

J. B. Kuhlemeier, L. Melse en H. van den Bergh\*\*

### Samenvatting

In dit artikel wordt verslag gedaan van een vergelijkend onderzoek naar de effectiviteit van twee leergangen Duits in de tweede klas van het Mavo. Nagegaan is welke invloed grammaticaal en communicatief gericht lesmateriaal heeft op de leerprestaties en de houding van de leerlingen. Het gaat daarbij om Kennzeichen D 2. Programm, een leergang die de overgang van de grammatica-vertaalmethode naar de communicatieve aanpak markeert, en de communicatieve leergang Kontakte. In de 2-Mavo-klassen die met Kennzeichen D werken, behalen de leerlingen hogere prestaties voor spelling, grammatica en woordkennis. Daar staat echter tegenover dat de 'Kontakte'-klassen zich onderscheiden door hogere prestaties op het vakonderdeel schrijven en door een positievere waardering voor de leergang. Geen verschillen zijn gevonden voor de onderdelen lezen, woordbegrip, luisteren en de beleving van het vak Duits. Dit patroon van uitkomsten komt overeen met verschillen tussen de leergangen in uitgangspunten, leerstof en didactiek.

Als de ene leergang effectiever is dan de andere, dan geldt dit meestal voor alle typen leerlingen en klassen in gelijke mate. Anders gezegd: de bijdrage van de begin-, leerling- en klaskenmerken aan de leeruitkomsten na één jaar onderwijs in het Duits verschilt doorgaans niet van leergang tot leergang. Het gevonden effectiviteitspatroon blijkt echter wel afhankelijk van de bevoegdheid, onderwijservaring en didactische oriëntatie van de docent die voor de klas staat.

\* Het onderzoek is uitgevoerd in opdracht van het Instituut voor Onderzoek van het Onderwijs (SVO-project 6125).

\*\* Met dank aan M. Zwarts en M. Ouborg voor hun commentaar op een eerdere versie van dit artikel.

### 1 Inleiding

In de basisvorming worden de communicatieve doelen van het vreemde-talenonderwijs sterk gepropageerd (Ministerie van Onderwijs en Wetenschappen, 1993). Het streven is leerlingen vanaf het begin teksten te laten lezen en beluisteren die authentiek zijn. Ook moeten zij zo spoedig mogelijk spreek- en schrijfp opdrachten uitvoeren die gebaseerd zijn op reële taalgebruikssituaties. Om deze doelen adequaat te kunnen onderwijzen is communicatief lesmateriaal noodzakelijk. Een belangrijke vraag is uiteraard wat de effectiviteit van zulk lesmateriaal is in vergelijking met leergangen die op een meer traditionele leest geschoeid zijn. In dit artikel doen wij verslag van een onderzoek naar de effectiviteit van twee leergangen Duits in de tweede klas van het Mavo<sup>1</sup>. Het gaat daarbij om de vrij traditionele leergang Kennzeichen D 2. Programm (KD) en de communicatieve leergang Kontakte (KO). Om deze veel gebruikte leergangen didactisch te kunnen plaatsen, gaan we eerst in op de ontwikkeling van het vreemde-talenonderwijs in Nederland gedurende de laatste drie decennia, waarbij we ons beperken tot de onderbouw van het voortgezet onderwijs (paragraaf 2). Deze beschouwing is beknopt gehouden; voor uitvoeriger informatie zie onder meer Koster en Matter (1983) en Van Els, Bongaerts, Extra, Van Os en Jansen-Van Dielen (1984). Vervolgens bespreken we de relevantie van het onderzoek voor de basisvorming en formuleren we enkele hypothesen aangaande de te verwachten verschillen tussen de twee leergangen (paragraaf 3). Daarna gaan we in op de methode van onderzoek (paragraaf 4) en bespreken we de bevindingen (paragraaf 5). Het artikel besluit met enkele conclusies en een discussie waarin we de bevindingen confronteren met onze verwachtingen (paragraaf 6).

## 2 Ontwikkelingen in het vreemde-talenonderwijs

### Grammatica-vertaalmethode

Vóór de invoering van de Wet op het Voorgezet Onderwijs (de Mammoetwet) werd de grammatica-vertaalmethode vrijwel overal in het voortgezet onderwijs toegepast. Kenmerkend daarvoor is de ordening van de leerstof volgens grammaticale progressie. Categorie voor categorie wordt behandeld: de vervoeging van het werkwoord, eerst in de tegenwoordige tijd, daarna in andere tijden, en bij Frans en Duits ook in verschillende modi, enzovoorts. De oefeningen zijn gericht op het aanleren van de grammaticale structuren die 'aan de beurt' zijn en op integratie ervan met het reeds geleerde. Daarbij gaat het hoofdzakelijk om vertaalopdrachten, meestal van het Nederlands naar de vreemde taal. Verder neemt het leren van woorden een belangrijke plaats in. Van de vier taalvaardigheden krijgt alleen lezen enige aandacht, de andere drie - luisteren, spreken en schrijven - komen amper aan bod.

De grammatica-vertaalmethode is eeuwenlang toegepast bij het onderwijs in de klassieke talen. Maar eigenlijk gold al in de vorige eeuw dat men deze talen slechts hoefde te kunnen lezen, dus was onderwijs in drie van de vier vaardigheden niet opportuun. Bij de moderne vreemde talen is er uiteraard wel behoefte aan verwerving van andere vaardigheden dan alleen lezen. Naast kritiek op de gebrekkige aansluiting van het vreemde-talenonderwijs bij de maatschappelijke behoeften was er - vóór invoering van de Mammoetwet - kritiek op de didactische aanpak. Het frontaal-klassikale grammatica-vertaalonderwijs werkte nog wel bij het betrekkelijk kleine deel dat in de jaren zestig vreemde talen leerde. Voor vele, doorgaans wat minder begaafde leerlingen die met de invoering van de Mammoetwet één of meer vreemde talen moesten leren, was deze aanpak ten enenmale ongeschikt.

### Audiolinguale methode

De veranderingen die samengingen met de invoering van de Mammoetwet lagen vooral op het vlak van grammatica, taalvaardigheid en nieuwe media. Natuurlijk bestreed niemand serieus dat taalvaardigheid ook een zekere grammaticale vaardigheid veronderstelde, maar wat

van de grammatica geleerd moest worden, wanneer en hoe, daarover liepen de meningen uiteen. De tendens dienaangaande in het begin van de jaren zeventig is als volgt samen te vatten: minder, later en niet (zo) expliciet. Grammatica moest meer 'gedoseerd' aan de orde komen: minder in de onderbouw en meer op momenten dat structuren uit het oogpunt van verwerving van taalvaardigheid nodig zouden zijn. Sommigen vonden dat grammaticale vormen en regels minder, of zelfs helemaal niet, expliciet geleerd moesten worden, maar impliciet verworven, min of meer analoog aan de manier waarop kinderen zich de grammaticale structuur van de moedertaal eigen maken.

Over het doel van vreemde-taalleren ontstond een vrij grote mate van consensus: de beheersing van de vier taalvaardigheden. De examenprogramma's voor de 'Mammoet'-schooltypen vermeldde dan ook dat in alle vier proeven van bekwaamheid moesten worden afgelegd. Verder won de opvatting veld dat de mondelinge vaardigheden van meet af aan een belangrijke plaats in de vreemde-taallessen moesten krijgen. Het talenpracticum, of alleen de bandrecorder, zouden daartoe effectieve mogelijkheden bieden, was de verwachting. Zo werd de audiolinguale methode geïntroduceerd. Deze methode gaat uit van een structuralistische kijk op taal (grammatica blijft dus een belangrijke rol spelen bij taalverwerving) en een op behavioristische theorieën gebaseerde methode van leren. In leerboeken volgens de audiolinguale methode is een luistertekst het uitgangspunt van nieuwe leerstof. Daarin komen nieuwe woorden aan de orde en nieuwe structuren. In de meest pure vorm worden de regels die daaraan ten grondslag liggen niet geëxpliciteerd, ze worden ingeoeffend door middel van drills. De verwachting is dat daarvan een zekere transfer uitgaat naar analoge zinnen, dat leerlingen zo stap voor stap spreekvaardigheid opbouwen en dat zij daarbij van lieverlede in staat zijn grammaticaal welgevormde zinnen te produceren, zonder veel expliciete kennis van de grammatica.

Over het algemeen liep een en ander uit op een teleurstelling. De drills werkten demotiverend, de transfer was onvoldoende en het impliciete grammatica-onderwijs voldeed bij velen niet. Er rees twijfel of het wel zo verstandig was grammaticaregels niet te expliciteren. Was het

eigenlijk niet zo dat regels de weg bij het leren van een taal bekorten? Het verkrijgen van empirische evidentie omtrent deze veronderstelling was niet eenvoudig. In de Verenigde Staten en Zweden werd onderzoek gedaan naar het effect van expliciet en impliciet grammatica-onderwijs (Smith, 1970; Björneberg, 1974; Von Elek & Oskarsson, 1975). Deze onderzoeken werden uitgevoerd in een normale onderwijssetting en men moest ervaren dat het door de veelheid van variabelen moeilijk was om effecten te kunnen toeschrijven aan de wijze van grammatica-leren. Toch kon voorzichtig worden geconcludeerd dat expliciet onderwijs, zeker bij oudere leerlingen en bij leerlingen met goede leercapaciteiten, wat betere resultaten opleverde dan impliciet onderwijs.

### Grammatica plus

Los van de wending die zich op het wetenschappelijke vlak begon te voltrekken, hadden de practici hun conclusie getrokken: men wenste het oude houvast van de grammatica terug. Bij de audiolinguale leergangen werden al spoedig grammatica-overzichten en schriftelijke grammatica-oefeningen gemaakt. Daarnaast verschenen er nieuwe leergangen waarin de sequentiëring van de leerstof op basis van de grammatica weer in ere hersteld was. Deze vonden veel aftrek. Eén van de meest succesvolle was Kennzeichen D. Deze leergang was echter niet zonder meer een replicatie van leergangen uit de jaren zestig. Een verworvenheid van de audiolinguale methode als aandacht voor luistervaardigheid bleef gehandhaafd. In de opvolger van Kennzeichen D, Kennzeichen D 2. Programm, is de aanpak enigszins in 'communicatieve' richting bijgebogen: oefening met de vaardigheden kreeg wat meer ruimte, maar de grammaticale leerstof bleef bepalend voor de ordening van deze leergang.

### Communicatieve benadering

Op de golf 'terug naar grammatica en vertaling' volgde de 'communicatieve wending'. Uitgaande van de gedachte dat taalleren vooral moet zijn het leren gebruiken van taal in situaties zoals deze in de werkelijkheid voorkomen, werden leergangen ontwikkeld waarin oefening met 'de vaardigheden' van meet af aan een grote rol speelt, terwijl grammatica minder volgens de traditionele systematiek wordt aange-

boden en er ook minder apart met grammatica geoefend wordt. De plaats van vertaal oefeningen is deels ingenomen door spreek- en schrijfopdrachten. De, uit het oogpunt van verspreiding, meest succesvolle representant van de communicatieve wending is de leergang Kontakte.

