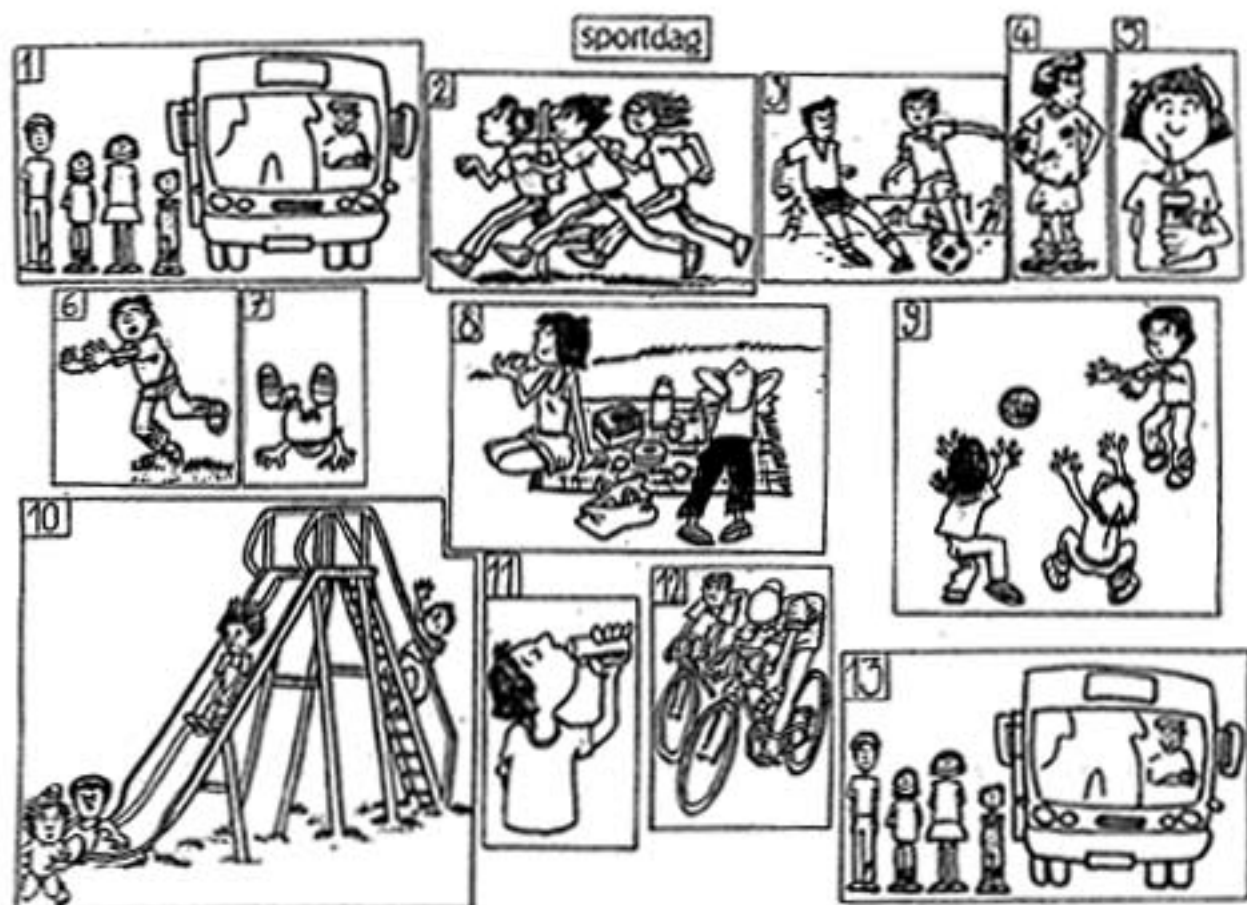
In de eerste opdracht moet de leerling een vriendje opbellen en uitnodigen voor een verjaardagsfeestje. De toetsafnemer zegt op voorhand duidelijk wat de getoetste leerling allemaal moet zeggen en vragen. Als de leerling de dingen in kwestie tijdens het telefoongesprek zelf zegt of vraagt, plaatst de afnemer een kruisje in de tweede kolom van tabel 1. Moet hij de vraag naar dit of dat onderdeel zelf stellen en krijgt hij er een correct antwoord op, dan zet hij een kruisje in de vierde kolom. Zo wordt de toets al inhoudelijk beoordeeld tijdens de afname. Tegelijkertijd wordt het gesprek ook opgenomen op band, zodat het geheel achteraf kan worden geverifieerd, op deugdelijkheid geëvalueerd en op andere punten beoordeeld.

