

Sturing van het onderwijsleerproces door middel van problemen: een veldexperiment¹

R.W. Tans², H.G. Schmidt³, B.E.J. Schade-Hoogeveen⁴ en W.H. Gijselaers³

ABSTRACT

In a field experiment, learning effects of a physiology course for physiotherapists were investigated. The aim of the study was to compare learning and achievement under a problem-based instructional mode with effects of direct instruction. In the problem-based learning condition, small tutorial groups analyzed problems, formulated learning goals and engaged in self-directed, independent learning activities. The problems used were constructed on the basis of objectives that also provided the framework for the direct instruction. Under the latter condition, students attended lectures and studied subject-matter prescribed by the teacher. The direct-instruction group performed better than the problem-based learning group on a multiple-choice test administered directly after the course. Ten weeks later, however, performance on a free recall test of two randomly selected samples showed the opposite result. Learning process data of the students under the problem-based condition indicate that the poor results on the multiple-choice test can be explained by the discrepancies between the learning goals on the one side and teaching objectives and test items on the other side. Students in the problem-based condition studied subject-matter that, although in itself relevant, was not foreseen by the teacher and, as a consequence, not tested. This non-intended learning comes into expression when students are asked to write down everything they remembered about a subject. Attention is paid to the implications of these results for research on the effects of 'open' instructional methods.

De idee dat het leerproces bevordert zou kunnen worden als lerenden in staat zouden worden gesteld hun eigen leerdoelen te formuleren en na te streven, is niet bepaald nieuw. Zij wordt al duidelijk gearticuleerd in het werk van de Amerikaanse psycholoog John Dewey (1910), en kan ook bij Piaget (1954) en Bruner (1959, 1961) worden teruggevonden. Deze auteurs verdedigen deze opvatting vanuit een *constructivistisch* perspectief op leren. In hun visie probeert iemand die iets leert, nieuwe informatie op actieve en herorganiserende wijze in te passen in al aanwezige cognitieve structuren. De uit deze activiteit resulterende representatie kan daarom niet zonder meer een complete en correcte afbeelding zijn van datgene wat ter lering werd voorgelegd, maar is veeleer een idiosyncratische constructie van het lerende individu (Rumelhart en Norman, 1978). Leren is in die opvatting het zelf ontdekken van structuur en regel in de werkelijkheid: kennis kan niet echt worden 'overgedragen', maar moet actief verworven worden.

Directe instructie is dus volgens Dewey minder zinvol dan leerlingen in de gelegenheid stellen in confrontatie met aan hen voorgelegde problemen, zelfstandig leerdoelen te formuleren en hen actief te laten zoeken naar informatie die het beste bij hun voorkennis aansluit. Informatie

Adres: H.G. Schmidt, Capaciteitsgroep Onderwijsontwikkeling en Onderwijsresearch, Postbus 616, 6200 MD Maastricht

¹ Met dank aan C. Salemans en B. Kerkhofs

² Universiteit van Amsterdam

³ Rijksuniversiteit Limburg

⁴ Akademie voor Fysiotherapie Leffelaar

die hen door de onderwijzende wordt voorgeschoteld in een door hem gekozen vorm, volgorde en moeilijkheidsgraad, kan nooit in dezelfde mate betekenisvol zijn als informatie die door de leerlingen zelf gezocht is (Kersh, 1958).

In hoeverre worden deze noties met betrekking tot de vermeende superioriteit van leerlinggecentreerde methoden door empirische gegevens ondersteund? Nauwelijks, moet worden gezegd. Een aantal recente overzichten van onderzoek naar effecten van traditioneel versus open onderwijs – in een enkel geval ongeveer 200 studies samenvattend – laten zien dat open onderwijs in de meeste gevallen wel een positieve invloed uitoefent op *attitudes* van lerenden ten opzichte van zichzelf, de leerstof en het onderwijs, maar dat van specifieke cognitieve effecten geen sprake lijkt te zijn (Giaconia en Hedges, 1982; Horwitz, 1979; Marshall, 1981; Peterson, 1979). Peterson (1979), die een meta-analyse uitvoerde op een groot aantal onderzoeken in dit veld, toonde aan dat traditioneel onderwijs over het geheel genomen tot iets betere prestaties leidt als het gaat om wiskunde en lezen, terwijl affectieve uitkomsten, zoals zelfconcept, attitudes tegenover het onderwijs, nieuwsgierigheid en onafhankelijkheid, in open onderwijs beter tot hun recht komen.

Moet op grond van deze data geconcludeerd worden dat de constructivistische opvattingen van Dewey en anderen in de praktijk van het onderwijs onvoldoende empirische ondersteuning vinden?

In deze bijdrage wordt de stelling verdedigd dat het debat over mogelijke cognitieve effecten van activerende onderwijsmethoden nog verre van beslist is. Het feit dat in verschillende evaluatiestudies nauwelijks prestatieverschillen tussen traditioneel en open onderwijs worden aangetoond, zou namelijk het gevolg kunnen zijn *van de wijze waarop in die onderzoeken leerwinsten werden gemeten*. Er zijn aanwijzingen dat bij de vaststelling van eventuele effecten van traditioneel en open, discovery en non-discovery onderwijs, veelal gebruik gemaakt wordt van toetsen die gebaseerd zijn op *de leerstofdoelstellingen van het conventionele onderwijs*. Wanneer leerlingen echter aangemoedigd worden hun eigen leerdoelen te verwoorden en na te streven, is het niet ondenkbaar dat leereffecten optreden op terreinen die – hoewel op zichzelf wellicht relevant – niet door de docent voorzien zijn. En datgene wat niet verwacht wordt, wordt gewoonlijk ook niet getoetst.