## 3 Basisvorming en lesmateriaal

Eén van de oogmerken van de basisvorming is verbetering van de kwaliteit van het onderwijs. Leermiddelen spelen in dat vernieuwingsproces een cruciale rol. Voor de onderwijskwaliteit zijn leerresultaten de belangrijkste indicator. Als de verwachting omtrent deze verbetering klopt, dan zou een leergang die meer overeenkomt met de doelen van de basisvorming tot betere resultaten moeten leiden dan een leergang die daar verder van afstaat. In de basisvorming wordt het vreemde-talen-onderwijs sterker gericht op dat wat je met de taal moet kunnen doen (dus op communicatie) dan tot heden in de onderbouw vaak het geval is. Vaardigheidsgericht onderwijs is hier het trefwoord. Andere vernieuwende kenmerken zijn, afgaande op de kerndoelen (Ministerie van Onderwijs en Wetenschappen, 1993):

- gebruik van authentiek tekstmateriaal;
- oefentaken die zijn geënt op levensechte communicatieve situaties;
- oriëntatie op wat verondersteld wordt leerlingen aan te spreken;
- het trainen van strategieën die communicatieve competentie vergroten;
- leerlingen bewust maken van het nut van het leren van moderne vreemde talen.

Kortom, het gaat erom het onderwijs interessanter en relevanter (meer praktisch gericht) te maken.

Wat in de kerndoelen van de basisvorming is verrat, is in Kontakte (KO) al min of meer in de praktijk gebracht. De leergang Kennzeichen D (KD), die volgens de auteurs de overgang van de grammatica-vertaalmethode naar de communicatieve aanpak markeert, staat daar veel verder van af. In Melse, Teunisse, Ouborg en Kuhlemeier (1992) zijn de overeenkomsten en verschillen tussen KD en KO vrij uitvoerig beschreven. We volstaan in dit verband met een

Tabel 1

Overeenkomsten en verschillen tussen Kenzeichen D en Kontakte

	Kenzeichen D 2. Programm	Kontakte
algemene typering	markeert overgang van grammaticaal naar communicatief onderwijs	communicatief
leerinhoud	nadruk op verwerving van woordkennis en grammaticakennis	nadruk op verwerving van communicatieve vaardigheden
leerstofopbouw	deels lineair (grammaticale progressie), deels concentrisch	concentrisch
didactische aanpak	overwegend deductief	inductief; spelelementen

schematische samenvatting en een korte beschrijving in Tabel 1.

Zoals in het voorgaande reeds is vermeld, biedt KD veel meer grammaticale leerstof dan KO, terwijl de vorm waarin dat gebeurt in KD traditioneler is; in KO is het grammaticaanbod geen belangrijk ordeningsprincipe meer. Ten tweede bevat KO communicatieve schrijftaken en KD niet. KD bevat wel allerlei schriftelijke oefeningen, maar daarbij hoeven de leerlingen niet zelf te formuleren. Ten derde zou men KO meer leerlinggericht kunnen noemen dan KD. Uiterlijke vormgeving, materiaalkeuze, presentatie van de leerstof en de oefenvormen zijn meer gericht op wat verondersteld wordt leerlingen van deze tijd aan te spreken. Aspecten die daarbij een rol spelen, zijn authenticiteit van de lees- en luisterteksten, attractieve illustraties, spelvormen, grappige elementen, communicatieve oefentaken e.d. Er zijn echter ook overeenkomsten, bijvoorbeeld wat betreft de vaardigheden die aan bod komen, de onderwerpen en de woordenschat.

Over de effectiviteit van de grammaticale en de communicatieve benadering van het vak Duits zijn voor zover bekend geen onderzoeksgegevens beschikbaar. De veel gebruikte leergangen KD en KO, waarvan inmiddels herziene versies voor de basisvorming het licht zagen, zijn tot op zekere hoogte exemplarisch voor deze twee visies op het vreemde-talenonderwijs. Met het oog op de basisvorming is het dan ook van belang om na te gaan wat de leereffecten ervan zijn na één jaar onderwijs. Op grond van de analyse van de inhoud van beide leergangen zijn vóór de uitvoering van

het onderzoek enkele hypothesen opgesteld. Deze hebben uitsluitend betrekking op de te verwachten verschillen en overeenkomsten in leeruitkomsten na één jaar onderwijs in het Duits.

- 1 Allereerst verwachtten we bij KD hogere grammacaprestaties. Aan dit vakonderdeel wordt in KD immers meer aandacht besteed dan in KO.
- 2 Ten tweede verwachtten wij dat KO-leerlingen hun leergang hoger waarderen dan KD-leerlingen. In KO wordt immers meer gebruik gemaakt van authentieke lees- en luisterteksten, attractieve illustraties, spelvormen, grappige elementen, communicatieve oefentaken e.d.
- 3 Ten derde verwachtten wij voor de onderdelen luisteren, lezen, spelling, woordbegrip en woordkennis geen substantiële verschillen in leeruitkomsten. Daartoe leken ons de leerstofinhoudelijke en vakdidactische verschillen tussen beide leergangen niet groot genoeg. Ook ten aanzien van de beleving van het vak Duits verwachtten wij geen essentiële verschillen tussen de beide leergangen. De mate waarin leerlingen een vak als plezierig, moeilijk of nuttig ervaren, is waarschijnlijk sterker afhankelijk van de docent die voor de klas staat dan van de gehanteerde leergang.
- 4 Ten aanzien van de schrijfprestaties stonden wij in dubio. Enerzijds zou de grammaticale kwaliteit van de schrijfproducten van de KD-leerlingen hoger kunnen zijn. In KD krijgt grammatica immers meer aandacht. Anderzijds zouden de KO-leerlingen een voorsprong kunnen hebben waar het de in-

houdelijke en structurele kwaliteit (vgl. par. 4.3) van de schrijfproducten betreft. Anders dan KD-leerlingen hebben zij immers ervaring met produktieve schrijftaken.

## 4 Methode van onderzoek

### 4.1 Proefpersonen en dataverzameling

Zoals hiervoor reeds is vermeld zijn voor het onderzoek twee leergangen gekozen die in de onderbouw van het voortgezet onderwijs zeer frequent gebruikt worden: Kennzeichen D en Kontakte (vgl. Kuhlemeier, 1989). In totaal hebben 53 klassen (35 docenten) van 28 categoriale Mavo-scholen aan het onderzoek deelgenomen. Van de 53 klassen werkten er 27 met KD en 26 met KO. Het aantal leerlingen varieert afhankelijk van het instrument van 1134 tot 1225. De dataverzameling vond plaats aan het begin en het einde van het schooljaar 1990-1991. Aan de hand van strikte aanwijzingen is de afname door de docent zelf verzorgd.

### 4.2 De vergelijking

In het onderhavige onderzoek is niet volstaan met één globale maat voor de effectiviteit van beide leergangen. Immers, als de ene leergang effectiever is dan de andere, hoeft dit niet op alle vakonderdelen tot uiting te komen. Er zijn negen indicatoren voor de effectiviteit van de beide leergangen na één jaar onderwijs verkregen. Met behulp van toetsen zijn de prestaties op zeven vakonderdelen in kaart gebracht: lezen, luisteren, woordbegrip (D-N), woordkennis (N-D), grammatica, spelling en schrijven. Via vragenlijsten is daarnaast aandacht besteed aan twee aspecten op het affectieve vlak: de beleving van het vak Duits en de waardering van het lesmateriaal.

In leergangvergelijkend onderzoek mogen geconstateerde verschillen in prestaties of attitudes niet zonder meer aan de leergang worden toegeschreven. Andere kenmerken dan de gehanteerde leergang kunnen de leeruitkomsten immers mede beïnvloeden. Wat met een bepaalde leergang in één jaar tijd haalbaar is, hangt onder meer af van de voorkennis van de leerlingen waarmee de docenten aan het begin van het schooljaar te maken krijgen (vgl. Jacobs-Hessing, 1989). Ook de intelligentie van de leerlingen doet er vaak toe (Fraser, Wal-

berg, Welch & Hattie, 1987), al lijkt dat bij de moderne vreemde talen minder het geval te zijn (Boogaards, 1982). Voorts spelen de attitudes aan het begin van het schooljaar een rol (Boogaards, 1982; Kuhlemeier, Van den Bergh & Teunisse, 1990). Als de cognitieve en affectieve bagage van de KD- en KO-klassen op voorhand zou verschillen, is een eerlijke vergelijking op basis van de leeruitkomsten na één jaar onderwijs niet zonder meer mogelijk. We hebben daarom het effect van de leergang vastgesteld na uitschakeling van de invloed van zoveel mogelijk andere effecten. Vandaar dat wij aan het begin van het schooljaar de vaardigheid in het lezen van Duitstalige teksten en het begrijpen van Duitse woorden hebben vastgesteld. Bovendien is van elke leerling de intelligentie, de doubleerstatus, de etniciteit en de beleving van het vak Duits in kaart gebracht. Deze variabelen vormen samen de begin- en leerlingkenmerken.

Ook school- en docentkenmerken stellen grenzen aan hetgeen met een bepaalde leergang in één jaar tijd haalbaar is (Good & Brophy, 1986). Zo blijkt uit onderzoek naar effectieve instructie dat de toegewezen leertijd een belangrijke voorspeller van leerprestaties is (o.a. Lugthart, Roeders, Bosker & Bos, 1989; Scheerens, 1989; Creemers, 1992). We mogen aannemen dat met vier lessen Duits per week een hoger eindniveau haalbaar is dan met drie uur. Docenten die over meer onderwijservaring beschikken, behalen veelal hogere leerresultaten (Lugthart e.a., 1987), wellicht omdat zij geleerd hebben om efficiënter met hun leergang om te gaan (Reints, 1989; Edelenbos, 1990). Ook behalen leerlingen bij hoger opgeleide docenten hogere prestaties (Fraser e.a., 1987), misschien omdat hoger bevoegden over meer vakkennis beschikken. Zou nu de ene leergang gebruikt worden door meer ervaren en hoger gekwalificeerde docenten, dan kan dit een eerlijke vergelijking in de weg staan. Vergelijking van de effectiviteit van leergangen moet derhalve plaatsvinden nadat rekening gehouden is met dergelijke verschillen in uitgangspositie en randvoorwaarden. Om deze reden zijn in ons onderzoek - naast de begin-, leerling- en klas-kenmerken - de volgende school- en docent-kenmerken meegenomen: het aantal lessen Duits per week (drie of vier), de schoolgrootte (het aantal leerlingen), de behandelde hoofd-

stukken van de leergang, de bevoegdheid van de docenten, de onderwijservaring, de didactische oriëntatie en de kwaliteit van de docenten in de ogen van de leerlingen.