TABEL 1: Scoreformulier 'Verjaardagsfeestje'

LEERLING ZELF		ANTWOORD OP VRAAG VAN DE TOETSAFNER	
Waarom?		Waarom bel je?	
Waar?		Waar is het feestje?	
Wanneer?		Wanneer is het feestje?	
Begin? Einde?		Om hoe laat begint het? Om hoe laat is het gedaan?	
Wat meebrengen?		Moet ik iets meebrengen?	

(2) *Sportdag*

In de tweede opdracht krijgen de leerlingen een reeks tekeningen te zien. Die beelden uit wat ze die dag zagezegd op de sportdag gedaan hebben. De getoetste leerling moet spelen dat hij aan zijn ouders vertelt wat hij allemaal gedaan heeft. Als hij belangrijke informatie vergeet te vermelden, moet de toetsafnemer ernaar vragen.

(3) *Speeltuin*

In de derde plaats moet de leerling spelen dat hij in een speeltuin een meisje van de glijbaan heeft zien vallen. Hij krijgt een tekening van vóór en na het ongeluk te zien. Hij moet bij een huis gaan aanbellen en om hulp vragen. De persoon die de deur opendoet (gespeeld door de toetsafnemer) heeft het ongeluk niet zien gebeuren. Als de leerling belangrijke informatie vergeet te geven, moet de toetsafnemer ernaar vragen.

(4) *Sportkamp*

Ten slotte krijgen de leerlingen een zeer vage affiche voor een sportkamp te zien. Ze moeten de organisator van het sportkamp opbellen en om meer informatie vragen. Op basis van de intuïtie van de makers van de toetsen én de proefafnames kan een lijst van noodzakelijk te stellen vragen worden gemaakt. Op die inhoudelijke parameters worden de leerlingen beoordeeld. Over de onderdelen waar de leerling niet uit zichzelf naar vraagt, moeten geen hints worden gegeven.

Als we deze vier taken naast de eindtermen voor spreken leggen, zien we meteen heel wat verbanden. Uiteraard staan deze taken voor het vierde leerjaar nog op een lager niveau dan wat in de eindtermen voor de leerlingen op het einde van het zesde leerjaar wordt vooropgesteld.

We hebben hier gewerkt met vier verschillende opdrachten om een zo groot mogelijke verscheidenheid in te bouwen. Dat die verscheidenheid ook relevant is, blijkt ook als

we het schema dat Van Gelderen (1992) samenstelde met 'parameters van taalgebruikssituaties voor de selectie en constructie van spreektaken' naast onze vier taken leggen. De taken variëren over criteria als doel, context, relatie spreker/luisteraar, bekendheid met de situatie, formaliteit enzovoort.

3 WAT EN HOE BEOORDELEN?

Het beoordelen van spreekvaardigheid kan in principe twee richtingen uitgaan. Men kan zich richten op aparte componenten van spreekvaardigheid zoals bijvoorbeeld de uitspraak of de grammaticale juistheid van uitingen. Of men kan vooral het gebruik van taal om een communicatieve doel te bereiken, beoordelen. Spreekvaardigheid wordt dan als één globale vaardigheid beschouwd en beoordeeld. Omdat het gebruik van taal in communicatie zowel voor het onderwijs (zie eindtermen) als voor de maatschappij het belangrijkste doel is, ligt het voor de hand om de tweede manier van beoordelen te kiezen. Dit sluit ook aan bij de manier waarop we het toetsen en beoordelen van de andere vaardigheden (luisteren, lezen en schrijven) in VLOT hebben opgevat.