Een serieuze test van deze veronderstelling vereist de beschikbaarheid van een meetprocedure waarmee aspecten van het leerproces op meer gedetailleerde wijze zichtbaar gemaakt worden dan gebruikelijk en mogelijk is met data, ontleend aan een afsluitende toets. Met name is daarbij behoefte aan een procedure waarmee in kaart gebracht kan worden in hoeverre door de docent geformuleerde onderwijsdoelen, feitelijke leeractiviteiten van leerlingen, en leerproducten – in de vorm van antwoorden op een afsluitende toets – elkaar dekken. Als de hierboven geformuleerde hypothese, dat in vormen van open onderwijs leerlingen leeractiviteiten ontplooiën die nòch door de docent verwacht, nòch met een toets gemeten worden, juist is, dan mag verwacht worden dat onderwijsdoelen, leeractiviteiten en toetsitems elkaar *slechts gedeeltelijk* overlappen.

Een voor de hand liggende techniek waarmee dit dekkingsprobleem nader onderzocht kan worden, is gebruik te maken van expert-beoordelaars, die uit zouden moeten maken of, en in hoeverre, leeractiviteiten overeenstemmen met onderwijsdoelen en de operationalisering daarvan in toetsitems (Rovinelli en Hambleton, 1977; Porter, Schmidt, Floden en Freeman, 1978). Deze techniek is in het hieronder beschreven onderzoek gebruikt. Een nadeel van deze methode is echter dat vaak weinig inzichtelijk is hoe experts tot hun oordeel komen, en dat de betrouwbaarheid daarvan (in psychometrische zin) niet altijd even hoog is. Daarom zal daarnaast een meer analytische techniek geïntroduceerd worden. Deze methode houdt in dat onderwijsdoelen, bestudeerde leerstof (of voornemens daartoe), en toetsitems worden ontleed in microproposities – of concepten – die vervolgens met elkaar vergeleken worden. Deze taak kan met hoge interbeoordelaarsovereenstemming worden uitgevoerd, zonder van experts in het betref-

fende leerstofdomein gebruik te hoeven maken (Schmidt en Tans, 1985).

In het veldexperiment, waarvan in deze bijdrage verslag gedaan wordt, werden effecten van een conventionele cursus over het onderwerp 'spieren', die onderdeel is van het curriculum van een opleiding voor fysiotherapie, vergeleken met die van een 'probleemgestuurde' cursus over dat onderwerp, die op dezelfde onderwijsdoelen gebaseerd was. Beide programma's werden met dezelfde meerkeuzetoets afgesloten. Onderwijsdoelen, leerdoelen (in de probleemgestuurde conditie) en toetsitems werden vervolgens nader geanalyseerd, om de volgende vragen te kunnen beantwoorden:

1. In hoeverre treden tussen beide condities verschillen op in leerresultaat?
2. Waardoor kunnen die verschillen verklaard worden?
3. In hoeverre ontplooiën leerlingen in de probleemgestuurde conditie leeractiviteiten die, in het licht van de onderwijsdoelen, als niet-verwacht kunnen worden gekarakteriseerd?
4. Hoe relevant zijn deze niet-verwachte leeractiviteiten in het licht van het bestudeerde onderwerp?

Een secundaire doelstelling van het onderzoek was de (predictieve) validiteit van de analytische methode voor leesstofanalyse vast te stellen.

METHODE

Proefpersonen.

Honderdtweëntwintig eerstejaars studenten aan een instelling voor hoger beroepsonderwijs (een academie voor fysiotherapie) namen deel aan het experiment. Zij werden per klas aselekt over de condities verdeeld. De samenstelling van de verschillende klassen – acht in totaal – was eveneens aselekt tot stand gekomen. Vijfenvestig studenten volgden klassikaal onderwijs in drie groepen. Deze conditie wordt verder de 'directe instructie' genoemd.

Zevenenzeventig studenten doorliepen de cursus 'spieren' op probleemgestuurde wijze, en werden daartoe ingedeeld in tien kleine groepen. Alle groepen werden door één docent begeleid.

Materiaal en procedure

Constructie van de cursussen. Aan een bekend handboek op het terrein van de fysiologie (Bernards en Bouman, 1976) werden door de docent 65 doelstellingen ontleend, betrekking hebbend op de fysiologie van spieren. Deze doelstellingen vormden de basis waarop voor beide cursussen het onderwijs geconstrueerd werd.

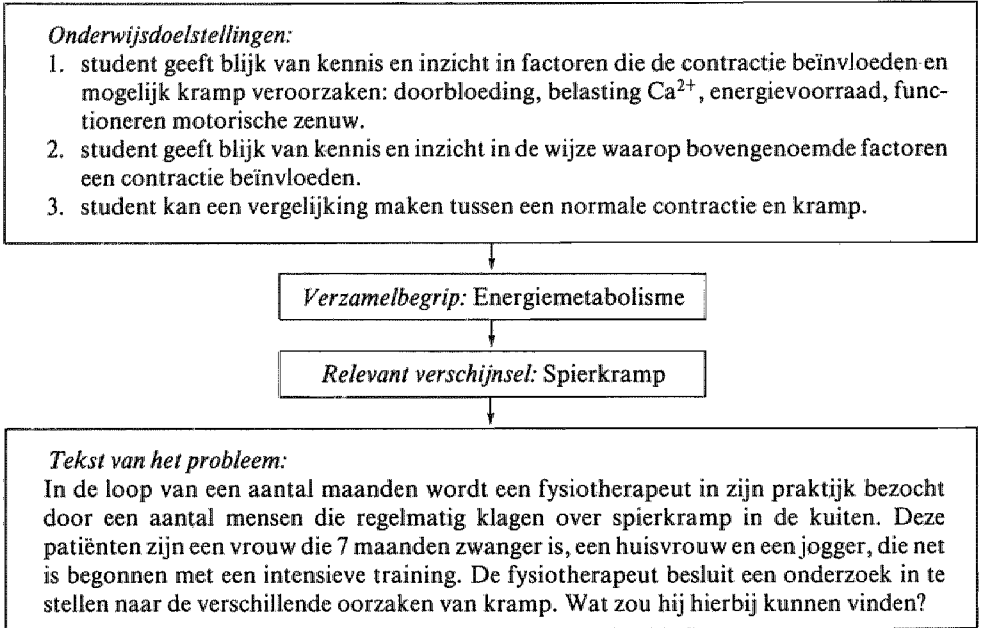
Onder de directe-instructie-conditie volgden de studenten gedurende zeven weken één maal per week gedurende twee uur een les over het onderwerp, bestudeerden als huiswerk opgegeven leerstof uit het handboek, en beantwoordden vragen over de stof die tijdens een volgende les door de docent besproken werden.

