

Evaluatie en implementatie

J.H. Slavenburg, *Onderwijsstimulering en gezinsactivering*. Effecten van programma's van het project Onderwijs en Sociaal Milieu op de schoolprestaties van eersteklassers lager onderwijs. Berichten summatieve evaluatie project Onderwijs en Sociaal Milieu 86/7. Proefschrift. S.V.O., Den Haag, 1986, 312 pag., ISBN 90-6472-075-4.

K.M. STOKKING

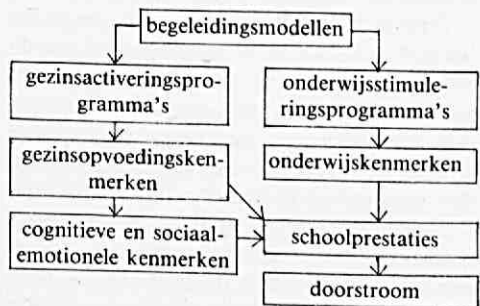
Vakgroep Onderwijskunde Rijksuniversiteit Utrecht

Het boek 'Onderwijsstimulering en gezinsactivering' is zowel het eerste eindverslag van het OSM-project als het proefschrift van de gene die een groot deel van de looptijd van dat project projectleider was. Het OSM-project is nog niet afgelopen. Het boek bevat primair een verslag van een summatief evaluatieonderzoek naar de effecten op de schoolprestaties aan het einde van klas 1 lagere school (nu groep 3 basisschool) van de OSM-programma's die voor de betreffende periode uit de schoolloopbaan (kleuterschool + eerste klas lagere school) zijn ontwikkeld.

Het boek begint met een hoofdstuk waarin doel, werkplan en werksituatie van het project Onderwijs en Sociaal Milieu worden samengevat. De aanzet tot het project is in eerste instantie gegeven door Grandia. In de jaren 1968-1973 bleken de doorstroomcijfers naar het A.V.O. van kinderen uit de volkswijken van Rotterdam (35%) duidelijk achter te blijven bij het landelijk gemiddelde over alle leerlingen (60%). OSM is opgezet om dit verschil weg te werken. Daartoe zijn vanaf 1974 op een planmatige, systematische en zoveel mogelijk wetenschappelijk verantwoorde wijze programma's ontwikkeld: onderwijsstimuleringsprogramma's voor de kleuterschool en alle klassen van de lagere school, gezinsactiveringsprogramma's, alsmede onderwijsbegeleidingsmodellen bij deze beide soorten programma's. Tevens zijn

in 1974 de uitgangspunten en criteria voor de evaluatie van de verschillende programma's vastgelegd. Centraal staat een zgn. verklaringmodel, waarin, gebaseerd op literatuuronderzoek, de belangrijkste vermoedelijke oorzaken van de milieuspecifieke doorstroom naar het voortgezet onderwijs zijn aangegeven, waaronder gezinsstructurele en gezinsopvoedingskenmerken, cognitieve en sociaal-emotionele kenmerken, onderwijskenmerken en schoolprestaties. De OSM-programma's dienden in te werken op de onderwijskenmerken en de gezinsopvoedingskenmerken. Het model is door mij globaal samengevat in Figuur 1.

Figuur 1 *Globaal onderzoeksmodel OSM-project*



Het OSM-project werkt(e) samen met 22 kleuterscholen en 24 lagere scholen. Het onderhavige verslag betreft zoals gezegd de periode tot en met klas 1 lagere school. Als evaluatiecriterium fungeert het behalen van tenminste het landelijk gemiddelde op genormeerde schoolprestatietoetsen aan het einde van klas 1. Dit criterium is afgeleid van het einddoel van OSM zoals hierboven ge-

noemd: het behalen van (tenminste) landelijk gemiddelde doorstroomcijfers naar het A.V.O.

Vervolgens wordt in een viertal hoofdstukken ingegaan op de opzet van het evaluatieonderzoek, de evaluatievragen, de evaluatiecriteria, de nadere specificatie van het verklaringsmodel en de te gebruiken meetinstrumenten en analyseprocedures.