Deze verklarende variabelen beschouwen we enerzijds als covariaten, om een eerlijker vergelijking tussen de beide leergangen mogelijk te maken. Anderzijds kunnen we daarmee nagaan in hoeverre de effectiviteit van een leergang afhankelijk is van bepaalde kenmerken van de leerlingen, klassen, docenten of scholen. Hier gaat het om de vraag naar de differentieële effectiviteit van een leergang voor verschillende groepen leerlingen, docenten of scholen. Daarvan is bijvoorbeeld sprake wanneer intelligente leerlingen bij de ene leergang naar verhouding meer opsteken dan bij de andere (een aptitude-treatment-interactie van intelligentie met leergang), of wanneer de leeruitkomsten van de onervaren docenten bij de ene leergang relatief hoger zijn dan bij de andere.

### 4.3 Instrumenten

Aan het einde van het eerste leerjaar Duits is voor elk van de vakonderdelen lezen, luisteren, woordbegrip (D-N), woordkennis (N-D), grammatica, spelling en schrijven een toets afgenomen. De beleving van het vak Duits en de waardering van het lesmateriaal zijn via vragen van het Likerttype in kaart gebracht. Met uitzondering van de spellingtoets, waarvan de meetnauwkeurigheid enigszins te wensen overlaat, is de betrouwbaarheid van de instrumenten redelijk tot goed te noemen. Voor precieze gegevens over de beoordeling, betrouwbaarheid en validiteit van de instrumenten wordt verwezen naar Melse e.a. (1992). Hieronder volgt een beknopte beschrijving van de negen effectmaten. Daarna beschrijven we de begin-, leerling-, klas-, docent- en schoolkenmerken die we als verklarende variabelen beschouwen.

#### De effectmaten

De *leestoets* bevat 22 opgaven, waarvan de eerste tien niet alleen aan het einde maar ook aan het begin van het schooljaar zijn afgenomen. Er zijn bij voorkeur authentieke teksten gekozen. Tevens is een zekere spreiding gerealiseerd wat betreft de functie (b.v. verstrooiing, zakelijke informatie), de bronnen (b.v. jeugdbladen,

kranten, brochures, folders), de soort (bericht, mededeling, advertentie e.d.) en de inhoud (o.a. onderwerpen over school, personen, hobby's, 'Landeskunde', toerisme) van de teksten. De vraagstelling is deels gericht op begrip van de hoofdzaak van de tekst (een soort van samenvatting) en deels op het zoeken van informatie.

De *luistertoets* bestaat uit 21 meerkeuzeopgaven. De toets is bestemd voor leerlingen met een geringe kennis van het Duits. Een volledig authentieke luistertekst zou voor de leerlingen te moeilijk zijn. Daarom is de luistertekst speciaal voor dit doel geschreven, waarbij naar een zo groot mogelijke levensechtheid gestreefd is. Wat betreft de inhoud is gekozen voor een situatie (vakantie in het Duitse taalgebied) die voor iedereen 'inleefbaar' is.

De toets *woordbegrip D-N* meet de vaardigheid in het afleiden van de betekenis van Duitse woorden uit de context en/of op grond van de woordvorm. Er zijn vijf teksten gekozen waarin 25 woorden onderstreept zijn. De taak van de leerling is de Nederlandse betekenis op te schrijven. In alle gevallen gaat het om een woord waarvan de betekenis uit de context is op te maken. Bij 12 van de woorden is de vorm en/of de betekenis geheel (b.v. 'Jugendherberge') of gedeeltelijk (b.v. 'Wettkämpfe', 'Schulzeugnis') verwant met het Nederlandse equivalent. De woordbegriptoets is zowel aan het begin als aan het einde van het schooljaar afgenomen.

De *woordkennistoets N-D* bevat 28 items. De woorden zijn ontleend aan (de eerste helft van) de beide leergangen. Bij de beoordeling van de toetsresultaten van de leerlingen is ook gelet op de spelling. Voor een woord dat goed vertaald en correct geschreven is, zijn twee punten gegeven; als het woord goed is maar de spelling incorrect, dan levert dat één punt op.

De *grammaticatoets* bevat 16 meerkeuzeopgaven, waarvan het merendeel betrekking heeft op de vervoeging van werkwoorden, een onderdeel waaraan beide leergangen aandacht besteden. In groepen van drie of vier zinnen is telkens een woord weggelaten. De leerling heeft tot taak aan te geven in welke van die zinnen een gegeven woord met een bepaalde grammaticale vorm past.

De *spelvaardigheid* van de leerlingen is gemeten met zestien driekeuze-opgaven. Elke

opgave bestaat uit een zin waarin woorden, die in beide leergangen voorkomen, zijn onderstreept. Eén van de onderstreepte woorden bevat een spelfout die leerlingen van het desbetreffende niveau regelmatig maken. De taak van de leerling is aan te geven welk woord incorrect is.

De *schrijfpdracht* plaatst de leerlingen in een (voor hen goed voorstelbare) communicatieve situatie. Het gaat om een contact met leerlingen van een Duitse school. Deze leerlingen komen op bezoek. In het kader daarvan schrijft elke leerling aan een Duitse leerling: iets over zichzelf, over thuis en over school. Een echte brief is het nog niet: aanhef en afsluiting worden niet gevraagd, en het begin is gegeven. Gebruik van hulpmiddelen zoals leerboeken en schriften is toegestaan. Voor de beoordeling golden enerzijds de criteria grammatica, spelling en woordgebruik, anderzijds inhoud en structuur. Bij de laatste twee criteria is niet alleen gelet op gebreken, maar ook op positieve aspecten (Melse e.a., 1992, p. 153 e.v.). Elk werkstuk is beoordeeld door dezelfde drie beoordelaars. Deze zijn vooraf getraind. Eén docent had tot taak (in één beoordelingsronde) de score te bepalen voor elk van de aspecten grammatica, spelling en woordgebruik, de beide anderen beoordeelden (ook in één ronde), onafhankelijk van elkaar, de inhoud en structuur van alle werkstukken. Hun scores zijn daarna gemiddeld.

De twee groepen criteria zijn door verschillende personen beoordeeld, omdat het voor beoordelaars vrijwel onmogelijk is meerdere criteria gelijktijdig te beoordelen en ze ook werkelijk los van elkaar te waarderen (vgl. Meuffels, 1986).

De *beleving van het vak Duits* is zowel aan het begin als het einde van het schooljaar gemeten met een aangepaste versie van de Belevingsschaal voor Engels (vgl. Teunisse, 1990; Kuhlemeier e.a., 1990). De 45 uitspraken van de schaal hebben betrekking op het plezier in Duits, de inzet voor het vak, de ervaren moeilijkheid van het vak en het nut ervan.

De *waardering van het lesmateriaal* (WL) is in kaart gebracht met een vragenlijst bestaande uit 13 uitspraken van het Likerttype. Er is onder meer gevraagd naar de waardering, de duidelijkheid en de leerzaamheid van de leergang.

**Begin-, leerling-, klas- en schoolkenmerken**  
Aan het begin van het schooljaar zijn vier indicatoren voor het beginniveau van de klassen verkregen. Naast de beleving van het vak Duits en de beginprestaties voor lezen en woordbegrip is ook de intelligentie van de leerlingen gemeten. In dit onderzoek is gebruik gemaakt van de derde subtest van de PSB, voluit: Prüfungssystem für Schul- und Bildungsberatung (Horn, 1969).

Van de leerlingen beschikken we over twee persoonlijke achtergrondkenmerken: dubleren en etniciteit. Op basis van deze leerlingkenmerken zijn twee maten voor de samenstelling van de klas geconstrueerd: de proportie dubblers en de proportie allochtonen.

Een maat voor de zogeheten institutionele leertijd is verkregen door de docenten te vragen naar het aantal lessen Duits per week (drie of vier lessen). Ook beschikken we over de schoolgrootte, het totaal aantal leerlingen in het schooljaar 1990-1991. Voor de wijze waarop de variabelen geconstrueerd zijn en enkele beschrijvende statistische gegevens wordt verwezen naar Bijlage 1.

#### **Docent- en implementatiekenmerken**

Een grove maat voor de implementatie van de leergang is ontleend aan de logboeken die de docenten gedurende het schooljaar moesten bijhouden. Daarbij is onder meer gevraagd naar het aantal hoofdstukken dat zij dat schooljaar behandelden. Op grond hiervan is de variabele 'proportie behandelde leerstof' geconstrueerd.

Er zijn vier docentkenmerken in het onderzoek betrokken: de onderrichtservaring, de onderrichtsbevoegdheid, de kwaliteit van de docent en de didactische oriëntatie. Evenals de onderrichtservaring en de onderrichtsbevoegdheid zijn de variabelen 'kwaliteit van de docent' en 'didactische oriëntatie' opgenomen vanwege de veronderstelde positieve relatie met leerresultaten. De kwaliteit van de docent is gemeten via een leerlingenvragenlijst. De leerlingen is onder meer gevraagd aan te geven hoe 'aardig' en vakbekwaam zij hun docent vinden. Met betrekking tot de didactische oriëntatie is de docenten gevraagd welke didactische benadering hen het meeste aanspreekt: de meer taalkundige (nadruk op verwerving van grammaticale structuren, woordkennis en

een correcte uitspraak) of de meer communicatieve (zo spoedig mogelijk beginnen met luisteren en spreken in levensechte situaties en met het lezen en schrijven van eenvoudige tekstjes). Over de relatie tussen de didactische oriëntatie en leerresultaten zijn voor zover ons bekend geen onderzoeksgegevens voorhanden. Plausibel lijkt echter de veronderstelling dat wanneer de opvattingen en de intenties van de docent overeenstemmen met die van de leergangauteurs er eerder sprake zal zijn van 'gebruik-zoals-bedoeld' (vgl. Gravemeijer, 1989) en van betere leerresultaten. Voor enkele technische en statistische gegevens zij verwezen naar Bijlage 1.

