

VERGELIJKENDERWIJS :

ONDERZOEK NAAR DE INVLOED VAN

METAFOREN OP HET LEREN

Proefschrift

ter verkrijging van de graad van

doctor in de sociale wetenschappen

aan de Katholieke Hogeschool Tilburg

op gezag van

de rector magnificus, prof.dr. G.M. van Veldhoven,

in het openbaar te verdedigen

ten overstaan van

een door het college van decanen aangewezen commissie

in de aula van de Hogeschool

op donderdag 15 januari 1981

te 16.15 uur precies

door

PETER ROBERT JAN SIMONS

geboren te Djakarta

Van Spaendonck drukkerij, Tilburg, december 1980

BIBLIOTHEEK DER
RIJKSUNIVERSITEIT
UTRECHT

Promotor: Prof.Dr. L.F.W. de Klerk

RIJKSUNIVERSITEIT UTRECHT



0241 7800

gec

VERGELIJKENDERWIJS :
ONDERZOEK NAAR DE INVLOED VAN
METAFOREN OP HET LEREN

Proefschrift

ter verkrijging van de graad van
doctor in de sociale wetenschappen
aan de Katholieke Hogeschool Tilburg

op gezag van

de rector magnificus, prof.dr. G.M. van Veldhoven,

in het openbaar te verdedigen

ten overstaan van

een door het college van decanen aangewezen commissie

in de aula van de Hogeschool

op donderdag 15 januari 1981

te 16.15 uur precies

door

PETER ROBERT JAN SIMONS

geboren te Djakarta

Van Spaendonck drukkerij, Tilburg, december 1980



VERKEER EN DE REIS

WINKELMAN VAN DE WINKELMAN
WINKELMAN VAN DE WINKELMAN

Proefschrift

dat verkrijging van de graad van
doctor in de sociale wetenschappen

aan de Katholieke Hogeschool Tilburg

door

de heer mr. J. M. van Veldhoven

in het openbaar te verdedigen

ten overstaan van

aan door het college van hoogleeraars aangewezen commissie

te de ziele van de hoogleeraar

op donderdag 15 januari 1951

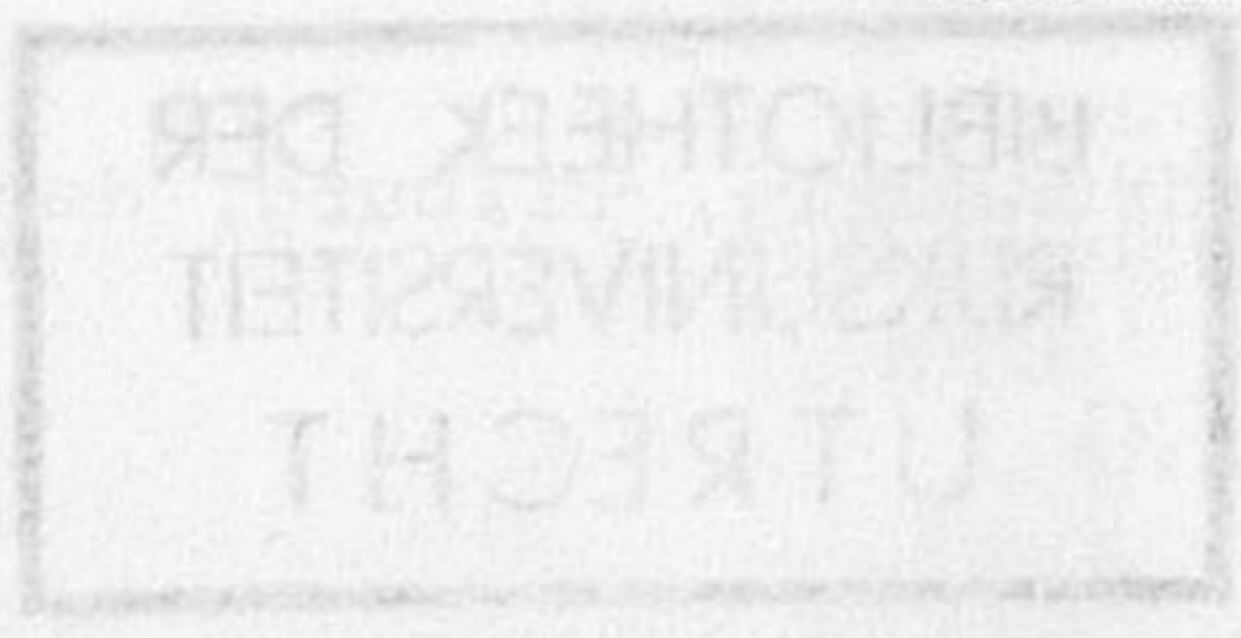
te 10.15 uur precies

door

PETER ROBERT JAN SIMONS

voorzitter van de commissie

Voor mijn vader



WOORDEN VAN DANK

Het is onmogelijk alle mensen die aan de totstandkoming van dit proefschrift hebben bijgedragen te vermelden. Een promovendus heeft vaak meer weg van een manager dan van een individueel werkend klerk. De enig echt individuele fase in de wordingsgeschiedenis van deze dissertatie was die van het schrijven van de eerste versie van dit boek. In alle andere fasen werkte ik samen met collega's, studenten, stagiaires en student-assistenten. Al deze mensen dank ik voor hun inbreng.