Als evaluatievragen worden genoemd, samengevat: worden landelijk gemiddelde schoolprestaties in klas 1 gehaald? zo nee, waardoor? zo ja, komt dat door de OSM-programma's? zo ja, wat is de goetkoopste deelverzameling van deze programma's die voldoende is om het beoogde effect te bereiken? Toegevoegd worden als criteria de afname van de uitstroom naar het buitengewoon onderwijs en van het zittenblijven. Apart behandeld worden de begeleidingsmodellen (daarover hieronder meer).

Landelijk gemiddelde schoolprestaties worden 'transfereffectcriteria' genoemd. Per OSM-programma is nl. tevens voorzien in programmadoe-effectiviteitscriteria. Deze zijn programma-specifiek en zijn gehanteerd bij de formatieve evaluatie van de programma's. Daarnaast spelen 'implementatiecriteria' een belangrijke rol: nagegaan moet worden of de programma's voldoende geïmplementeerd zijn, omdat men er anders niet de effecten van kan bepalen. Als norm geldt hier dat 70% van de in het programma voorgeschreven activiteiten moet zijn uitgevoerd.

De opzet van het summatieve evaluatieonderzoek voorziet in de mogelijkheid om de effectiviteit van de programma's vast te stellen via een quasi-experiment, al dan niet met een controlegroep (van vergelijkbare, niet aan het OSM-project deelnemende scholen). Als meest adequaat wordt echter het werken met een verklaringsmodel gezien (een met statistische technieken te analyseren causaal model), o.a. omdat op die manier meer variabelen kunnen worden opgenomen en omdat overige evaluatievragen (indien geen effect, waardoor; indien wel, hoe goedkoper) daarmee tevens beter kunnen worden beantwoord.

In het grootste hoofdstuk van het boek wordt het te hanteren verklaringsmodel nader uitgewerkt, op basis van allerlei onderzoeksliteratuur m.b.t. onderwijskenmerken, leerlingkenmerken en gezinskenmerken. Ver-

geleken met het oorspronkelijke model zijn variabelen weggelaten, variabelen toegevoegd, relaties tussen variabelen weggelaten, relaties tussen variabelen toegevoegd, en relaties tussen variabelen en indicatoren (c.q. meetinstrumenten) gewijzigd.

In het vierde speciaal aan de evaluatieopzet gewijde hoofdstuk wordt een overzicht gegeven van alle ontwikkelde en gekozen meetinstrumenten en worden de te hanteren analyseprocedures besproken. De voorkeur gaat uit naar de Lisrel-programmatuur; men onderkent echter wel de strenge assumpties die daaraan ten grondslag liggen en het risico dat Lisrel daardoor niet bruikbaar kan blijken te zijn.

Ten slotte wordt in een tweetal hoofdstukken ingegaan op de resultaten. Deze resultaten worden eerst gegeven per OSM-programma, d.w.z. met betrekking tot de mate van implementatie en programmadoel-effectiviteit (deze laatste zoals opnieuw gemeten tijdens het summatief evaluatieonderzoek). Van de 7 betrokken onderwijsstimulerings- en gezinsactiveringsprogramma's bleken slechts 2 (onderwijsstimulerings-)programma's voldoende geïmplementeerd en programmadoe-effectief te zijn. De begeleidingsmodellen bleken slecht geïmplementeerd te zijn; wel was er een gematigd positief verband tussen de mate van begeleiding en de mate waarin de onderwijsstimulerings- en gezinsactiveringsprogramma's waren geïmplementeerd.

Het transfereffectcriterium werd niet gehaald; de gemiddelde prestaties op de beschikbare drie genormeerde leestoetsen van de ± 350 kinderen die tot en met de 1e klas lagere school gevolgd konden worden lagen onder het landelijk gemiddelde, en voor 2 van de 3 tevens onder het gemiddelde van de ± 200 kinderen op de 15 vergelijkings-scholen. Conform de evaluatieopzet werd vervolgens geprobeerd om via causale analyse na te gaan waar de oorzaken voor dit falen kunnen worden gevonden. Gebruik van Lisrel leidde, ondanks een aantal vereenvoudigingen en wijzigingen, schipbreuk. Met zgn. 'softmodellering' procedures werd achterhaald dat dominerende factoren lijken te zijn: intelligentie, cognitief ontwikkelingsniveau, werkhouding, beroep en opleiding ouders en het stimuleren van de cognitieve ontwikkeling van hun kinderen door de ouders.