#### 4.4 Statistische analyse

##### *Overwegingen vooraf*

Verschillen tussen leergangen in effectiviteit zijn tot op heden vrijwel uitsluitend geanalyseerd met behulp van technieken van unilevelanalyse, zoals ANOVA en ANCOVA. Doorgaans wordt daarbij uitgegaan van de aanname dat de data verzameld zijn als een enkelvoudige random steekproef uit de populatie van leerlingen. In ons onderzoek zijn echter scholen/klassen geselecteerd. Omdat leerlingen in één klas ten gevolge van selectie- en onderwijsfactoren meer op elkaar lijken dan leerlingen van verschillende klassen is de aanname van onafhankelijke waarnemingen niet erg realistisch. Voor de effecten van klas-, docent- of schoolvariabelen geeft unilevelanalyse dan standaardfouten die te optimistisch zijn (vgl. Raudenbush & Bryk, 1988). Verder hebben we te maken met variabelen op verschillende niveaus. Voor de afhankelijke variabelen zijn er net zoveel (deels afhankelijke) waarnemingen als er leerlingen zijn. Voor klas-, docent- of schoolvariabelen (leergang, ervaring, bevoegdheid, schoolgrootte e.d.) zijn er veel minder waarnemingen, namelijk net zoveel als er klassen, docenten of scholen in het onderzoek vertegenwoordigd zijn. De geschatte standaardfouten op verschillende niveaus moeten dus op een verschillend aantal waarnemingen gebaseerd zijn. In de door ons gehanteerde methode van analyse - multilevelanalyse - wordt rekening gehouden met zowel het 'clustereffect' als met het verschillend aantal waarnemingen op leerling- en klasniveau. Derhalve

worden correctere schattingen van de standaardfouten verkregen (Goldstein, 1987).

##### *Geanalyseerde modellen*

Voor elk van de negen effectmaten wordt een afzonderlijke univariate multilevelanalyse uitgevoerd. Overeenkomstig de hiërarchische structuur van onze data zouden we in de analyse ten minste drie niveaus moeten onderscheiden: leerlingen, klassen en scholen. De beperkte omvang van onze responsgroep laat het niet toe om een betrouwbaar onderscheid te maken tussen klas- en schoolniveau. We beperken ons daarom tot twee niveaus: leerlingen (genest binnen klassen) en klassen. Alle continue variabelen zijn gestandaardiseerd (met een gemiddelde van nul en een variantie van één).

In het zogeheten *onconditionele model (model 0)* worden van elke effectmaat alleen het gemiddelde en de varianties op leerling- en klasniveau geschat. De uitkomst van het onconditionele model is van belang, omdat de geobserveerde tussen-klassenvariantie de bovengrens vormt van hetgeen we op klasniveau met alle begin-, leerling-, klas-, school- en docentkenmerken en de leergang gezamenlijk kunnen verklaren. Door de tussen-klassenvariantie te delen door de totale variantie verkrijgen we inzicht in de relatieve grootte van de geobserveerde verschillen tussen klassen.

In een volgende analyse (*model 1*) wordt het 'bruto-verschil' tussen beide leergangen geschat. Dit wil zeggen: het verschil op de effectmaten vóór uitschakeling van de invloed van de andere verklarende variabelen. Daartoe voegen we aan het onconditionele model twee regressiegewichten toe: de één representeert het gemiddelde voor KD en de ander het gemiddelde voor KO.

In de hierna beschreven modellen 2 t/m 4 schatten we de invloed van de in paragraaf 4.3 beschreven verklarende variabelen. Voor elke verklarende variabele zijn er twee regressiegewichten<sup>2</sup>: één voor KD en één voor KO.

In het tweede model (*model 2*) nemen we de begin-, leerling-, klas- en schoolkenmerken op; de docent- en implementatiekenmerken blijven nog buiten beschouwing.

In het derde model (*model 3*) breiden we model 2 uit met de docent- en implementatiekenmerken; met uitzondering van de leergang



zijn nu alle verklarende variabelen in de analyse opgenomen.

In het vierde model (*model 4*) voegen we de leergang toe. Het verschil tussen KD en KO in gemiddeld niveau wordt hiermee niet alleen gecorrigeerd voor verschillen tussen beide groepen in cognitieve en affectieve Ausgangssituatie aan het begin van het schooljaar, maar ook voor verschillen in de samenstelling van de klassen naar sekse, etniciteit en schoolgrootte en voor verschillen tussen de groepen docenten qua bevoegdheid, onderwijservaring en didactische oriëntatie. Het verschil tussen de twee regressiegewichten die de gemiddelden van beide leergangen representeren, mogen we nu het 'netto-verschil' noemen. Dit wil zeggen: het verschil op de effectmaat na uitschakeling van de gezamenlijke invloed van alle begin-, leerling-, klas-, school- en docentkenmerken.

Ten slotte zijn, volgens een hier niet nader beschreven stapsgewijze procedure (zie Melse e.a., 1992), de modellen 2 tot en met 4 vereenvoudigd door niet-significante ( $p < .05$ ) regressiegewichten uit de analyse te verwijderen.

De regressiegewichten laten zien wat de invloed is op de effectmaat van elke verklarende variabele afzonderlijk (rekening houdend met alle overige predictoren in het model). Ze geven geen antwoord op de vraag in hoeverre de niveauverschillen tussen klassen aan de leergang dan wel aan andere kenmerken zijn toe te schrijven. Deze vraag beantwoorden we aan de hand van het percentage verklaarde variantie. Een eenduidige maat voor het interpreteren van de grootte van effecten is het percentage<sup>3</sup> van de tussen-klassenvariantie dat verklaard wordt:  $R^2_{\text{klas}}$ . In de evaluatie van de verschillende modellen staat de toename van  $R^2_{\text{klas}}$  centraal<sup>4</sup>. Als de tussen-klassenvariantie klein is in verhouding tot de totale variantie suggereert  $R^2_{\text{klas}}$  een groot effect. Vandaar dat wij ook het percentage van de totale variantie rapporteren:  $R^2_{\text{totaal}}$ .

De grootte van het gezamenlijke effect van de begin-, leerling-, klas- en schoolkenmerken evalueren we aan de hand van de toename van  $R^2_{\text{klas}}$  van model 2 ten opzichte van het onconditionele model (model 0). Omdat dit laatste model geen verklarende variabelen bevat, is het percentage verklaarde variantie uiteraard nul. Het percentage onverklaarde variantie uit model 2 vormt de bovengrens van hetgeen we - gegeven ons model - met alle docent- en leer-

gangkenmerken kunnen verklaren (vgl. Willms & Raudenbush, 1989; Bosker, 1991; Bryk & Raudenbush, 1992). De additionele invloed van alle docent- en implementatiekenmerken gezamenlijk evalueren we aan de hand van de toename van de  $R^2_{\text{klas}}$  van model 2 naar model 3. We krijgen dan een indruk van de bijdrage van de docent onder constanthouding van de Ausgangssituatie en de randvoorwaarden waarmee hij of zij geconfronteerd wordt. De onverklaarde variantie uit model 3 vormt de bovengrens van hetgeen in principe door de leergang (en andere kenmerken die we niet gemodelleerd hebben) verklaard kan worden. Tegen dit laatste percentage evalueren we ten slotte het effect van de leergang (de toename van  $R^2_{\text{klas}}$  van model 3 naar model 4).

## 5 Onderzoeksresultaten

In deze paragraaf gaan we eerst in op de 'bruto-verschillen' (model 1) en de 'netto-verschillen' (model 4) tussen KD en KO in het niveau dat door de leerlingen na één jaar onderwijs in het Duits gerealiseerd wordt (paragraaf 5.1). Daarna bespreken we de invloed van de verklarende variabelen op de leeruitkomsten en gaan we na in hoeverre de effectiviteit van de beide leergangen verschillend is voor verschillende groepen leerlingen, klassen, scholen en docenten (paragraaf 5.2). Tot slot bespreken we in hoeverre de verschillen tussen klassen in gemiddeld niveau zijn toe te schrijven aan de leergang dan wel aan de andere verklarende variabelen (paragraaf 5.3).

### 5.1 Bruto- en nettoverschillen tussen KD en KO

Tabel 2 geeft inzicht in de bruto- en nettoverschillen tussen beide leergangen. Dit wil zeggen: de verschillen in gemiddeld niveau aan het einde van het schooljaar vóórdat en nadat rekening gehouden is met verschillen tussen de KD- en KO-groep in de cognitieve en affectieve positie aan het begin van het schooljaar en met verschillen in de samenstelling van de leerling-, klas- en docentenpopulatie<sup>5</sup>.

Uit Tabel 2 blijkt dat de KO-leerlingen vóór correctie hogere scores behalen op de vakonderdelen schrijven (.94), spelling (.18), de beleving van het vak Duits (.23) en de waardering

Tabel 2

Gemiddeld prestatieverschil tussen KD en KO vóór en na correctie voor begin-, leerling-, klas-, docent- en schoolkenmerken (LE: lezen; WB: woordbegrip; LU: luisteren; SV: schrijven; WK: woordkennis; GR: grammatica; SP: spelling; BE: vakbeleving; WL: waardering lesmateriaal)

	LE	WB	LU	SV	WK	GR	SP	BE	WL
Vóór correctie: KO minus KD	.04	.16	.18	.94 <sup>1</sup>	-.33 <sup>2</sup>	-.14	.18 <sup>2</sup>	.23 <sup>2</sup>	.73 <sup>1</sup>
Na correctie: KO minus KD	-.02	.11	.08	.58 <sup>1</sup>	-1.11 <sup>1</sup>	-1.63 <sup>1</sup>	-1.01 <sup>1</sup>	-.09	.67 <sup>1</sup>

<sup>1</sup>  $p < .01$ ; <sup>2</sup>  $p < .05$ .

van het lesmateriaal (.73). Voor woordkennis zijn de prestaties evenwel lager (.33). Wat betreft lezen, woordbegrip, luisteren en grammatica is er geen verschil tussen beide leergangen aantoonbaar.

Na correctie presteren de KO-leerlingen hoger op het vakonderdeel schrijven (.58). Ook blijken zij hun lesmateriaal hoger te waarderen (.67). Daar staan echter lagere prestaties voor woordkennis (1.11), grammatica (1.63) en spelling (1.01) tegenover. Deze effectiviteitsverschillen tussen de twee leergangen zijn zonder meer groot te noemen (vgl. Cohen, 1977). Geen leergangverschillen zijn aantoonbaar bij de onderdelen lezen, woordbegrip, luisteren en vakbeleving.

Vergelijking van de gemiddelden vóór en na correctie laat in vier gevallen relatief grote verschillen zien. Bij woordkennis, grammatica en spelling is het verschil tussen KD en KO na correctie groter dan ervoor. Bij vakbeleving verdwijnt het significante brutoverschil na correctie. Aan het einde van de volgende paragraaf gaan we hier nader op in.

## 5.2 De verklarende variabelen

In hoeverre vallen de resultaten op de eindmeting te voorspellen met de begin-, leerling-, klas-, school- en docentkenmerken? En in hoeverre zijn de leergangen differentieel effectief voor verschillende groepen leerlingen, klassen, scholen en docenten? In de rapportage van de bevindingen worden regressiegewichten tussen de -.20 en .20 niet besproken. Enerzijds omdat deze gewichten een kleine invloed suggereren. Anderzijds omdat de gewichten voor de klas-, docent- en schoolkenmerken (en sommige leerlingkenmerken) op weinig waarnemingen gebaseerd zijn en dus weinig robuust

zijn. De tussen haakjes weergegeven getallen betreffen de regressiegewichten zoals geschat in model 4. Voor een overzicht van alle significante regressiegewichten wordt verwezen naar Bijlage 2.