Wanneer we opteren voor die functionele visie op spreekvaardigheid, zullen de criteria bij het beoordelen voornamelijk betrekking hebben op de communicatieve kwaliteit van het spreken. Meer bepaald: is het spreken van die kwaliteit dat de spreker erin slaagt te functioneren in communicatieve situaties? Maar wat verstaan we nu precies onder die communicatieve kwaliteit? In de eindtermen vinden we alleen de vage formulering 'het gepaste register kunnen hanteren'. Slaat dit op inhoud of vorm, of op allebei?

Uit de manier waarop de eindtermen voor spreken zijn ingevuld, leiden we af dat dit zeker ook op inhoud moet slaan. Toch is ook de vorm (waaronder we even heel algemeen 'verstaanbaarheid' verstaan) van belang, meer nog dan bij schrijfvaardigheid. Immers, bij schrijven heb je hopelijk een welwillende lezer die alles zal doen (bijvoorbeeld herlezen) om te begrijpen wat er staat. Een luisteraar, hoe welwillend ook, kan een onverstaanbare uiting geen tweede keer beluisteren om er toch iets van te maken. (Wij nemen de uitspraken van de kinderen wel op, maar dat is niet de normale communicatieve situatie, waar we toch zoveel mogelijk willen bij aansluiten.) Spraaktechnische vaardigheden (zoals formulering en uitspraak) die met vorm te maken hebben, zijn dus belangrijk om de luisteraar niet voor begripsmoeilijkheden te plaatsen.

Het beoordelen van spreekvaardigheid moet aan een aantal eisen voldoen:

- *betrouwbaarheid* (d.w.z. dat de oordelen van verschillende beoordelaars op basis van dezelfde beoordelingswijzer moeten overeenstemmen);
- *validiteit* (d.w.z. dat de beoordeling voldoende informatiewaarde over het niveau van de spreekvaardigheid moet opleveren, en dat deze informatie genuanceerd moet zijn);
- *efficiëntie* (d.w.z. dat het beoordelen zo weinig mogelijk extra tijd en inspanning mag kosten).

In Nederlandstalige en Engelstalige literatuur vonden we een aantal voorstellen voor beoordelingsschema's. De meeste schema's voldoen echter niet in het licht van bovenstaande en andere overwegingen. Veelal is de beoordeling volgens deze schema's (vaak puntenscalen) erg globaal en levert ze dus weinig specifieke informatie over

spreekvaardigheid op. Dat komt de validiteit van de beoordeling niet ten goede. Andere schema's zijn dan weer zo uitgebreid dat ze nauwelijks hanteerbaar zijn voor het beoordelen van de toch wel beperkte mondelinge uitingen van lagereschoolleerlingen.

4 EEN ALGEMEEN, FUNCTIONEEL BEOORDELINGSSHEMA

Het 'algemeen, functioneel beoordelingsschema' van Van Gelderen (1992) vonden we echter wel de moeite van het uitproberen waard. We wilden weten of we met dat schema de communicatieve kwaliteit van spreekprestaties zouden kunnen nagaan. Die communicatieve kwaliteit van spreken is bij Van Gelderen het overkoepelende criterium waaruit alle andere criteria zijn afgeleid. Zowel inhoud als vorm (meer spraaktechnische aspecten) komen erin aan bod. De vier dimensies waaruit het schema bestaat, staan voor de verschillende soorten communicatieve functies. Ze kunnen per twee samengenomen worden: *referentie* en *overdracht* enerzijds, die communicatie substantieel bepalen, en *spreekgemak* en *verstaanbaarheid* anderzijds, die nodig zijn voor de kwaliteit van communicatie. De verschillende criteria binnen elke dimensie zijn relevant voor een zo groot mogelijke variatie van spreesituaties.