De eindconclusie luidt dat het beïnvloeden van de cognitieve en sociaal-emotionele leerlingkenmerken en de gezinsopvoedingskenmerken door middel van daarop gerichte OSM-programma's onvoldoende is gelukt. In andere Nederlandse en in Amerikaanse projecten werden echter amper betere resultaten geboekt, zo blijkt uit een bespreking daarvan door de auteur. Veel programma's boeken wel korte-termijneffecten. Op minder-programmaspecifieke meetinstrumenten zijn deze echter, na korte of langere tijd, niet meer terug te vinden. De auteur besluit met een aantal aanbevelingen. De goed gestructureerde en doeleffectief gebleken OSM-programma's worden aanbevolen voor gebruik op scholen in onderwijsvoor-rangsgebieden. Voor het overige wordt vooral gewezen op de noodzaak van voortgezet onderzoek en ontwikkelingswerk. Het boek bevat in bijlage een opsomming van de belangrijkste kenmerken van de OSM-programma's, inclusief de verantwoording van het t.a.v. een aantal programma's genomen besluit om deze door een educatieve uitgever te laten verspreiden.

Het boek munt uit door een zeer overzichtelijke hoofdstukkenindeling. Kenmerkend is ook de verregaande explicitering van allerlei genomen besluiten. Er komen amper storende onduidelijkheden in voor – éénmaal spreken tekst en tabel elkaar tegen (p. 157) – en een gedeelte van ongeveer een pagina komt tweemaal in het boek voor (p. 54-55 en 99-100). Het ontbreken van typfouten is een genoegen, de regelmatig onduidelijke alinieë-ring niet: de tekst gaat soms bladzijden achter elkaar door en inhoudelijke cesuren vallen soms midden in een regel.

Het summatief evaluatieonderzoek van het OSM-project valt op door het vooraf aangeven van exacte evaluatiecriteria, het trekken van conclusies via complexe causale modellen, de aandacht voor de mate van implementatie van de programma's, en de negatieve eindconclusie. Bij deze vier aspecten staan we hieronder stil.

Slavenburg heeft geprobeerd om vooraf exacte evaluatiecriteria te formuleren. Dat gebeurde per programma, in de vorm van programmadoeleffectiviteitscriteria, en voor het project als geheel, in de vorm van landelijk gemiddelde schoolprestaties en

doorstroomcijfers zoals boven genoemd. Gedurende de looptijd van het OSM-project heeft de discussie over predictivistische versus meer op interpretatie achteraf gebaseerde evaluatieonderzoeksmethodologie zich voortgezet, in Nederland onder meer door Hofstee (1980) en Stokking (1984). We kunnen constateren dat in het OSM-project weliswaar exacte criteria zijn geformuleerd, maar op tal van punten toch achteraf interpretatie noodzakelijk bleek en soms water in de wijn werd gedaan. De auteur is daar overigens heel eerlijk over. Kennelijk is de praktijk toch sterker dan de leer. Bovendien ligt nader interpreteren van onderzoeksuitkomsten voor de hand als je er zoveel mogelijk van wilt leren.

Ten behoeve van dit laatste is gebruik gemaakt van de mogelijkheden tot constructie en analyse van causale modellen. De opeenvolgende versies van het bovengenoemde 'verklaringsmodel' vormen de rode draad door het boek (pp. 20, 25, 34, 118, 174, 176, 180, 183-187). Naar mijn smaak dragen alle veranderingen en vaak ad hoc aanpassingen (om technische redenen), hoe zorgvuldig ook gedocumenteerd, niet echt bij tot grotere duidelijkheid over de werkzame oorzakelijke mechanismen. Kennelijk zijn geavanceerde technieken als Lisrel gewoon niet zo geschikt in de levensechte onderzoekssituaties waarmee we in het onderhavige type onderwijs-onderzoek nu eenmaal te maken hebben. Overigens is het, op een iets fundamenteeler niveau bekeken, al moeilijk genoeg om de mogelijke causale samenhangen tussen drie variabelen plausibel te maken en inhoud te geven (zie bijvoorbeeld Tacq, 1984). Zo leidt o.a. het werken met niet-recursieve modellen (waarvan in OSM sprake was door het veronderstellen van een wederzijdse beïnvloeding tussen cognitieve en sociaal-emotionele leerlingkenmerken) tot behoorlijke conceptuele problemen. Terzijde dient opgemerkt dat men zich kan afvragen of het verklaringsmodel in de OSM-evaluatie wel is gebruikt voor het opsporen van mogelijke oorzaken van het (eerst eventuele, later geconstateerde) niet behalen van de doelstelling. Door via een causaal model tot de conclusie te komen dat factoren als intelligentie en werkhouding van de leerling meer variantie binden dan de mate van implementatie van de OSM-programma's is toch niet de oorzaak aange-

wezen van het feit dat deze programma's niet meer effect hadden?