## Beginkenmerken

Aan het begin van het eerste leerjaar Duits blijken de Mavo-klassen reeds over enige kennis op het terrein van lezen en woordbegrip te beschikken. Althans, de prestaties liggen hoger dan men op grond van de raadkans zou verwachten. Bovendien blijkt deze beginkennis reeds significant te variëren van klas tot klas (ook nadat rekening gehouden is met zittenblijven). Dit is opmerkelijk, aangezien de klassen, afgezien van zittenblijvers, op dat moment nog geen expliciet onderwijs in het Duits hebben gehad.

Van de vier beginkenmerken is het *begrip van Duitse woorden*, zoals gemeten aan het begin van het schooljaar, veruit de beste voorspeller van het bereikte onderwijsniveau na één jaar onderwijs in het Duits (gecorrigeerd voor de invloed van de overige verklarende variabelen). Leerlingen met hoge prestaties voor woordbegrip op de beginmeting behalen hogere prestaties voor leesvaardigheid (.38), woordbegrip (.52), luisteren (.45), woordkennis (.43), grammatica (.31) en spelling (.25). De *leesprestaties*, de *intelligentie* en de *beleving van het vak Duits* zijn voor de uiteindelijke prestaties van minder groot belang. Echter, de vakbeleving aan het begin van het schooljaar draagt wel bij tot een betere voorspelling van de vakbeleving aan het einde van het schooljaar en de waardering van het lesmateriaal. Veel leerlingen die Duits aan het begin van het schooljaar plezierig, nuttig, relevant e.d. vinden, doen dat ook nog aan het einde van het schooljaar (.44). Bo-

vendien zijn leerlingen met een gunstigere uitgangspositie op affectief gebied meer geneigd hun leergang aan het einde van de rit positief te waarderen (.20).

### **Leerlingachtergrondkenmerken**

De scores van de *zittenblijvers*, de leerlingen die het eerste leerjaar Duits voor de tweede keer doen, zijn niet of nauwelijks onderscheidbaar van die van de niet-zittenblijvers (gegeven de overige variabelen in het model). Eén regressiegewicht is het vermelden waard. Alleen bij KO heeft het zittenblijven een additieve invloed op de prestaties voor woordkennis (.27); bij KD is een dergelijke invloed niet aantoonbaar.

Verschillen tussen *autochtone en allochtone leerlingen* zijn bij vier van de negen effectmaten geconstateerd. Het betreft woordbegrip, luisteren, woordkennis en vakbeleving. Zowel bij KD als KO behalen allochtone leerlingen lagere scores voor woordbegrip (.24) en luisteren (.45). Inspectie van de univariate samenhang tussen etniciteit en woordbegrip laat zien dat de achterstand van de allochtone leerlingen op dit onderdeel gedurende het schooljaar groter wordt (van -.17 naar -.35 standaarddeviatie). Dit effect geldt voor beide leergangen in ongeveer gelijke mate. Bij woordkennis is er evenwel sprake van een voorsprong van allochtone leerlingen (.27); bij beide leergangen kennen zij aan het einde van het schooljaar meer woorden N-D dan de autochtone leerlingen. Ten slotte blijken allochtone KO-leerlingen hun lesmateriaal hoger te waarderen dan hun autochtone klasgenoten (.34).

### **Klas- en schoolkenmerken**

De *samenstelling van de klas*, zoals deze tot uiting komt in de proportie zittenblijvers en allochtonen, is niet of nauwelijks van belang voor de leeruitkomsten op cognitief en affectief gebied. Ook de invloed van de *grootte van de school* is verwaarloosbaar. Op grotere categoriale Mavo-scholen zijn de leeruitkomsten na één jaar onderwijs in het Duits niet hoger of lager dan op kleinere categoriale Mavo's.

Verwacht mag worden dat leerlingen die vier *lesuren Duits per week* krijgen hoger presteren dan leerlingen die slechts drie lessen Duits hebben. Vreemd genoeg is er bij vijf van de zeven cognitieve maten geen enkele relatie

met de leerresultaten aan het einde van het schooljaar aantoonbaar. De twee wel significante regressiegewichten zijn leergangspecifiek, en bovendien negatief! Bij KD scoort de vier-uursvariant lager op het vakonderdeel schrijven (.37), terwijl dit bij KO voor de luisterprestaties geldt (.31).

### **Docent- en implementatiekenmerken**

De *onderwijservaring* van de docent blijkt bij vier van de zeven cognitieve maten een rol te spelen. Hiervan voldoet slechts één regressiegewicht aan ons criterium. Bij KO behalen de ervaren gebruikers hogere prestaties met hun leerlingen voor spelling dan hun minder ervaren KO-collega's (.23).

De *onderwijsbevoegdheid* van de docent speelt een rol bij de vakonderdelen schrijven, woordkennis, grammatica en leesvaardigheid. Zowel voor KD- als KO-docenten geldt dat een hogere bevoegdheid samengaat met hogere schrijffprestaties. Bij de gemiddelde eerstegrader behalen de Mavo-leerlingen hogere schrijffprestaties dan bij de gemiddelde derdegrader ( $2 * .24 = .48$ ). De overige drie relaties zijn leergangspecifiek in de zin dat de bevoegdheid er bij KO-docenten niet toe lijkt te doen, maar bij de KD-docenten wel. In vergelijking met KD-tweedegraders realiseren KD-eerstegraders hogere scores voor woordkennis (.39) en voor grammatica (.48), maar lijken iets minder effectief op het terrein van de leesvaardigheid (.20). Een inhoudelijke interpretatie is moeilijk te geven. Toch handhaven we de regressiegewichten in het model omdat we de verschillen tussen beide leergangen willen corrigeren voor verschillen in de samenstelling van de groepen docenten.

De *kwaliteit van de docent*, zoals ervaren door de leerlingen, draagt bij tot een betere voorspelling van de leeruitkomsten op affectief gebied. Leerlingen van de als 'aardig' en 'vakbekwaam' ervaren docenten onderscheiden zich van andere leerlingen door een positievere beleving van het vak Duits (.40) en een hogere waardering van het lesmateriaal (.36 voor KD en .47 voor KO). Voor prestaties na één jaar onderwijs Duits is de ervaren kwaliteit van de docent van ondergeschikt belang.

De *didactische oriëntatie* van de docent blijkt bij drie van de negen effectmaten van belang (onder constanthouding van de overige

variabelen). Twee ervan zijn het vermelde waard. KD-docenten die meer affiniteit hebben met een communicatieve aanpak van het Duits realiseren lagere prestaties bij schrijven (.20) en grammatica (.20) dan hun KD-collega's die meer taalkundig gericht zijn.

Voor het *aantal hoofdstukken* dat een docent aan het einde van het schooljaar behandeld heeft, kan geen relatie met de cognitieve effectmaten aangetoond worden. Docenten die een relatief groot deel van het boek behandelen, behalen geen lagere of hogere leerresultaten dan docenten die tegen het einde van het schooljaar minder ver zijn gekomen. Ook voor de waardering van de leergang is de proportie behandelde leerstof van ondergeschikt belang.

Eerder in dit artikel constateerden wij dat het effectiviteitsverschil tussen KD en KO deels afhangt van de uitgevoerde correctie. Na correctie voor verschillen tussen beide leergangen in uitgangssituatie en randvoorwaarden was het verschil tussen KD en KO bij woordkennis, grammatica en spelling groter dan ervoor. Nadere analyse wijst uit dat dit verschijnsel niet zozeer kan worden toegeschreven aan een differentieel effect van de begin-, leerling- of klaskenmerken als wel aan de docentkenmerken bevoegdheid, ervaring en didactische oriëntatie. Het effect van deze docentkenmerken, die ongelijk over beide condities verdeeld zijn (zie Bijlage 1), verschilt van leergang tot leergang. Bij vakbeleving bleek het significante brutoverschil na correctie te verdwijnen. Nadere analyse leert dat dit vooral voor rekening komt van de vakbeleving aan het begin van het schooljaar.

### 5.3 Leergang en verschillen tussen klassen

In paragraaf 5.2 is de invloed van de leergang en de andere verklarende variabelen op de leeruitkomsten beschreven in termen van het gemiddeld niveau van leerlingen. Er wordt echter geen eenduidig antwoord gegeven op de vraag

naar het effect op de verschillen tussen *klassen* in gemiddeld cognitief of affectief niveau. Daarover wordt in deze paragraaf gerapporteerd.

#### *Verschillen tussen klassen*

Verschillen tussen klassen in gemiddeld prestatieniveau kunnen voor de leerprestaties van leerlingen van groot praktisch belang zijn (Purkey & Smith, 1982; 1983; Scheerens, 1989). Vooraf daarom eerst een belangrijke vraag. In hoeverre bestaan er verschillen tussen klassen in het gerealiseerde eindniveau na één jaar onderwijs in het Duits? De percentages tussenklassenvariantie, zoals geschat in het onconditionele model (model 0), zijn weergegeven in Tabel 3.

Uit Tabel 3 blijkt dat de 2-Mavoklassen systematisch van elkaar verschillen. Dit geldt zowel voor de zeven taalvaardigheidstoetsen als voor de twee affectieve effectmaten. Het cognitieve en affectieve niveau dat bereikt wordt aan het einde van het eerste leerjaar Duits verschilt van klas tot klas. In Tabel 3 is verder te zien dat de verschillen tussen klassen niet voor ieder instrument even groot zijn. Voor woordkennis en schrijven zijn deze bijvoorbeeld aanzienlijk groter dan voor spelling en lezen. Gezien het feit dat het gaat om categoriale Mavo's - een homogene onderzoeksgroep - is de hoogte soms opmerkelijk. Zeker in vergelijking met hetgeen uit ander taalvaardigheidsonderzoek bekend is (o.a. Blok & Eiting, 1988; Kuhlemeier & Van den Bergh, 1989).

#### *Begin-, leerling-, klas- en schoolkenmerken (model 2)*

Het percentage tussen-klassenvariantie dat verklaard wordt door verschillen in uitgangssituatie en randvoorwaarden (model 2) is weergegeven in de eerste rij van Tabel 4.