Het schema is algemeen bedoeld. Dat maakt het ook heel goed hanteerbaar. Het bevat richtlijnen voor het maken van specifieke beoordelingsschema's voor specifieke taken (vooral monologe taken). Het uitwerken van een specifiek schema bestaat erin het meest relevante criterium per dimensie te selecteren. De relevantie van een criterium wordt bepaald door de aard van de specifieke taak. Eén criterium per dimensie volstaat wegens de bijzonder hoge correlaties tussen oordelen op basis van verschillende criteria. Deze werkwijze laat toe om oordelen over spreekvaardigheid in diverse communicatieve situaties onderling te vergelijken, door telkens terug te grijpen naar de vier dimensies.

Op basis van een onderzoek waarin dit beoordelingsschema is uitgetoetst, stelt Van Gelderen (1992) dat het schema tegemoetkomt aan de eisen van betrouwbaarheid, validiteit en efficiëntie. Toch lijken de verschillende criteria binnen de dimensies onderling ons niet altijd verschillende informatie op te leveren. Een aantal criteria binnen de categorieën *verstaanbaarheid* en *spreekgemak* lijken bijvoorbeeld erg op elkaar. Zelf leerde Van Gelderen uit onderzoek dat oordelen over de dimensies *overdracht* en *spreekgemak* erg hoog correleren. Het onderscheid tussen de dimensies is dus betrekkelijk en er is sprake van overlap.

We beschouwen spreken als één globale vaardigheid en willen ze ook zo globaal mogelijk, maar toch voldoende genuanceerd scoren. Het schema van Van Gelderen splitst de vaardigheid naar ons gevoel toch iets te veel op in aparte aspecten. En ten slotte: het selecteren van een relevant criterium bij een specifieke opdracht lijkt geen probleem voor de dimensie *referentie*. Voor de overige dimensies is dat echter veel minder evident.

Rekening houdend met onze bedenkingen bij het schema van Van Gelderen hebben we een eigen versie van het functioneel beoordelingsschema opgesteld:

FUNCTIONEEL BEOORDELINGSSCHEMA

INHOUD	
-	Goed onderscheid tussen hoofd- en bijzaken
-	Relevantie
-	Waarheidsgetrouwheid en volledigheid van informatie
-	Logische opeenvolging van passages
-	Duidelijkheid van woordbetekenissen en verwijzingen tussen zinnen
VORM	
-	Communicatief adequaat: te begrijpen door een welwillende luisteraar
-	Afhankelijk van de inhoud: is de inhoud er niet of is de inhoud fout, dan kan de vorm ervan ook niet beoordeeld worden
SPREEKGEMAK	
-	Vlotheid van spreken

NAAR: Van Gelderen (1992).

Met dit beoordelingsschema willen we spreektoetsen ontworpen voor VLOT evalueren: de spreektoetsen zullen bestaan uit spreekprestaties binnen een hele waaier van specifieke communicatieve situaties.

Het schema is in de eerste plaats gericht op de communicatieve kwaliteit van spreekprestaties. Voor deze communicatieve kwaliteit veronderstellen we dat de drie categorieën *inhoud*, *vorm* en *spreekgemak* essentieel zijn. Een verdere uitbreiding lijkt ons niet nodig omwille van de validiteit: oordelen op basis van dit schema leveren volgens ons voldoende genuanceerde informatie over de spreekvaardigheid van een leerling op. En die uitbreiding is zeker niet wenselijk met het oog op efficiëntie: leerkrachten moeten binnen een redelijk tijdsbestek de spreekprestaties van hun leerlingen kunnen beoordelen.

Net zoals het functionele schema van Van Gelderen (1992) is dit een algemeen schema. Per spreektoets zullen we op basis van dit schema een specifiek beoordelingschema opstellen, aangepast aan de opdracht(en) van een bepaalde spreektoets.