Dit laatste leidt ons tot het derde kenmerkende aspect in de evaluatie van OSM, nl. de aandacht voor de implementatie van de programma's. Zeer terecht gaat de auteur ervan uit dat het bepalen van de mate waarin de programma's geïmplementeerd zijn een voorwaarde is om hun effecten te kunnen bepalen, immers: anders weet je niet waarvan je bezig bent de effecten te bepalen (p. 45, 54-55). Met andere woorden: alleen produkt-evaluatie is niet genoeg (vgl. Hofstee, 1982; Stokking, 1983). Hoewel gesteld wordt dat het niet zozeer de verklaringswaarde van de mate van implementatie voor het effect is waar het om gaat, maar meer het documenteren van het feitelijk uitgevoerde programma (p. 55), is het wel deze mate van implementatie die in de geanalyseerde causale modellen wordt meegenomen. (Overigens verwijst de auteur naar het enige hem bekende Nederlandse onderzoek waarin de relatie tussen mate van implementatie en schoolprestatie is nagegaan en geen verband werd gevonden; in het GEON-project bleek er sprake te zijn van een positief verband tussen de mate waarin de leerkrachten de cursussen hadden doorgewerkt en de hoogte van de IQ-winst bij de kinderen; zie Stokking en De Vries, 1981.) Ook op een andere manier is de auteur niet consequent op dit punt, nl. door te rapporteren over de programma-doeleffectiviteit van de programma's als gegeven naast de implementatie. Maar hoe kan een programma 'niet doeleffectief' worden genoemd als het 'onvoldoende geïmplementeerd' is? (zie p. 155 e.v.).

Daarmee komen we bij het laatste opvallende punt in dit onderzoeksverslag, nl. de negatieve eindconclusie. Allereerst moet worden opgemerkt dat het misschien een beetje ver gaat om de uitkomsten op een drietal leestoetsen de doorslag te laten geven (omdat op dat moment alleen voor het vak lezen genormeerde toetsen beschikbaar waren, p. 51). Het achterblijven bij het landelijk gemiddelde (p. 169) had bovendien wel gerelateerd mogen worden aan de achterstand die voorafgaande aan de inbreng van OSM op de scholen bestond. Het vasthouden aan het geformuleerde criterium is heel nobel, maar inzicht in wat OSM nu wél heeft bereikt wordt nu onvoldoende geboden. In

dit verband moet zeker worden ingegaan op de hierboven reeds genoemde mate van implementatie van de programma's. Het is toch wel heel opvallend dat bij de 7 geëvalueerde programma's sprake is van een perfect samengaan van de variabelen 'mate van implementatie' (onvoldoende/voldoende) en 'doeleffectiviteit' (nee/ja). Waarom dan niet uit de causale analyses, waaruit een positieve relatie tussen schoolprestaties (cq. leesprestaties) resp. cognitieve kenmerken en de mate van implementatie van het lees- resp. rekenprogramma naar voren komt, de conclusie getrokken dat een hoge mate van implementatie wel degelijk tot effecten leidt? Temeer waar op p. 189 uit een aanvullend uitgevoerde discriminantanalyse ineens wel de conclusie wordt getrokken dat een optimale tot maximale implementatie er kennelijk toe doet. Het is toch belangrijk om op zo'n resultaat verder te bouwen? Wellicht was er t.a.v. verschillende programma's sprake van een 'restriction of range', konden zij niet effectief zijn doordat hun implementatiewaarde te laag bleef. Bovendien had kunnen worden nagegaan waardoor de mate van implementatie van één bepaald programma nu juist negatief samenhangt met het criterium (p. 187), en had in de analyses rekening kunnen worden gehouden met de mate waarin de *leerlingen* feitelijk in de programma's hebben geparticipeerd (zoals ook in de GEON-evaluatie is gedaan; zie Stokking en De Vries), in plaats van alleen te kijken naar de door leerkrachten uitgevoerde activiteiten: gezien de wél gerapporteerde migratie op de scholen zal de implementatie op leerlingniveau nog zijn achtergebleven bij die op leerkrachtniveau. Ook opvallend is de hoge correlatie tussen de mate van implementatie van het lees- en het rekenprogramma (.73, p. 303): was de ene leerkracht een betere implementeerder dan de andere? En ten slotte is het typisch dat aan het aantal uitbreidingen aan een programma door een leerkracht een maximum van 10% werd gesteld (p. 57), terwijl de auteur elders vermeldt dat goede aanpassingen tot betere prestaties kunnen leiden (p. 100). Het 'weten welk programma je evalueert' (p. 55) is met het aanhouden van een 'getrouwheidsperspectief' (p. 56) onvoldoende recht gedaan. Als een (OSM-)programma ertoe leidt dat een leerkracht door aanpassing een evt. groter effect bereikt, is dat toch