Onder de probleemgestuurde conditie analyseerden de studenten in kleine groepen, en in diezelfde tijdperiode, één maal per week één of meer problemen en formuleerden op basis van deze analyse een aantal leerdoelen. Deze leerdoelen werden vervolgens door middel van zelfstudie nagestreefd. Daarbij kon gebruik gemaakt worden van een verzameling handboeken die in de bibliotheek beschikbaar waren.

De problemen die in deze conditie als startpunt voor leeractiviteiten dienden, werden op de volgende wijze geconstrueerd: De onderwijsdoelstellingen werden gegroepeerd in een aantal bovengeordende categorieën, zoals 'energiemetabolisme', 'excitatie' en 'circulatie'. Voor elk van deze categorieën werd een verschijnsel of verzameling verschijnselen uit de werkelijkheid gezocht, van belang voor fysiotherapeuten in spé, dat als een concrete manifestatie van die categorie beschouwd kon worden. Het resulterende probleem tenslotte, bestond uit de beschrij-

ving van dat verschijnsel, geplaatst in een context waarvan kon worden aangenomen dat deze door de studenten als betekenisvol herkend zou worden. In tabel 1 is deze constructiewijze aan de hand van een voorbeeld uitgewerkt.

Tabel 1: Voorbeeld van de constructie van een probleem



Op deze wijze werden 16 problemen geconstrueerd die met de geformuleerde doelstellingen correspondeerden.

De studenten in de probleemgestuurde conditie beschikten niet over doelstellingen, noch over verzamelbegrippen. Hen werden alleen de problemen voorgelegd met het verzoek ze te 'analyseren', dat wil zeggen: verklaringen te bedenken in termen van onderliggende processen, mechanismen of principes. De leerlingen maakten daarbij gebruik van een systematische werkprocedure, die hen voorafgaande aan de cursus geleerd werd. Deze procedure helpt bij het transformeren van een probleem in een verzameling leerdoelen die vervolgens door middel van individuele studie nagestreefd worden. (Zie voor meer uitgebreide beschrijvingen Barrows, 1983; Barrows en Tamblyn, 1980; Schmidt en Bouhuijs, 1983; Schmidt en De Volder, 1984). Het was de groepen toegestaan een keuze te maken uit de voorgelegde problemen.

De *toets* bestond uit 60 items, ieder met vier antwoordalternatieven, die verondersteld werden de 65 doelstellingen volledig te dekken. De interne consistentie van deze toets was gelijk aan $\alpha = 0,78$ (berekend over de totale onderzoekspopulatie). Hij werd onmiddellijk na afloop van de onderwijsperiode afgenomen.

Om een indruk te krijgen van de lange-termijn-effecten van het gegeven onderwijs werd tien weken na de meerkeuzetoets een 'free-recall'-toets afgenomen bij een klein deel van de studenten. Deze toets was bedoeld om kennis van een aantal centrale begrippen uit de cursus te meten en bestond uit de volgende open vraag: "Schrijf *alles* op wat je weet van de begrippen: 1.

spierkracht, 2. rusttonus, en 3. contractie". Hij werd afgenomen bij twaalf aselekt gekozen studenten: zes uit de experimentele, en zes uit de controleconditie. Deze studenten ontvingen voor hun medewerking een kleine vergoeding. Ze kregen 20 minuten tijd om hun antwoord op te schrijven. De op deze wijze verkregen teksten werden ontleed in proposities. Een propositie is meestal een onderwerp-gezegde combinatie die één bepaald idee uitdrukt. Van elk van deze proposities werd vervolgens vastgesteld of hij inhoudelijk correct dan wel incorrect was. Per proefpersoon werd tenslotte het totaal aantal correcte proposities berekend. De interbeoordelaarsovereenstemming voor deze procedure was gelijk aan 0,92.

Tenslotte werden alle *leerdoelen* verzameld die door de probleemgestuurde groepen geproduceerd werden. Deze leerdoelen werden beschouwd als indicatoren voor de feitelijke leeractiviteiten die onder deze conditie ontplooid werden. Nagegaan werd vervolgens in hoeverre de door de docent geformuleerde doelen – verderop steeds onderwijsdoelstellingen genoemd –, de leerdoelen van de studenten in de probleemgestuurde conditie, en de toetsitems elkaar dekten. Deze vergelijking vond op twee manieren plaats: met behulp van beoordelaars, en met behulp van een tekstreductietechniek.

De *beoordeling* verliep als volgt: twee beoordelaars die goed thuis waren in het betreffende leerstofgebied stelden onafhankelijk van elkaar vast welke van de onderwijsdoelstellingen door welke groepen als leerdoel genoemd werden. Op deze wijze kon per probleem en per groep een percentagescore berekend worden die de mate aangeeft waarin onderwijsdoelstellingen in leerdoelen werden afgebeeld. De interbeoordelaarsovereenstemming voor deze procedure was 0,73.