Mijn promotor dank ik voor de begeleiding die hij mij heeft gegeven bij het opzetten en uitvoeren van de onderzoeken, maar vooral voor de intensieve en kritische aandacht die hij aan de rapportage heeft besteed.

Veel dank ben ik ook verschuldigd aan Prof.Dr. E.J. Bijnen, die enkele methodologische en onderzoekstechnische gedeelten van dit proefschrift becommentariëerde, en aan Prof.Dr. R.A. de Moor die mij vooral veel leerde over leesbaarheid en stijl van schrijven.

Vakzwager Hans Lodewijks dank ik voor de vele stimulerende discussies die wij over ons vak in het algemeen en over mijn onderzoeksproject in het bijzonder hebben gevoerd en voor het beschikbaar stellen van les- en toetsmateriaal.

Johan van der Sanden ben ik erkentelijk voor zijn steun bij het verwerken van de gegevens en bij het verbeteren van de tekst. Frank Goyarts dank ik voor de minitieuze arbeid die hij verrichtte bij het samenstellen van de literatuurlijst en bij het corrigeren van typefouten.

De experimenten werden opgezet en uitgevoerd in samenwerking met de volgende doctoraalstudenten: J. Biskop, W. Burgmans, P. Domen, T. van Erve, R. Feijen, A. Fijten, R. Janssen, R. Kleinmoedig, M. Koene, H. Kreutzer, A. Timmermans, J. Tooten, E. Truin, L. Wagemans en A. Zagers. Ik dank hen voor de enthousiaste wijze waarop zij in mijn project hebben meegewerkt.

Drs. G. van de Veer van de Vrije Universiteit van Amsterdam dank ik voor het beschikbaar stellen van zijn vertaling van de

Spy Ring History Test en ook Dr. M. Boekaerts ben ik erkentelijk voor het beschikbaar stellen van haar testmateriaal.

De volgende scholen dank ik voor hun bereidheid mee te werken aan de onderzoeken: Aloysius Mavo, Angela Mavo, St. Canisius Mavo, Hart van Brabant Mavo, Heuvelse Maria Mavo, St. Joseph Mavo, St. Odulphus Lyceum, Paulus Lyceum.

Voor het typen van instructies, lesmateriaal, toetsen en eerdere versies van dit proefschrift ben ik dank verschuldigd aan Mevr. R. Bergmans, Mevr. C. Grbič, Mevr. T. Lodewijks, Mevr. H. v.d. Hout en Mej. W. Boers. Mej. A. Craane ben ik erg dankbaar voor het lenen van haar typemachine. Het voorliggende boek werd - op uitstekende wijze en met onvoorstelbaar weinig fouten - getypt door Mevr. I. Maas. De figuren werden getekend door A. van Helfteren en de tekening op de voorpagina is van Kaj Surink.

Last but not least bedank ik Marijke voor haar morele steun en haar kritische vragen.