prima? In deze optiek vertegenwoordigt adaptatie van een curriculum potentieel een zekere toegevoegde waarde (vgl. Stokking, 1986). Je moet het als ontwikkelaar nu eenmaal doen met en via de leerkracht.

In de Nederlandse situatie is er bovendien de mogelijkheid om de begeleider zo effectief mogelijk in te schakelen. Vanaf de derde versie van het verklaringsmodel (p. 118) worden de implementatiewaarden van de begeleidingsmodellen echter uit de analyses wegge laten. Dat is begrijpelijk omdat het model daarmee vereenvoudigd werd en zeker ook gezien het feit dat deze begeleidingsmodellen slecht geïmplementeerd bleken te zijn of op dit punt niet goed beoordeelbaar (p. 142-144). Maar de positieve samenhang die er desalniettemin bleek te zijn tussen de mate van begeleiding en de mate van implementatie van de programma's is toch wederom iets om op voort te bouwen? En het voorkomen van een negatieve relatie bij één programma (p. 166) zou toch nader onderzocht moeten worden, om daar zoveel mogelijk van te leren? De auteur volstaat nu met een moeizame discussie over de 'bruikbaarheid' van de ontwikkelde begeleidingsmodellen (pp. 31-33, 54, 58, 61-62, 72). Eerst lijkt het te gaan over de mate waarin deze het handelen van de begeleider kunnen sturen, later blijkt een begeleidingsmodel bruikbaar te zijn als het model zelf zowel als het onderwijsstimulerings- of gezinsactiveringsprogramma waarop het betrekking heeft voldoende geïmplementeerd zijn, nog weer later als er tussen beide een causaal verband bestaat, hoewel de causale bewijsvoering 'van minder belang' is en daarom 'éénvoudiger' kan plaatsvinden. Waarom niet, evenals t.a.v. de programma's, ook t.a.v. deze modellen gesteld dat een voldoende implementatie ervan voorwaarde is voor het kunnen doen van uitspraken over hun effectiviteit (nl. het bereiken van een voldoende implementatie van de programma's)? 'Bruikbaarheid' in de ogen van begeleiders zou dan gezien kunnen worden als voorwaarde voor de implementatie van deze modellen.

Slavenburg plaatst de OSM-resultaten terecht in de context van de Amerikaanse onderzoeksliteratuur op dit terrein en van de overige projecten in Nederland op dit gebied (IPA, GEON). Hij rapporteert dat een ver-

gelijkende evaluatie van IPA, GEON en OSM er destijds niet inzat (p. 35). Deze discussie werd destijds gevoerd in termen van 'externe evaluatie' (zie Stokking en De Vries, 1981). Na de negatieve beoordeling van de IPA-evaluatie (zie Van den Berg en Welten, 1981) had 'Den Haag' daar echter geen behoefte meer aan. In de Sociaalwetenschappelijke Raad van de Koninklijke Nederlandse Academie van Wetenschappen is deze wens echter steeds blijven leven, hetgeen recent heeft geleid tot een afzonderlijke studie (zie Scheerens, 1986).