Bij het vaststellen van de mate waarin leerdoelen en toetsitems elkaar dekten werd een enigszins andere weg gevolgd. Voor elk van de items werd bepaald hoeveel van de probleemgestuurde groepen leerstof, waarover met dat item getoetst werd, bestudeerd zouden kunnen hebben, gegeven hun leerdoelen. De interbeoordelaarsovereenstemming daarvoor was gelijk aan 0,70.

De *tekstreductietechniek*, die voor de analyse van het leerproces werd gebruikt, is vergelijkbaar met methoden die door Breuker (1980) en Dansereau en Holley (1982) zijn voorgesteld. Deze procedure heeft de volgende vorm: Zinnen kunnen getransformeerd worden tot lijsten van concepten: werkwoorden, zelfstandige naamwoorden, bijvoegelijke naamwoorden. Een dergelijke lijst kan vervolgens gereduceerd worden door alleen die concepten te accepteren die een directe relatie onderhouden met een bepaald van te voren gekozen thema of onderwerp. Begrippen die dat niet in het bijzonder doen, of die gebruikt worden om de integriteit van zinnen in stand te houden, worden verwijderd. Op deze wijze kunnen niet alleen studieteksten geanalyseerd worden, maar ook andere elementen van een onderwijsleerproces, zoals onderwijsdoelstellingen, leerdoelen en toetsitems. De onderwijsdoelstellingen, geformuleerd in het kader van het onderwerp 'spieren', die afgedrukt zijn in tabel 1 kunnen bijvoorbeeld gereduceerd worden tot de volgende lijst met zeven concepten:

contractie
kramp
doorbloeding
belasting
Ca²⁺
energievoorraad
motorische zenuw

Op deze wijze werden zowel onderwijsdoelstellingen als leerdoelen en toetsitems gereduceerd. De drie resulterende verzamelingen concepten werden met elkaar vergeleken om na te gaan in hoeverre zij elkaar dekten. De interbeoordelaarsovereenstemming voor deze procedure was gelijk aan 0,87.

RESULTATEN

Leerproducten

Onder dit hoofd zullen de resultaten van de beide toetsingen gepresenteerd worden. Tabel 2 geeft de resultaten weer behaald door beide groepen op de 60-item meerkeuzetoets.

Tabel 2: Gemiddelde aantallen goede antwoorden (met standaarddeviaties) op de meerkeuzetoets.

| | \bar{X} | SD | N |
|-------------------------|-----------|------|-----|
| Probleemgestuurde groep | 27,72 | 4,88 | 74 |
| Controlegroep | 39,24 | 4,64 | 45 |
| Totaal | 32,08 | 7,37 | 119 |

Op deze gegevens werd een variantieanalyse uitgevoerd. Het resultaat daarvan is $F(1,117) = 162,16$; $p < 0,0001$. Studenten die gedurende zes weken een serie lessen over het onderwerp 'spieren' gevolgd hebben, leveren dus een significant betere prestatie op de toets dan studenten die in kleine groepen aan problemen met betrekking tot hetzelfde onderwerp gewerkt hebben.

In tabel 3 worden de resultaten weergegeven van de 'free-recall' toets die tien weken na afsluiting van het onderwijs bij twaalf aselekt gekozen proefpersonen werd afgenomen (één persoon produceerde een onbruikbaar protocol).

Tabel 3: Gemiddelde aantallen correcte proposities (met standaarddeviaties), geproduceerd tien weken na afloop van het onderwijs.

| | \bar{X} | SD | N |
|-------------------------|-----------|------|----|
| Probleemgestuurde groep | 31,50 | 8,26 | 6 |
| Controlegroep | 21,60 | 4,56 | 5 |
| Totaal | 27,00 | 8,32 | 11 |

Beide gemiddelden verschillen eveneens significant van elkaar: $F(1,9) = 5,66$; $p < 0,005$, nu echter in omgekeerde richting. De probleemgestuurde groep herinnert zich gemiddeld genomen meer van de drie kernbegrippen waarnaar gevraagd werd, dan de groep studenten die aan de directe-instructie-conditie heeft deelgenomen.

Leerprocesgegevens

Op basis van de procesgegevens, verzameld onder de probleemgestuurde conditie, werden onderwijsdoelstellingen, leerdoelen en toetsitems met elkaar vergeleken. Doel was na te gaan in welke mate deze drie elementen van het onderwijsleerproces met elkaar in overeenstemming waren. Eerst zullen de resultaten van het onderzoek naar het verband tussen onderwijsdoelstellingen en leerdoelen gepresenteerd worden: In hoeverre worden de eerste in de laatste gepresenteerd?

De studenten in de probleemgestuurde groepen produceerden in totaal 314 leerdoelen, dat is gemiddeld 31,4 doelen per groep. Per probleem werden, rekening houdend met het feit dat studenten een keuze konden maken uit het aanbod, 5,06 leerdoelen geformuleerd. Tussen de groepen en tussen problemen bestonden aanzienlijke verschillen in aantallen geproduceerde doelen.

In tabel 4 zijn percentages weergegeven die de mate aanduiden waarin naar het oordeel van de deskundigen de leerdoelen van een groep de onderwijsdoelstellingen dekken die aan het betreffende probleem ten grondslag lagen. Groep 1 bijvoorbeeld reproduceerde als leerdoel 66,7 procent van de onderwijsdoelstellingen op basis waarvan probleem 1 geconstrueerd was. Lege cellen geven aan dat het desbetreffende probleem niet door de betreffende groep geanalyseerd is.