4.1 Inhoud

De eerste beoordelingscategorie, *'inhoud'*, sluit aan bij de dimensie *'referentie'* in het schema van Van Gelderen (1992). Het gaat hier om wat gezegd wordt. Voor elke van de vier opdrachten van onze spreektoets selecteerden we het criterium *'volledigheid van informatie'*. Per opdracht gingen we na wat een leerling moet zeggen opdat hij het opgelegde communicatieve doel (het doel van de opdracht) zou bereiken. Bij opdracht 2 selecteerden we ook nog het criterium *'logische samenhang van passages'*. De onderscheiden items (passages van de sportdag) moeten dus ook in een logisch – niet noodzakelijk chronologisch – verband verteld worden.

Bij het beoordelen bekeken we per opdracht of een leerling de verschillende items al dan niet vermeldde. Werd een item vermeld, dan scoorden we daarvoor een 1. Werd het item pas vermeld nadat de toetsafnemer ernaar vroeg, dan scoorden we er 0,5 voor. We willen immers het spontaan vermelden van een item hoger waarderen dan het vermelden na een vraag van de toetsafnemer.

4.2 Vorm

De categorie 'vorm' richt zich op de manier waarop de inhoud geformuleerd wordt: is de inhoud op een communicatief adequate manier geformuleerd zodat die te begrijpen is door een welwillende luisteraar? (Score per item: 1 of 0.) Deze categorie is afhankelijk van de categorie 'inhoud': is de inhoud er niet of is die fout, dan kan de vorm ervan ook niet beoordeeld worden. Deze categorie moet ook per taak worden geconcretiseerd, afhankelijk van de leeftijd van de leerlingen en het onderwijs dat ze al achter de rug hebben. Zo is het evident dat aan de uitingen van leerlingen van het vijfde leerjaar hogere eisen gesteld worden dan aan leerlingen van een lager leerjaar.

In het geval van de bovenstaande toets voor het vierde leerjaar hadden we bijvoorbeeld vooropgesteld dat de uitingen van de leerlingen moesten voldoen aan het elementaire niveau van de zin: onderwerp + persoonsvorm (in de juiste volgorde en gecorreleerd) + ... Dat bleek echter niet mogelijk, omdat de uitingen van de leerlingen vaak geen volzinnen zijn. Sommige leerlingen houden het echt heel minimaal en in principe kan je dat niet fout rekenen. Kleine vormelijke of uitspraakfouten (bijvoorbeeld 'eerstens' in plaats van 'eerst') worden evenmin fout gerekend. We zijn van oordeel dat een dergelijke kleine fout in deze context niets afdoet aan de begrijpelijkheid van de uiting.

4.3 Spreekgemak

De categorie 'spreekgemak' is in ons geval globaler opgevat dan wat Van Gelderen (1992) bedoelt. In principe wordt hier steeds hetzelfde (namelijk de vlotheid waarmee iemand spreekt) beoordeeld. Het is niet haalbaar en zeker niet zinvol om dit voor elk stukje inhoud te beoordelen. We doen het eenmaal per gehele taak. In het eerste geval zou het veel te sterk doorwegen in de totale beoordeling. Er kunnen ook weer naargelang van de leeftijd en het niveau van de getoetste leerlingen strengere of minder strenge eisen aan deze categorie gesteld worden.

In ons voorbeeld letten we bijvoorbeeld specifiek op:

- *haperen*: 0 als een leerling herhaaldelijk hapert;
- *tempo*: 0 als het tempo onnatuurlijk of onaangepast is (te traag of te snel);
- *stemvolume*: 0 als een leerling té stil praat;
- *natuurlijkheid*: 0 als een leerling te onnatuurlijk spreekt, bijvoorbeeld te monotoon, te 'verkapt', te nonchalant (bij opdracht 3 bijvoorbeeld).