Bij alle negen effectmaten leveren de begin-, leerling-, klas- en schoolkenmerken een bijdrage (zie Tabel 4). Het niveau dat de klas na

**Tabel 3**  
Percentage tussen-klassenvariantie voor de negen effectmaten, zoals geschat in het onconditionele model (LE: lezen; WB: woordbegrip; LU: luisteren; SV: schrijven; WK: woordkennis; GR: grammatica; SP: spelling; BE: vakbeleving; WL: waardering lesmateriaal)

LE	WB	LU	SV	WK	GR	SP	BE	WL
8	16	13	31	25	10	7	14	28

**Tabel 4**

Toename van het percentage verklaarde tussen-klassenvariantie  $R^2_{\text{klas}}$  ten gevolge van de begin-, leerling-, klas- en schoolkenmerken (model 2 versus 0), de docent- en implementatiekenmerken (model 3 versus 2) en de leergang (model 4 versus 3); tussen haakjes: toename percentage verklaarde totale variantie; LE: lezen; WB: woordbegrip; LU: luisteren; SV: schrijven; WK: woordkennis; GR: grammatica; SP: spelling; BE: vakbeleving; WL: waardering lesmateriaal)

	LE	WB	LU	SV	WK	GR	SP	BE	WL
Begin-, leerling-, klas- en schoolkenmerken	34 (21)	24 (32)	30 (31)	39 (18)	15 (25)	18 (12)	10 (6)	26 (31)	4 (17)
Docent- en implementatiekenmerken	2 (0)	0 (0)	17 (1)	30 (10)	2 (3)	5 (1)	16 (2)	56 (14)	32 (15)
Leergang	0 (0)	0 (0)	0 (0)	7 (3)	13 (4)	21 (3)	11 (2)	0 (0)	42 (11)

één jaar onderwijs bereikt, wordt dus mede bepaald door de uitgangspositie aan het begin van het schooljaar, de samenstelling van de groep leerlingen en kenmerken van de school. Het effect is het grootst voor schrijven (39%), lezen (34%), luisteren (30%), de beleving van het vak Duits (26%) en woordbegrip (24%). Het percentage NIET-verklaarde tussen-klassenvariantie  $[(1 - R^2) * 100]$  varieert van 61 procent voor schrijven tot 96 procent voor de waardering van het lesmateriaal. Ook na correctie voor het effect van de begin-, leerling-, klas- en schoolkenmerken (die voor een docent moeilijk te beïnvloeden zijn) blijft er doorgaans nog een aanzienlijke hoeveelheid onverklaarde variantie over. Alleen deze resterende variantie kan – gegeven ons model – door docent en leergang verklaard worden.

### Docent- en implementatiekenmerken (model 3)

Te verwachten valt dat de verschillen tussen klassen in gerealiseerd eindniveau mede afhankelijk zijn van kenmerken van degene die voor de klas staat. Nu gaat het dus om de vraag wat de docent toevoegt aan de leeruitkomsten, boven hetgeen kan worden toegeschreven aan de uitgangspositie aan het begin van het schooljaar, de samenstelling van de groep leerlingen en kenmerken van de school. De tweede rij van Tabel 4 geeft inzicht in het additieve effect van de docent ( $R^2_{\text{model3}}$  minus  $R^2_{\text{model2}}$ ). De invloed van de docent is bij acht van de negen effectmaten traceerbaar. Ook hier geldt dat het additieve effect niet in alle gevallen even groot is. De invloed van de docent is het grootst bij de beleving

van het vak Duits, waar de verklaarde tussen-klassenvariantie stijgt van 26 naar 82 procent. Ook bij luisteren, grammatica, spelling en de waardering van het lesmateriaal doet het er toe welke docent er voor de klas staat, terwijl het effect bij lezen en woordkennis verwaarloosbaar klein is. Dat er voor woordbegrip geen docenteffect aantoonbaar is, komt niet als een volslagen verrassing. Het afleiden van de betekenis van Duitse woorden uit de context wordt immers niet zo expliciet onderwezen.

Ook na uitschakeling van het effect van alle verklarende variabelen behalve de leergang resteert er nog een aanzienlijke hoeveelheid variantie die nog niet verklaard is. Het percentage NIET-verklaarde tussen-klassenvariantie  $[(1 - R^2) * 100]$  uit model 3 varieert namelijk van 18 procent voor de beleving van het vak Duits tot 83 procent voor woordkennis. Dit is de variantie die in beginsel door de leergang verklaard zou kunnen worden.

### Leergang (model 4)

In hoeverre is de leergang van belang voor de leeruitkomsten? De toename van de verklaarde tussen-klassenvariantie ten gevolge van de leergang ( $R^2_{\text{model4}}$  minus  $R^2_{\text{model3}}$ ) is weergegeven in de derde rij van Tabel 4. Daaruit blijkt de leergang er bij vijf van de negen effectmaten toe doet. Weer geldt dat het additieve effect niet overal gelijk is. Bij de waardering van het lesmateriaal bedraagt de toename maar liefst 42 procent (van 36% naar 78%). Maar ook bij grammatica (21%), woordkennis (13%), spelling (11%) en schrijven (7%) lijkt de gehanteerde leergang van belang voor het niveau dat

de klas na één jaar onderwijs behaalt.

Een relativerende kanttekening is hier op zijn plaats. Zoals gezegd is het additieve effect van de leergang geëvalueerd als de toename van de verklaarde tussen-klassenvariantie na uitzuivering van het effect van de begin-, leerling-, klas-, school- en docentkenmerken. Hoewel theoretisch juist, geeft deze procedure misschien een te optimistisch beeld van het effect van de leergang in de onderwijspraktijk. Een voorbeeld mag dit duidelijk maken. Zoals gezegd heeft de leergang - gezien de toename van de verklaarde tussen-klassenvariantie - een aanzienlijk additief effect op de grammaticaprestaties ( $R^2_{\text{model4}}$  minus  $R^2_{\text{model3}} = 21$ ). Maar zoals eveneens in Tabel 4 is weergegeven, bedraagt de overeenkomstige toename van de totale variantie (d.w.z. tussen- en binnenklassenvariantie gezamenlijk) slechts drie procent. Aangezien de leergang alleen van invloed is op de tussen-klassenvarianties (aptitude-treatment-interacties daargelaten) kan deze hooguit tien procent van de totale variantie in de grammaticaprestaties naar zich toe trekken [de tussen-klassenvariantie bedraagt immers slechts tien procent van de totale variantie (zie Tabel 3)]. Alleen in het geval van de waardering van het lesmateriaal voegt de leergang meer dan tien procent aan de verklaarde totale variantie toe, maar daar kan de leergang in theorie maar liefst 28 procent van de totale variantie binden (zie Tabel 3).

## 6 Conclusies en discussie

De belangrijkste conclusie is dat de leergang 'ertoe doet'. Na correctie voor verschillen in uitgangspositie, randvoorwaarden en docentkenmerken blijkt een duidelijk verschil tussen KD en KO aantoonbaar bij schrijven, woordkennis, grammatica, spelling en de waardering van de leergang. Dit rechtvaardigt de conclusie dat het cognitieve en affectieve niveau van 2-Mavo-klassen na één jaar onderwijs in het Duits mede afhangt van de gebruikte leergang<sup>6</sup>.

Opmerkelijk is het differentiële karakter van het geconstateerde effectiviteitsverschil. Het is namelijk niet zo dat de ene leergang over de hele linie betere leeruitkomsten te zien geeft

dan de andere. Overeenkomstig onze eerste hypothese hebben de KD-klassen een voorsprong op de KO-klassen waar het de grammaticaprestaties betreft. Dit lijkt goed te verklaren vanuit de grotere aandacht voor dit vakonderdeel in KD.

Overeenkomstig de tweede hypothese wordt KO door de leerlingen duidelijk hoger gewaardeerd dan KD. Dit is waarschijnlijk toe te schrijven aan de speelsere vormgeving van KO, het gebruik van authentieke teksten, communicatieve oefentaken en de toepassing van spelvormen.

Conform onze derde hypothese constateren wij geen substantiële verschillen in leeruitkomsten op de onderdelen luisteren, lezen, woordbegrip en vakbeleving. Dat de KO-klassen het beter doen op de vakonderdelen spelling en woordkennis is niet door ons voorspeld. Toch lijkt deze bevinding, achteraf gezien, niet in strijd met de intuïtie. In KD heeft immers het leren van de traditionele deelvaardigheden veel gewicht.

Ten aanzien van schrijfvaardigheid waren er twee, onderling tegenstrijdige hypothesen. Ten eerste zouden de KD-leerlingen een voorsprong kunnen hebben op het terrein van de kwaliteit van de schrijfproducten (vanwege de nadruk op grammatica in KD). Ten tweede zouden de KO-leerlingen het beter kunnen doen waar het de inhoudelijke en structurele kwaliteit van de schrijfproducten betreft (vanwege de nadruk op productief schrijven in KO). Voor deze tweede hypothese is in het onderzoek enige ondersteuning gevonden. De KO-klassen blijken namelijk hogere schrijfprestaties te behalen dan de KD-klassen. Hiervoor is een plausibele verklaring te geven. De functionele toetsvorm - een briefje schrijven aan een Duitse leeftijdsgenoot - wordt namelijk alleen in KO geoefend. Kortom: het patroon van uitkomsten weerspiegelt overeenkomsten en verschillen tussen de leergangen in uitgangspunten, leerstofaanbod en didactiek (vgl. Walker & Schaffarzick, 1974).

De conclusies van dit onderzoek lijken in strijd met ander Nederlands leergangvergelijkend onderzoek, waarin geen noemenswaardige effectiviteitsverschillen tussen leergangen aangetoond konden worden (o.a. Harskamp & Suhre, 1986; Batenburg, Van den Berg & Ede-

lenbos, 1986; Meyer, De Leeuw, Groen & Perrenet, 1987; Van den Berg, 1987; Van den Berg & Harskamp, 1985; Edelenbos, 1990; Creemers, 1992). Een mogelijke verklaring voor deze discrepantie is wellicht dat Duits een relatief instructiegevoelig vak is waar leerlingen buiten de school doorgaans weinig mee in aanraking komen. Bovendien, anders dan bij veel ander leergangvergelijkend onderzoek het geval is, betreft het hier een effectiviteitsverschil ten gevolge van slechts één jaar onderwijs. Ook zouden de inhoudelijk-didactische verschillen tussen KD en KO groter kunnen zijn dan tussen veel van de leergangen uit de hiervoor aangehaalde onderzoeken. Tussen leergangen die slechts op marginale punten van elkaar verschillen, kan men nu eenmaal geen grote effectiviteitsverschillen verwachten (Chall, 1983; Walker & Schaffarzick, 1974). Tot slot is de effectmeting in ons onderzoek gedifferentieerder geweest dan veelal het geval is. Hadden wij bijvoorbeeld volstaan met lezen, woordbegrip, luisteren en vakbeleving, dan hadden wij geen effectiviteitsverschillen geconstateerd.