Tabel 4: Mate van overeenstemming tussen onderwijsdoelstellingen en leerdoelen volgens globale beoordelingsmethode uitgedrukt in een percentage.

| GROEP | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Gemiddelde |
|----------|------|-----|-----|------|------|------|-----|------|-----|------|------------|
| PROBLEEM | | | | | | | | | | | |
| 1 | 66,7 | 100 | 100 | 66,7 | 66,7 | 66,7 | 100 | 33,3 | | 66,7 | 74,1 |
| 2 | | | | | | | | | | | |
| 3 | 50 | | | | | | | | 100 | | 75 |
| 4 | | 100 | | 100 | 28,6 | | 100 | | | | 82,2 |
| 5 | 20 | | 40 | | | 40 | | 100 | 80 | 100 | 63,3 |
| 6 | | | | | | | | 33,3 | | 33,3 | 33,3 |
| 7 | 50 | 50 | 50 | 75 | 50 | | 50 | 75 | 50 | | 56,3 |
| 8 | | | 80 | | | 80 | 80 | | | 80 | 80 |
| 9 | | 75 | | | | | | | 50 | | 62,5 |
| 10 | | | | | | 100 | 100 | 100 | | | 100 |
| 11 | 100 | | 50 | | 50 | 50 | | | 50 | 50 | 58,3 |
| 12 | | | 100 | 50 | | | | | | | 75 |
| 13 | | | | | | | | 66,7 | | | 66,7 |
| 14 | 75 | 75 | | 100 | 100 | | | | | | 87,5 |
| 15 | | 40 | 40 | 60 | | 40 | 40 | | 40 | 70 | 47,1 |
| 16 | 100 | | | | | | | 100 | | | 100 |

Uit deze tabel kan worden afgeleid dat de gemiddelde dekking van onderwijsdoelstellingen door leerdoelen 68,3 procent is (met een standaarddeviatie van 24,6). Dus: onderwijsdoelstellingen werden door de beoordelaars in ongeveer zeventig procent van de gevallen in de leerdoelen van de groepen teruggevonden. In ruwe aantallen uitgedrukt bleken 248 door de groepen geproduceerde leerdoelen samen te vallen met datgene wat door de docent bedoeld was. Dat betekent echter dat naar het oordeel van de deskundigen 66 leerdoelen – dat is bijna een kwart van het totaal aantal geproduceerde – zich buiten het door de docent afgepaalde domein bewegen.

Met de resultaten van de tekstreductietechniek werd op dezelfde wijze omgegaan als met die van de beoordelingsmethode. Per probleem en per groep werd een percentagescore berekend die aangaf in welke mate de door de docent in haar doelstellingen genoemde concepten ook door de studenten genoemd werden. Op deze wijze ontstond een matrix, vergelijkbaar met tabel 4. De gemiddelde dekking van concepten ontleend aan de onderwijsdoelstellingen door concepten ontleend aan de leerdoelen was gelijk aan 39,6 procent (standaarddeviatie 22,9). De mate van dekking vastgesteld met de reductiemethode is dus geringer dan die vastgesteld door de beoordelaars. Dat heeft waarschijnlijk te maken met het feit dat de tekstreductietechniek 'strenger' is bij het beantwoorden van de vraag wat overeenstemming is. Alleen als concepten identiek of elkaars parafraze zijn, dekken ze elkaar. Beoordelaars kijken soms eerder naar wat bedoeld is dan naar wat er staat.

In hoeverre vormden de items waarmee verworven kennis getoetst werd een representatieve afspiegeling van de leeractiviteiten van de studenten in de probleemgestuurde conditie?

Teneinde na te gaan in welke mate de zestig toetsitems door de leerdoelen gedekt werden, bepaalden de beoordelaars voor elk van de items hoeveel groepen de leerstof die op dat item betrekking had, bestudeerd zouden kunnen hebben (voorzover dat van hun leerintenties kon worden afgeleid). Als alle groepen die stof bestudeerd hadden, was de dekking voor dat item honderd procent. De op deze wijze berekende gemiddelde dekking per item is dan 20 procent, dat wil zeggen dat gemiddeld genomen de leerstof betrekking hebbend op een bepaald item door twee (van de tien) groepen zeker bestudeerd werd (standaarddeviatie is 15). De resultaten van de tekstreductietechniek laten een gemiddelde dekking van 36,3 procent zien (met een standaarddeviatie van 12,6).

Tenslotte werd nagegaan in hoeverre de leerdoelen die door de studenten geformuleerd werden, als een valide indicator van hun werkelijke studieactiviteiten beschouwd konden worden. Daartoe is een productmoment correlatiecoëfficiënt berekend tussen de frequentie waarmee een leerdoel genoemd werd en de prestatie op het ermee corresponderende item*. Deze correlatie kan beschouwd worden als de (predictieve) validiteit van de leerdoelen als indicator van leeractiviteiten.

De correlatie tussen beide variabelen was gelijk aan 0,42, wanneer leerdoelen en items geanalyseerd worden met behulp van beoordelaars, en aan 0,44 wanneer dat gebeurt volgens de tekstreductietechniek. In het licht van het feit dat het bij leerdoelen gaat om een *afgeleide* variabele, zijn deze validiteiten als bevredigend te beschouwen.

DISCUSSIE

In de prelude tot deze bijdrage werd onderzoek geciteerd waaruit bleek dat voorzover verschillen in het cognitieve domein konden worden aangetoond tussen conventioneel onderwijs en open onderwijs, deze meestal ten nadele van de laatste variant uitvielen (Giaconia en Hedges, 1982; Horwitz, 1979; Marshall, 1981; Peterson, 1979). Daarbij moet echter wel aangetekend worden dat die verschillen klein waren en vooral in het wiskunde- en leesonderwijs werden aangetroffen.

In het hier gerapporteerde veldexperiment, waarin een probleemgestuurde cursus spierfysiologie werd vergeleken met een meer directe onderwijsaanpak, werden vergelijkbare resultaten gevonden. Studenten die een serie lessen over het onderwerp volgden, gecombineerd met huiswerk, leverden een niet onaanzienlijk betere prestatie op een meerkeuzetoets dan studenten die problemen analyseerden en op grond daarvan zelfgekozen onderwerpen bestudeerden.