Daarnaast scoren we ook 0 als de leerling wel goed begint, maar stilvalt naar het einde toe en het initiatief helemaal aan de toetsafnemer overlaat. Als een leerling bij een opdracht op geen van deze criteria negatief scoorde, kreeg hij voor die opdracht een 1 voor 'spreekgemak'.

De strengheid waarmee je dit derde criterium beoordeelt, verschilt van opdracht tot opdracht: bijvoorbeeld bij opdracht 1 kun je hogere eisen stellen aan vlotheid dan bij opdracht 4 (wachten om na te denken is daar heel normaal). Ook bij opdracht 2 moet het eerder vlot gaan, omdat de tekeningen houvast bieden.

4.4 Totaalbeeld

De spreektoets bestaat uit 18 items voor 'inhoud', hetzelfde aantal items voor 'vorm' en vier items voor 'spreekgemak'. In ons onderzoek hebben we de prestaties van de leerlingen op de toets op band opgenomen. Nadien hebben we aan de hand van dit schema 30 opgenomen spreektoetsen gescoord binnen een tijdsbestek van 6 uren (5 leerlingen per uur). Het beoordelen van de categorie 'inhoud' gebeurde echter al tijdens de toetsafname zelf. Bij het beluisteren van de opgenomen spreekprestaties hebben we dit wel nog eens herbekeken om eventuele fouten eruit te halen. Deze manier van scoren binnen dit tijdsbestek lijkt ons haalbaar voor iedere leerkracht van een klas van gemiddeld 20 à 25 leerlingen. Dit pleit voor de efficiëntie van ons beoordelingschema.

Met het schema van Van Gelderen (1992), zijn studies erover en onze bedenkingen daarbij in het achterhoofd, nemen we aan dat de drie categorieën van ons beoordelingschema genoeg en voldoende genuanceerde informatie (gezien in het licht van de eindtermen) over de spreekvaardigheid van leerlingen oplevert. Het scoren van de 30 toetsen hebben we met twee gedaan. Over het algemeen stemden onze oordelen heel goed overeen. In een volgende stap zullen we met het schema een aantal toetsen apart scoren om de betrouwbaarheid ervan beter te kunnen nagaan.

De 30 gescoorde leerlingen haalden een gemiddelde totaalscore van 27,8 op 40 (70%). Per opdracht geeft dat het volgende beeld: (1) 8,7 op 11 (79%); (2) 7,2 op 9 (80%); (3) 6,1 op 9 (67%); (4) 5,8 op 11 (53%). De opdrachten gaan van gemakkelijk naar moeilijk. Dat is gedeeltelijk te verklaren doordat er bij de eerste opdrachten een 'herkansing' is wanneer de toetsafnemer naar relevante items vraagt. Opdracht 4 is zonder meer de moeilijkste voor de leerlingen, omdat ze het minst vertrouwd zijn met de situatie en snel tevreden zijn over het aantal vragen dat ze hebben gesteld.

LITERATUUR

- Colpin, M., S. Gysen e.a. (red.): *VLOT. Volgstelsel Lager Onderwijs Taalvaardigheid*. Deurne: Wolters/Plantyn, 1997.
- Colpin, M., B. Linsen & K. Van Gorp: Taakgericht taalvaardigheidsonderwijs in de basisschool. In: *Gids voor het basisonderwijs*. Diegem: Wolters/Kluwer, sept. 1997, p. 149-175.
- Gelderen, A. van: *De evaluatie van spreekvaardigheid in communicatieve situaties*. Academisch proefschrift. Amsterdam: SCO-Kohnstamm Instituut, Universiteit van Amsterdam, 1992 (publicatie 303).
- Ontwikkelingsdoelen en eindtermen. Informatiemap voor de onderwijspraktijk*. Gewoon basisonderwijs. Brussel: Ministerie van de Vlaamse Gemeenschap, Departement Onderwijs, Centrum voor Informatie en Documentatie, 1998.