Het bereikte niveau na één jaar onderwijs blijkt mede bepaald te worden door de onderwijservaring, de bevoegdheid en de kwaliteit van de docent zoals ervaren door de leerlingen. Zo is de voorsprong van KD op KO bij woordkennis en grammatica niet onafhankelijk van de onderwijsbevoegdheid. KD-docenten realiseren hogere woordkennis- en grammacaprestaties dan hun KO-collega's, maar er bestaan ook verschillen binnen de groep KD-docenten: KD-eerstegraders presteren hoger dan KD-tweede-graders; in de groep KO-docenten is er voor de vakonderdelen woordkennis en grammatica geen effect van bevoegdheid aantoonbaar. Bij schrijven en grammatica blijkt de meer grammaticale leergang KD in handen van communicatief ingestelde docenten minder effectief dan in handen van KD-docenten die meer taalkundig georiënteerd zijn. Wellicht is het gebruik van een traditionele leergang voor communicatieve doeleinden minder geschikt. Of behaalt een docent met een leergang die hem of haar 'op het lijf geschreven is' betere leerresultaten dan iemand die er minder affiniteit mee heeft? Voorts is de samenhang tussen de waardering van het lesmateriaal en de kwaliteit van de docent bij KO (nog) sterker dan bij

KD. Kortom: het is strikt genomen niet geheel juist te spreken van *de* effectiviteit van *de* leergang. Het gaat er ook om *wie* de leergang hanteert.

Het verschil in effectiviteit tussen beide leergangen hangt niet of nauwelijks samen met de gemeten begin-, leerling- en klaskenmerken. Slechts twee interacties zijn het vermelden waard. Zittenblijven heeft alleen bij KO een positief additief effect op de prestaties voor het vakonderdeel woordkennis. De voorsprong van KO waar het de waardering van het lesmateriaal betreft, is voor de allochtone leerlingen extra groot. Kortom: als de ene leergang effectiever is dan de andere, dan geldt dit doorgaans voor alle typen leerlingen en klassen in gelijke mate. Het effect van de begin-, leerling- en klaskenmerken verschilt meestal niet van leergang tot leergang. Van differentiële effectiviteit voor verschillende groepen leerlingen en klassen lijkt derhalve nauwelijks sprake.

Opvallend is dat klassen die vier lessen Duits per week hebben geen betere prestaties leveren dan de klassen met drie lessen. Dit is één van de meest onverwachte bevindingen van het onderzoek. Volgens de logboekgegevens behandelen de docenten van de 4-uursvariant wel meer onderdelen (leesteksten, oefeningen e.d.) dan hun collega's van de 3-uursvariant, maar het rendement van deze meerinvestering blijft geheel uit, het komt tenminste niet tot uiting in de leeruitkomsten. Naar een verklaring kan men slechts gissen. Wellicht wordt de tijd in de 4-uursvariant wat minder effectief besteed? Of krijgen de leerlingen in de 3-uursvariant ter compensatie meer huiswerk op?

### **Besluit**

In een recent SVO-review over instructie-effectiviteit (Creemers, 1992) is vastgesteld dat de gehanteerde leergang van groot belang is voor de leerstofkeuze en de didactische vormgeving van het onderwijs. Het belang voor de leerresultaten van de leerlingen zou gering zijn. De bevindingen van de onderhavige studie - de ene leergang genereert een ander patroon van leeruitkomsten dan de andere - geven reden tot enig optimisme. Verder vormen de uitkomsten een steun in de rug voor degenen die menen dat communicatief en leerlinggericht vreemde-

talenonderwijs in de basisvorming vanuit het oogpunt van rendement verantwoord is. Het in dit onderzoek gevonden patroon van uitkomsten geeft wel aanleiding tot de veronderstelling dat de gepropageerde accentverschuiving naar de integrale, communicatieve vaardigheden ten koste kan gaan van de beheersing van de traditionele deelvaardigheden.

## Noten

- 1 Van het eindrapport 'Vergelijking van enkele leergangen Duits voor de tweede klas' van L. Melse, F. Teunisse, M. Ouborg en J.B. Kuhlmeier zijn nog enkele exemplaren (gratis) te bestellen bij het Instituut voor Toetsontwikkeling (telefoon 085-521261)
- 2 Een opmerking over de interpretatie van de regressiegewichten is hier op zijn plaats. Merk op dat het hier coëfficiënten in een multivariate samenhang betreft. Deze geven aan welke bijdrage de verklarende variabelen in onderlinge samenhang leveren aan de voorspelling van de leeruitkomsten. De grootte van een regressiegewicht geeft inzicht in het effect op de leeruitkomsten onder constanthouding van alle andere verklarende variabelen in de analyse. Een niet-significant regressiegewicht hoeft niet altijd op het ontbreken van een univariate samenhang (b.v. correlatie of effectgrootte) te duiden. Als bijvoorbeeld etniciteit voor de leerresultaten van groter belang is dan sociaal milieu, zou de oorspronkelijke correlatie tussen milieu en prestaties kunnen verdwijnen op het moment dat deze gecorrigeerd wordt voor de invloed van etniciteit. Ook het omgekeerde is mogelijk: een effect verschijnt pas na correctie voor andere variabelen in het model. Zo zou het effect van de leergang pas op kunnen treden nadat gecorrigeerd is voor een mogelijk ongelijke uitgangssituatie aan het begin van het schooljaar. Met andere woorden: regressiegewichten kunnen niet zonder meer univariaat geïnterpreteerd worden.
- 3 Het percentage verklaarde variantie ( $R^2$ ) is het kwadraat van de correlatie tussen de geobserveerde scores en de op grond van het model voorspelde scores.
- 4 Melse e.a. (1992) gebruiken een andere maat voor het evalueren van de grootte van de effecten, namelijk de daling van de geobserveerde tussen-klassenvariantie. Dit leidt tot andere per-

centages verklaarde variantie dan de in dit artikel gerapporteerde percentages. Het patroon van uitkomsten blijft evenwel gelijk.

- 5 De regressiegewichten van de verklarende variabelen zijn geschat als afwijking van hun gemiddelden (vgl. Prosser, Rasbash & Goldstein, 1991, p. 75). Het verschil tussen de beide regressiegewichten voor KD en KO representeert derhalve het verschil tussen beide leergangen in gemiddeld niveau voor de gemiddelde leerling, docent en school.
- 6 Bij deze conclusie is een relativerende opmerking op zijn plaats. In leergangvergelijkend onderzoek kunnen vele variabelen de uitkomsten beïnvloeden. Volledige zekerheid of deze alle (volledig) onder controle zijn gebracht dan wel geen effect hebben gehad op de leeruitkomsten is er niet. Daarom past een zeker voorbehoud bij de uitkomsten van de vergelijking. Opmerkelijk lijkt in dit verband dat een leergangeffect niet optreedt bij de vakonderdelen waar een correctie voor verschillen in beginniveau op het desbetreffende vakonderdeel is uitgevoerd (lezen, woordbegrip en vakbeleving). Misschien dat de gevonden leergangeffecten bij schrijven, grammatica en spelling zouden verdwijnen wanneer ook voor deze vakonderdelen een indicatie van het beginniveau was verkregen. Daarbij moet opgemerkt worden dat het natuurlijk ten enenmale onmogelijk is om het beginniveau van bijvoorbeeld de schrijfvaardigheid Duits vast te stellen; aan het begin van de tweede klas beheersen de leerlingen deze vaardigheid immers nog niet. Wel zou het mogelijk zijn geweest de leerlingen aan het begin van het schooljaar een schrijftoets Nederlands voor te leggen en na te gaan of de KD- en KO- klassen wat dit betreft verschillen. Een dergelijk verschil is evenwel niet erg aannemelijk aangezien er op de voormeting geen verschillen tussen de KD- en KO-klassen aangetoond konden worden, noch bij lezen noch bij woordbegrip.

## Literatuur

- Batenburg, Th. A. van, Berg, G. van den, & Edelenbos, P. (1986). *Vergelijking van taalmethoden in het basisonderwijs*. Eindrapport deel I: Rapportage onderzoeksresultaten. Groningen: RION.
- Berg, G. van den (1987). *Effectief evalueren. Een empirische studie naar de doelmatigheid van aan-*



- wijzingen voor de evaluatiepraktijk. Lisse: Swets & Zeitlinger.
- Berg, G. van den, & Harskamp, E. G. (1985). *Twee onderwijsleerpakketten economie voor Mavo vergeleken: onderzoeksverslag SVO-project 1151 (een vooronderzoek)*. Groningen: RION.
- Björneberg, B. (1974). *A follow-up study in teaching foreign language grammar*. Gothenburg: University of Gothenburg.
- Blok, H., & Eiting, M. H. (1988). De grootte van schooleffecten: Hoe verschillend presteren leerlingen van verschillende scholen? *Tijdschrift voor Onderwijsresearch*, 13, 16-30.
- Bosker, R. J. (1991). De consistentie van schooleffecten in het basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 16, 206-218.
- Boogaards, P. (1982). *Moderne vreemde talen op school: een studie betreffende de determinanten van succes bij het leren van Frans in de brugklas*. 's-Gravenhage: SVO.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: application and data analysis methods*. Newsbury Park: Sage Publications.
- Chall, J. S. (1983). *Learning to read: the great debate* (updated edition). New York: McGraw-Hill Book Company.
- Creemers, B. P. M. (1992). *Effectieve instructie: een empirische bijdrage aan de verbetering van het onderwijs in de klas*. 's-Gravenhage: Instituut voor Onderzoek van het Onderwijs.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Edelenbos, P. (1990). *Leergangen voor Engels in het basisonderwijs vergeleken*. Groningen: RION.
- Elek, T. von, & Oskarsson, M. (1975). *Comparative method experiments in foreign language teaching, the final report of the GUME/Adults Project*. Gothenburg: University of Gothenburg.
- Els, T. van, Bongaerts, T., Extra, G., Os, C. van, & Jansen-Van Dielen, A. M. (1984). *Applied linguistics and the Learning and Teaching of Foreign Languages*. London: Edward Arnold.
- Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Synthesis of educational productivity research. *International Journal of Educational Research*, 11, 145-252.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin.
- Good, T., & Brophy, J. (1986). School effects. In M. Wittrock (Ed.). *Handbook of research on teaching*. Third edition (pp. 570-602). New York: McMillan.
- Gravemeijer, K. P. E. (1989). Methodegebruik, een ongewisse factor in de vernieuwing van het reken-wiskundeonderwijs. In A. Reints, R. A. de Jong & N. A. J. Lagerweij (Red.). *Om de kwaliteit van het leermiddel* (pp. 138-150). Tilburg: Zwijssen.
- Harskamp, E. G., & Suhre, C. J. M. (1986). *Vergelijking van rekenmethoden in het basisonderwijs*. Eindrapport. Groningen: RION.
- Horn, W. (1969). *Prüfsystem für Schul- und Bildungsberatung P-S-B*. Göttingen: Verlag für Psychologie Dr. C. J. Hogrefe.
- Jacobs-Hessing, E. G. (1989). *Leerlingverschillen en onderwijsaanbod*. Utrecht: Uitgeverij w.c.c.
- Koster, C. J. & Matter, J. F. (1983). *Vreemde talen leren en onderwijzen*. Groningen: Wolters-Noordhoff.
- Kuhlemeier, J. B. (1989). *Inventarisatie van leergangen in Mavo, Havo en Vwo. Schooljaar 1989-1990*. Arnhem: Cito.
- Kuhlemeier, J. B., & Bergh, H. van den (1989). *De Proefpeiling Nederlands: een onderzoek naar de haalbaarheid van peilingsonderzoek in het voortgezet onderwijs*. Bijlage bij Specialistisch Bulletin Nr. 74. Arnhem: Cito.
- Kuhlemeier, J. B., Bergh, H. van den, & Teunisse, F. (1990). Interne structuur en constructvaliditeit van belevingsschalen voor Wiskunde en Engels. *Tijdschrift voor Onderwijsresearch*, 15, 110-122.
- Lugthart, E., Roeders, P. J. B., Bosker, R. J., & Bos, K. T. (1989). *Effectieve schoolkenmerken in het voortgezet onderwijs*. Deel 1: Een literatuuroverzicht. Groningen: RION.
- Melse, L., Teunisse, F., Ouborg, M., & Kuhlemeier, J. B. (1992). *Vergelijking van enkele leergangen Duits voor de tweede klas*. Arnhem: Cito.
- Meuffels, B. (1986). Globaal versus analytisch beoordelen. *Moer*, 1, 9-14.
- Meyer, J., Leeuw, J. de, Groen, E., & Perrenet, Chr. (1987). Vergelijking van effecten van wiskundemethoden. In E. Harskamp & W. J. Nijhoff (Red.). *Curriculum: ontwerp, implementatie en evaluatie* (pp. 157-170). Lisse: Swets & Zeitlinger.
- Ministerie van Onderwijs en Wetenschappen (1993). *Besluit kerndoelen en adviësurentabel basisvorming 1993-1998*. Zoetermeer: Ministerie van Onderwijs en Wetenschappen.
- Prosser, R., Rasbash, J., & Goldstein, H. (1991). *Software for three-level analysis. Users' guide for V.2*. London: Institute of Education.