Men kan de nu gerapporteerde OSM-evaluatie prijzen om z'n objectiviteit. Daarbij moet echter niet alleen worden afgegaan op het feit dat een direct betrokkene, die er veel in heeft geïnvesteerd, een negatief resultaat rapporteert. Objectiviteit, 'het voorwerp van studie recht doen', houdt naar onze mening ook in dat wordt nagegaan waardoor begeleidingsmodellen en programma's onvoldoende werden geïmplementeerd, of een grotere implementatie tot meer effect zou kunnen leiden, en hoe deze grotere implementatie zou kunnen worden bereikt. De conclusie van de auteur is wel erg pessimistisch. Het is terecht dat hij op grond van dit onderzoek met weinig concrete aanbevelingen komt voor scholen in onderwijsvoor-rangsgebieden. Maar de onderzoeksresultaten moeten er niet toe leiden het hoofd in de schoot te leggen.

Wat is nu de waarde van dit boek? Het is primair een vrij gedetailleerd en technisch onderzoeksverslag, waarin op tal van punten wordt verwezen naar andere OSM-publicaties. De auteur heeft veel onderwijswetenschappelijke en methodologisch-statistische literatuur verwerkt. Toch is het boek heel leesbaar gebleven. Voor een eerste kennismaking met het OSM-project als wetenschappelijk ontwikkelings- en onderzoeksproject is het wel geschikt, al vereist het daarvoor wellicht de nodige voorkennis en is het voor dat doel wat omvangrijk. Wie kennis wil nemen van de zeer uiteenlopende wetenschappelijke en praktisch-organisatorische invalshoeken die bij een dergelijk project aan de orde zijn, kan bij dit boek uitstekend terecht.

Literatuur

- Berg, A. van den & V.J. Welten, *Het Innovatieproject Amsterdam gewogen*. Den Haag: 1981.
- Hofstee, W.K.B., *De empirische discussie. Theorie van het sociaalwetenschappelijk onderzoek*. Meppel: 1980.
- Hofstee, W.K.B., *Evaluatie: een methodologische analyse. Tijdschrift voor onderwijsresearch*. 1982, 7, 193-202.
- Scheerens, J., *Enhancing educational opportunities for disadvantaged learners. A review of Dutch research on compensatory education and educational development policy*. Den Haag: 1986.
- Stokking, K.M., *Reactie op Hofstee: 'Evaluatie:*

- een methodologische analyse'. Tijdschrift voor onderwijsresearch*, 1983, 8, 34-38.
- Stokking, K.M., *Interpreteren en evalueren. Methodologie rondom de uitkomsten van onderzoek*. Deventer: 1984.
- Stokking, K.M., *Hoe bereiken we de scholen? Een studie naar de verspreiding van vernieuwingen in het Nederlandse onderwijs*. Tilburg: 1986.
- Stokking, K.M. & A.K. de Vries, *Een luis in de pels*. Eindverslag van het GEON-project. Den Haag: 1981.
- Tacq, J., *Causaliteit in sociologisch onderzoek. Een beoordeling van causale analysetechnieken in het licht van wijsgerige opvattingen over causaliteit*. Deventer: 1984.

Naar beter onderwijs

Redactie: J.C. Voogt en A. Reints

Inhoud

De auteurs beschrijven in deze bundel resultaten van hun onderwijskundige bemoeienis met de kwaliteit van het onderwijs. Zij geven daarbij aan welke moeilijkheden zich voordoen om kwaliteit van onderwijs te bepalen en verbeteringen te realiseren. Beter onderwijs omvat meer dan hogere cijfers voor lezen, rekenen, schrijven en taal bij leerlingen. Met deze bijdrage beogen de auteurs een overzicht te geven van het complexe terrein van de onderwijsverbetering. Op basis van feitenmateriaal en theoretische overwegingen worden wensen voor de toekomst geformuleerd.

De bundel Naar beter onderwijs kost:
f 39,-.

Voor wie interessant?

De inhoud van deze bundel is van belang voor onderwijsgevend, schoolleiders en

-bestuurders, begeleiders en opleidingsdocenten, inspecteurs en beleidsambtenaren, onderwijskundigen en onderzoekers.

Andere uitgaven in de reeks onderwijsinnovatie

- Hoe bereiken we de scholen - K.M. Stokking.
- Onderwijsinnovatie in verschuivend perspectief - R.M. van den Berg en R. Vandenbergh.
- Grootschaligheid in de onderwijsvernieuwing - R.M. van den Berg en R. Vandenbergh.

Voor meer informatie:

Zwijsen

5000 AV Tilburg postbus 805
013 - 353635

ook via de boekhandel