Merkwaardig was dat tien weken na afloop van het experiment een omgekeerd effect gevonden werd in de "free recall"-protocollen van twee aselekt gekozen steekproeven uit de onderzoekspopulatie. De studenten uit de probleemgestuurde groep wisten zich meer te herin-

* De redenering daarbij is als volgt: Als leerdoelen een valide indicatie zijn voor feitelijke leeractiviteiten, dan moet de frequentie waarmee ze genoemd worden, een indicator zijn voor de mate waarin leerstof waarnaar ze verwijzen bestudeerd is. Dit laatste bepaalt de hoogte van de prestatie op de met die leerstof corresponderende items.

Als maat voor de prestatie van de experimentele groep op een item werd niet de p-waarde van dat item genomen. Een lage p-waarde kan immers zowel beduiden dat het betreffende item door de experimentele groep niet of nauwelijks bestudeerd is, als dat het wel bestudeerd is, maar te moeilijk was. Een betere schatter van de prestatie van de experimentele groep op een bepaald item is daarom de betreffende p-waarde, gedeeld door de p-waarde van de controle-groep voor dat item: P_e/P_c . Een eenvoudig voorbeeld kan dat verduidelijken: voor item 8 uit de toets was $P_e = 0,18$ en $P_c = 0,82$. Voor item 29 was $P_e = 0,16$, maar $P_c = 0,13$. Op grond hiervan kan geconcludeerd worden dat item 8 door de experimentele groep niet of nauwelijks bestudeerd is terwijl item 29 te moeilijk was. De moeilijkheidsgraad van het item wordt dus met behulp van de ratio 'weggedeeld'.

neren van drie onderwerpen die in de cursus een centrale rol gespeeld hadden, dan de representanten van de controlegroep.

Hoe zijn deze op het eerste gezicht toch strijdige uitkomsten met elkaar te rijmen? Als toeval buiten beschouwing gelaten wordt, doen zich in eerste instantie twee mogelijke verklaringen voor. De eerste probeert de ongerijmdheid te verklaren uit de aard van de geheugentaken die aan de betrokken proefpersonen werd voorgelegd; de tweede verklaring zoekt het in de aard van de door hen verworven kennis.

Eerst enkele opmerkingen over de aard van de geheugentaken. Een "free recall"-opdracht, een taak dus waarbij proefpersonen gevraagd wordt alles wat zij zich herinneren op te schrijven, doet over het algemeen een groter beroep op de terughaalstrategieën ("retrieval strategies") die personen beschikbaar hebben, dan een herkenningstaak zoals het zoeken van het juiste antwoord-alternatief uit een aantal aangeboden alternatieven. Dat komt omdat bij een herinneringstaak geheugensteuntjes ('cues') ontbreken die reconstructie van het geleerde mogelijk maken, zoals dat bij een herkenningstaak in overvloedige mate het geval is. Een betere 'recall' duidt dan ook meestal op een betere *organisatie* van de kennis in het geheugen. Een dergelijke betere organisatie komt tot uitdrukking in meer verschillende 'retrievalpaden' (Anderson en Reder, 1979). In deze opvatting zou de probleemgestuurde groep minder kennis bezitten (getuige het resultaat op de meerkeuzetoets); maar wat aan kennis aanwezig is, beter gestructureerd hebben (hetgeen in de betere 'free recall' tot uitdrukking komt).

De tweede hypothese veronderstelt dat de meerkeuzetoets minder adequaat datgene meet wat de probleemgestuurde groep geleerd heeft dan de 'recall'-taak. In die opvatting bestaan er niet zozeer kwantitatieve als wel kwalitatieve verschillen tussen beide condities: de probleemgestuurde groep heeft niet zozeer 'minder' geleerd dan de groep die onder de directe-instructie conditie werkte, maar heeft zich deels met onderwerpen beziggehouden die door de docent niet voorzien waren. Op de meerkeuzetoets kunnen studenten in deze conditie slechts laten zien wat geleerd is voorzover dát overeenkomt met wat de docent verwachtte. Voor een recall taak geldt deze beperking in minder sterke mate. Daar ligt het initiatief bij degene die schrijft. Alle kennis, mits correct en relevant, 'telt mee'.

Voor deze laatste hypothese werd in de leerprocesgegevens enige ondersteuning gevonden. Onderwijsdoelstellingen, leerdoelen (als indicatoren voor feitelijke leeractiviteiten) en toets-items bleken elkaar inhoudelijk maar in beperkte mate te dekken. Dat verschijnsel werd zowel geconstateerd door inhoudsdeskundige beoordelaars als met behulp van een tekstreductietechniek.

In figuur 1 is de mate van overlap grafisch weergegeven. De data zijn ontleend aan de analyse met behulp van de reductiemethode.

De figuur kan als volgt worden gelezen: In de eerste kolom staan alle concepten die in het materiaal zijn teruggevonden. In kolom 2, 3 en 4 wordt aangegeven of en in hoeverre het betreffende concept in respectievelijk doelstellingen, leerdoelen en items werd waargenomen. De lengte van de balk geeft een indicatie van de frequentie waarmee dat concept voorkwam. Lege plekken duiden aan dat geen corresponderende concepten gevonden werden.

Figuur 2 illustreert de centrale these van deze bijdrage, namelijk dat lerenden in een open context idiosyncratische leerdoelen formuleren die hen ertoe brengen zich met leeractiviteiten te engageren waarop de docent niet verdacht is en die daarom niet getoetst worden. De figuur biedt daarmee een mogelijke verklaring voor de vanuit constructivistisch perspectief teleurstellende resultaten van onderzoek naar effecten van activerende onderwijsmethoden. Zij suggereert dat de aard van die uitkomsten inderdaad een functie zou kunnen zijn van de wijze waarop in dat onderzoek leerwinsten geoperationaliseerd en gemeten zijn.