Purkey, S. C., & Smith, M. S. (1982). To soon to cheer?: synthesis of research on effective schools. *Educational Leadership*, 40, 64-69.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: a review. *Elementary School Journal*, 83, 427-452.

Raudenbush, S. W., & Bryk, A. (1988). Methodological advances in studying effects of schools and classrooms on student learning. In E. Rothkopf (Ed.). *Review of research in education*. Vol. 15 (pp. 423-475). Washington, D.C.: American Educational Research Association.

Reints, A. (1989). Over het gebruik van methodes door leerkrachten. In A. Reints, R. A. de Jong & N. A. J. Lagerweij (Red.). *Om de kwaliteit van het leermiddel* (pp. 127-137). Tilburg: Zwijssen.

Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvattingen en analyses van onderzoeksresultaten*. 's-Gravenhage: SVO.

Smith, Ph. (1970). *A comparison of the cognitive and audiolingual approaches to foreign language instruction: the Pennsylvania foreign language project*. Philadelphia: Center for Curriculum Development.

Teunisse, F. (1990). *De beleving van Engels. Verslag van een vooronderzoek naar de psychometrische en onderwijskundige kwaliteit van de Belevingschaal voor Engels*. Arnhem: Cito.

Walker, D. F., & Schaffarzick, J. (1974). Comparing curricula. *Review of Educational Research*, 44, 83-111.

Willms, J. D., & Raudenbush, S. W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, 26, 209-232.

Manuscript aanvaard 10-1-1994

## Auteurs

**J. B. Kuhlemeier** (1952) is als onderwijskundig onderzoeker werkzaam bij het Instituut voor Toetsontwikkeling te Arnhem.

Adres: CITO, bedrijfspgroep PPON, Postbus 1034, 6801 MG Arnhem.

**L. Melse** (1944) is als vakmedewerker Duits en onderzoeker verbonden aan het Instituut voor Toetsontwikkeling te Arnhem.

Adres: CITO, bedrijfspgroep Talen, Postbus 1034, 6801 MG Arnhem.

**H. van den Bergh** (1957) is als docent en onderzoeker verbonden aan de afdeling Taalbeheersing van de Rijksuniversiteit Utrecht.

Adres: Rijksuniversiteit Utrecht, Centre for Language and Communication, Trans 10, 3512 JK Utrecht.

## Abstract

### **Effects of a grammar-based and a communicative curriculum on cognitive and affective learning outcomes**

J. B. Kuhlemeier, L. Melse & H. van den Bergh. *Pedagogische Studiën*, 1994, 71, 284-303.

This article reports on a comparative curriculum evaluation study conducted in the second year of intermediate general education (Mavo). The purpose was to measure the impact of two curricula for learning German as a foreign language on learning outcomes. In a pretest-posttest design the effects of a communicative and a grammar-based curriculum on cognitive and affective learning outcomes were measured. Multilevel analyses revealed significant and substantial differences (while controlling statistically for the differential effects of student, class, school and teacher characteristics). Students following the communicative course did better on communicative writing and were more satisfied with their curriculum materials; students following the grammar-based course achieved higher scores on spelling, grammar and vocabulary. No differences were found on reading, word guessing, listening and the attitude towards German. In addition, the impact of the curricula on learning outcomes appeared to be rather stable across different types of students (e.g. entry achievement, entry attitude towards German, intelligence, staying down a class, ethnicity).

# Bijlage 1

## Constructie van de leerling-, klas-, school- en docentvariabelen

### *Leerling- en klaskenmerken*

Voor doubleren en etniciteit zijn dummies geconstrueerd met de waarden 0 voor een leerling die in de tweede klas niet is blijven zitten en 1 voor een doubleur, respectievelijk 0 indien de leerling thuis Nederlands, Duits, Frans of Engels spreekt en 1 indien er thuis een andere taal gesproken wordt. De variabelen 'proportie doubleurs' en 'proportie allochtonen' zijn verkregen door middel van aggregatie van de individuele leerlingkenmerken naar het niveau van de klas. De proportie doubleurs varieert van .00 tot .38 met een gemiddelde van .05; de proportie allochtonen van .00 tot .43 met een gemiddelde van .15.

### *School-, implementatie- en docentkenmerken*

Voor de institutionele leertijd is een dummy gebruikt, met als waarden 0 indien de leerlingen drie lessen Duits krijgen en 1 indien zij vier lessen per week ontvangen. Van KD volgde 44 procent van de leerlingen de 4-uursvariant en van KO 43 procent.

De schoolgrootte is ontleend aan een bestand van het Ministerie van Onderwijs en Wetenschappen en varieert in de KD-conditie van 145 tot 460, met een gemiddelde van 286; in de KO-conditie van 120 tot 446, met een gemiddelde van 302.

De variabele 'proportie behandelde leerstof' is geconstrueerd door de ratio te nemen van het totaal aantal behandelde hoofdstukken en het totaal aantal hoofdstukken van de desbetreffende leergang. Van KD behandelen de docenten gemiddeld 80 procent, terwijl hun collega's van KO gemiddeld 81 procent van hun leergang behandelen.

De onderwijservaring van de docent, gemeten op een vier-puntsschaal, is tamelijk ongelijk verdeeld: de KO-docenten beschikken over meer ervaring dan de KD-docenten. Voor de KD-klassen bedraagt de gemiddelde ervaring 2.3 (*sd*: .7) en in de KO-klassen 3.2 (*sd*: .9) op de vier-puntsschaal.

De bevoegdheid van de docent is gemeten op een driepuntsschaal, met de waarde 1 voor eerste-graders, 2 voor tweedegraders en 3 voor derdegraders. De bevoegdheid is ongelijk verdeeld over de beide leergangen. Voor de KD-klassen is de gemiddelde bevoegdheid 1.7 (*sd*: .5) en voor KO 2.3 (*sd*: .7).

De kwaliteit van de docent als persoon en als vakdidacticus is gemeten met behulp van een leerlingenvragenlijst bestaande uit 13 uitspraken van het Likerttype. De gemiddelden voor de KD- en KO-klassen bedragen respectievelijk 36.3 (*sd*: 4.3) en 36.8 (*sd*: 7.7).

De docenten is ook gevraagd welke didactische benadering hen het meeste aanspreekt: de meer taalkundige (nadruk op verwerving van grammaticale structuren, woordkennis en een correcte uitspraak) of de meer communicatieve (zo spoedig mogelijk beginnen met luisteren en spreken in levensechte situaties en met het lezen en schrijven van eenvoudige tekstjes). Deze didactische oriëntatie is gemeten op een vijf-puntsschaal met als de uitersten 'taalkundige benadering' en 'communicatieve benadering'. Zoals mag worden verwacht, geven de KD-docenten in meerderheid de voorkeur aan de taalkundige benadering, terwijl de voorkeur van de KO-docenten juist meer tendeert in de communicatieve richting. De gemiddelden op de schaal zijn voor KD-klassen lager dan voor KO-klassen: 2.52 (*sd*: .9) versus 3.42 (*sd*: 1.2).

## Bijlage 2

De invloed van de begin-, leerling-, klas-, school- en docentkenmerken op de leeruitkomsten (per verklarende variabele zijn de regressiegewichten voor 'KD' en 'KO' in een afzonderlijke rij weergegeven; niet-significante regressiegewichten zijn uit de analyse verwijderd en derhalve niet gerapporteerd)

	LE	WB	LU	SV	WK	GR	SP	BE	WL
<i>Begin- en leerlingkenmerken</i>									
Lezen	.12	.08							
KD				.07					
KO			.18	-.08					
Woordbegrip	.38	.52	.45	.16	.43	.31	.25		
Intelligentie			.10						
KD								-.09	
KO								.44	.20
Beleving	.07			.09	.09	.07			
Doubleren		-.17							
KD									
KO					.27				
Etniciteit		-.24	-.45		.27				
KD									
KO									.34
<i>Klas- en schoolkenmerken</i>									
Prop. doubleurs									
KD									-.16
KO									
Prop. allochtonen			.08			.09			
Aantal lessen									
KD				-.37					
KO			-.31						
Schoolgrootte									
KD	.18								
KO									
<i>Docent en implementatie</i>									
Prop. hoofdst.									
KD									-.16
KO									
Ervaring						.14			
KD	-.14		-.11						-.06
KO							.23		
Bevoegdheid				-.24					
KD	.20				-.39	-.48			
KO									
Kwaliteit							.09	.40	
KD				.16	.11				.36
KO									.47
Did. oriëntatie									
KD				-.20		-.20			
KO							.12		