De vraag kan nu gesteld worden hoe zinvol die niet-verwachte uitkomsten van het leren in een probleemgestuurde context zijn in het licht van het onderwijs als geheel. Immers: als datgene waaraan onder die conditie deels gewerkt is, als niet ter zake doende 'Spielerei' gekarakteriseerd

| onderwijsdoelstellingen | leerdoelen | items |
|-------------------------|------------|----------------------------|
| CONTRACTIE | 00 2 | 00000000 9 |
| KRAMP | 0 1 | 00000000 10 |
| DOORBLOEDING | 000000 7 | 0000000000000000 21 |
| BELASTING | 000 3 | 000000 7 |
| Ca ²⁺ | 0 1 | 00 2 |
| ENERGIE | 00000000 9 | 00000000 8 |
| MOTORISCHE ZENUW | 000 3 | 00 2 |
| SPIER | 000 3 | 00000000...000000000000 47 |
| TEMPERATUUR | 000 3 | 00000000 9 |
| TONUS | 0000 5 | 0000000000000000 19 |
| ATP | 0 1 | 00 2 |
| AFVALSTOFFEN | 0 1 | 00 2 |
| VISCOSITEIT | 0 1 | 0 1 |
| ELASTICITEIT | 00 2 | 0 1 |
| STIJFHEID | 00 2 | 00 3 |
| PIJN | 000 3 | 00000000 9 |
| KRACHT | 0000000 8 | 00000000 19 |
| RECROITMENT | 00 2 | 00 3 |
| MYOFASIBACK | 0 1 | 00000000 8 |
| SYMPTOMEN | 0 1 | 00 2 |
| OWESPANNING | 0 1 | 00000 5 |
| FYSIOTHERAPIE | 000 3 | 000000000000 12 |
| MOTORUNIT | 0 1 | 0000000000 11 |
| SPIERVEZELS | 0000 4 | 00000 6 |
| ZENUWVEZELS | 0 1 | 0000000000 15 |
| BEDEELGERIGHEID | 0 1 | 00000000 8 |
| CONDITIE | 00 2 | 00000000 8 |
| ADENHALING | 0 1 | 00000 5 |
| TRAINING | 00 2 | 0000000 7 |
| STOPWISSELING | 0 1 | 00000 5 |
| MASSAGE | 00 2 | 000000000 9 |
| VERLAMING | 0 1 | 00000 5 |
| BOTULINE-TOXINE | 0 1 | 0000 4 |
| ACETYLCHOLINESTERASE | 0 1 | 0000 4 |
| CURARE | 0 1 | 0000 4 |
| REK | 0 1 | 0000 4 |
| VET | 0 1 | 000 3 |
| DIRET | 0 1 | 0 1 |
| HERSTEL | 0 1 | 0 1 |
| RESPIRATIE-QUOTIENT | 0 1 | 0 1 |
| WARMING-UP | 0 1 | 000000 6 |
| SPORT | 0 1 | 000000000 10 |
| U. K. G. | 0 1 | 0000000 7 |
| OEFPENINGEN | 0 1 | 0000000 7 |
| ATROFIE | 0 1 | 00000 5 |
| REFLEX | 0 1 | 00 2 |
| ARBEID | 0 1 | 000 3 |
| ACETYLCHOLINE | 0 1 | 00000 6 |
| MELKZUUR | 0 1 | 0000 4 |
| α - MOTORNEURON | 0 1 | 00000000 9 |
| GETRAINDEHEID | 0 1 | 0 1 |
| CENTRAAL ZENUWSTELSEL | 0 1 | 0 1 |
| SPIERSPOEL | 0 1 | 0 1 |
| MYOSINE | 0 1 | 0 1 |
| KOOLHYDRAAT | 0 1 | 000 3 |

Figuur 1: Onderlinge dekking van onderwijsdoelstellingen, leerdoelen en toets items.

moet worden, dan zou men langs een omweg alsnog tot de conclusie moeten komen dat deze benadering inefficiënt leren bevordert.

Teneinde deze vraag voor het onderhavige materiaal te onderzoeken, werd de twee vakdeskundigen gevraagd de niet-verwachte leerdoelen onafhankelijk van elkaar op relevantie te beoordelen. Criterium voor relevantie was de mate waarin vervulling van zo'n leerdoel een bijdrage levert aan de theoretische en professionele ontwikkeling van de desbetreffende studenten. Met andere woorden: hebben zij de verworven kennis in hun latere fysiotherapeutische praktijk nodig? De interbeoordelaars overeenstemming voor deze taak was gelijk aan 67 procent. Verschil van mening werd door middel van discussie opgelost.

77 Procent van de niet-verwachte leerdoelen bleek in die zin als relevant of tamelijk relevant te worden aangemerkt. Het ging daarbij in meerderheid om doelen die op een of andere wijze met ziektebeelden of met fysiotherapeutische behandelingsmethoden in verband gebracht konden worden. Een voorbeeld is: "Wat is decubitus?"

Geconstateerd kan dus worden dat een deel van de discrepanties tussen docent en studenten

| | | |
|----------------------------|----------|-----------------|
| AEROOB | o 1 | o 1 |
| ANAEROOB | o 1 | oooo 4 |
| ZUURSTOF | ooo 3 | o 1 |
| RUSTLENGTE | o 1 | o 1 |
| NEUROTRANSMITTER | o 1 | o 1 |
| VOEDING | oooooo 6 | o 1 |
| STRESS | ooo 3 | |
| COORDINATIE | o 1 | |
| ZWELLING | o 1 | |
| INSUFFICIENTIE | o 1 | |
| CO ₂ | o 1 | |
| BANDEN | o 1 | |
| KAPSELS | o 1 | |
| SPASME | o 1 | |
| BOT | o 1 | |
| THERMOSE | o 1 | |
| NECROSE | o 1 | |
| SPIERVERSTERKING | ooo 3 | |
| GEMRICHTEN | oooo 4 | |
| BINDWEEFSEL | o 1 | |
| PSYCHISCH | oooooo 6 | |
| DECUBITUS | oooooo 6 | |
| SPIITVOETEN | o 1 | |
| IMMOBILITEIT | oo 2 | |
| PUNCTIEVERLIES | oo 2 | |
| CONTRACTUUR | oo 2 | |
| PRIME-MOVER | o 1 | |
| MEMBRAN | oo 2 | |
| ZWEEFSLAG | o 1 | |
| HYPNOSE | o 1 | |
| YOGA | o 1 | |
| TOPOGRAFIE | o 1 | |
| MOTORISCHE VOORHOORNCCEL | | oo 2 |
| SARCOTUBULAIR SYSTEEM | | o 1 |
| REFRACTAIRE PERIODE | | o 1 |
| ACTIEPOTENTIAAL | | oooooooooooo 11 |
| K ⁺ - EFFLUX | | o 1 |
| MEMBRAAN | | oo 2 |
| PRESYNAPTISCHE MEMBRAAN | | o 1 |
| RELAXATIE | | oo 2 |
| SARCOPLASMATISCH RETICULUM | | oooo 4 |
| RECEPTOR | | o 1 |
| SARCOMER | | o 1 |
| SUNNATIE | | o 1 |
| NA ⁺ - INFLUX | | o 1 |
| POLARISATIE | | ooo 3 |
| REMSTOF | | o 1 |
| EINDPLAATPOTENTIAAL | | o 1 |
| GLYCOGEN | | o 1 |
| LACTAAT | | o 1 |
| SARCOLEMA | | oo 2 |
| MYOGLOBINE | | o 1 |
| CREATINEFOSFAAT | | o 1 |
| ADEMHALINGSKETEN | | o 1 |
| KREBSCYCLUS | | o 1 |
| GLYCOLYSE | | o 1 |
| SKELET | | o 1 |

terug te voeren is op de interesse die studenten aan de dag leggen voor praktische toepassingen van fysiologisch-anatomische kennis. Blijkbaar benutten de studenten de vrijheid die hen geboden wordt voor een deel om een beter inzicht te verwerven in hun latere beroepspraktijk. Gezien vanuit het perspectief waaruit studenten een bepaalde opleiding kiezen, is dat een begrijpelijke voorkeur, die echter in de onderwijsdoelstellingen van de onderzochte cursus onvoldoende werd verdisconteerd.

REFERENTIES

- Anderson, J.R. en L.M. Reder (1979). An elaborative processing explanation of depth of processing. In: L.S. Cermak en F.I.M. Craik (Eds.). *Levels of processing in human memory*. Hillsdale, N.J.: Lawrence Erlbaum.
- Barrows, H.S. (1983). Problem-based, self-directed learning. *Journal of the American Medical Association*, 250, 3077.
- Barrows, H.S. en R.M. Tamblyn (1980). *Problem-based learning*. New York: Springer Publishing.

- Bernards, J.A. en L.N. Bouman (1979). *Fysiologie van de mens*. Utrecht: Bohn, Scheltema en Holkema.
- Breuker, J. (1980). *In kaart brengen van leerstof*. Utrecht: Het Spectrum.
- Bruner, J.S. (1959). Learning and thinking. *Harvard Educational Review*, 29, 184-192.
- Bruner, J.S. (1961). The act of discovery. *Harvard Educational Review*, 31, 21-32.
- Dansereau, D.F. en C.D. Holley (1982). Development and evaluation of a text mapping strategy. In: A. Flammer en W. Kintsch (Eds.) *Discourse processing*. Amsterdam: North-Holland Publishing.
- Dewey, J. (1910). *How we think*. Boston: Heath & Co.
- Giaconia, R.M. en L.V. Hedges (1982). Identifying features of effective open education. *Review of Educational Research*, 52, 579-602.
- Horwitz, R.A. (1979). Psychological effects of the 'open classroom'. *Review of Educational Research*, 49, 71-86.
- Kersh, B.J. (1958). The adequacy of "meaning" as an explanation for superiority of learning by independent discovery. *Journal of Educational Psychology*, 49, 282-292.
- Marshall, H.H. (1981). Open classrooms: has the term outlived its usefulness? *Review of Educational Research*, 51, 181-192.
- Peterson, P.L. (1979). Direct instruction reconsidered. In: Peterson en H.L. Walberg (Eds.). *Research on teaching*. California: McCutchan.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- Porter, A.C., W.H. Schmidt, R.E. Floden en D.J. Freeman (1978). *Impact on what?: the importance of content covered*. East Lansing: The Institute for Research on Teaching.
- Rovinelli, R.J. en R.K. Hambleton (1977). On the use of content specialists in the assessment of criterion-referenced test item validity. *Tijdschrift voor Onderwijsresearch*, 2, 49-60.
- Rumelhart, D.E. en D.A. Norman (1978). Accretion, tuning and restructuring. In: J.W. Cotton en R. Klatzky (Eds.). *Semantic factors in cognition*. Hillsdale: Lawrence Erlbaum.
- Schmidt, H.G. en P.A.J. Bouhuijs (1983). *Onderwijs in taakgerichte groepen*. Utrecht: Het Spectrum.
- Schmidt, H.G. en M.L. de Volder (Eds.) (1984). *Tutorials in problem-based learning*. Assen: Van Gorcum.
- Schmidt, H.G. en R. Tans (1985). *A simple method for the content analysis of subject-matter in educational texts*, Rijksuniversiteit Limburg.

Manuscript ontvangen 3-4-1985

Definitieve versie ontvangen 22-10-1985