INHOUDSOPGAVE

	<u>pagina</u>
Woorden van dank	1
<u>I. Theoretisch deel</u>	
1. <u>Inleiding</u>	7
1.1 Het project "informatieverwerkingsprocessen en onderwijs"	8
1.2 Enkele algemene uitgangspunten met betrekking tot onderwijspsychologisch onderzoek	12
1.3 Het deelproject "de betekenissen van metaforen in het onderwijs"	14
2. <u>Aptitude-treatment-interaction: Op zoek naar een afleidingsstrategie voor zinvolle interacties</u>	17
2.1 Enkele wijd verbreide misvattingen met betrekking tot ATI	18
2.2 Waarom ATI-onderzoek?	24
2.3 Startpunten voor een ATI-theorie	27
2.4 Acht beginselen voor ATI-onderzoek	34
2.5 De correspondentie-analysestrategie	44
2.6 De correspondentie-analysestrategie in deze dissertatie	52
3. <u>Theoretische functies van metaforen</u>	54
3.1 Vergelijkend leren: Definiëring, categorisering en een procesbeschrijving	54
3.2 Waarom metaforen belangrijk zijn: theoretische functies	57
3.2.1 Ortony's argumenten	57
3.2.2 Metaforen en schematheorieën	61
3.2.3 Metaforen en Ausubel's subsumptietheorie	63
3.2.4 Metaforen en de assimilatie-coderingstheorie van Mayer	68
3.2.5 Metaforen en de generatieve theorie van Wittrock	70
3.2.6 Metaforen en de conversatietheorie van Pask	71
3.2.7 Metaforen en de "hypostatization"-hypothese van Davidson	72
3.2.8 Samenvatting	74
3.3 Argumenten tegen metaforen	75
3.4 Conclusie	78

	<u>pagina</u>
4. <u>Empirisch onderzoek naar vergelijkend leren</u>	80
4.1 Onderzoek naar vergelijkend leren	80
4.1.1 Comparative organizers	80
4.1.2 Concrete modellen, analogieën en spelen	82
4.1.3 Verbale analogieën	87
4.1.4 De invloed van een relatief eenvoudige tekst op het leren van een analoge abstracte tekst	89
4.1.5 De rol van analogieën bij het oplossen van problemen	92
4.1.6 Spontaan gebruik van analogieën	94
4.1.7 Overig onderzoek naar metaforen en analogieën	94
4.2 Evaluatie van het tot nu toe verrichte onderzoek naar vergelijkend leren	95
II. <u>Methodologisch deel</u>	100
5. <u>Onderzoekstechnische problemen en de verwerking van gegevens van Aptitude-Treatment-Interaction research</u>	100
5.1 Inleiding	100
5.2 Methodologische aanbevelingen en de rol van de voortoets	102
5.2.1 Methodologische aanbevelingen met betrekking tot pre-instructievariabelen	102
5.2.2 De betekenis van de voortoets in het pretest-posttest-control-group-design	107
5.3 Het verwerken van gegevens van ATI-experimenten door middel van multiple regressie-analyse	112
5.4 Analyseschema	127
6. <u>Het controleren van studietijden</u>	129
6.1 Methoden om studietijden te controleren	129
6.2 Voor- en nadelen van de verschillende studietijdcontrolemethoden	131
6.3 Indirecte studietijdinvloeden	134
6.4 Het kiezen van een studietijdcontrolemethode: Vijf verschillende vragen	138
6.5 Oplossingen voor de gesignaleerde problemen	142
6.6 De studietijdcontrolemethoden in deze dissertatie	144

	<u>pagina</u>
III. <u>Onderzoeksdeel</u>	148
7. <u>Drie experimenten: Invloeden van metaforen op studieprestaties en studietijden bij verschillende leerstijlen</u>	148
7.1 Experiment 1: Operation learning versus comprehension learning	148
7.1.1 Vraagstellingen en hypothesen	148
7.1.2 Operation versus comprehension learning: Theorie en onderzoek	151
7.1.3 Methode	153
7.1.4 Resultaten van experiment 1	161
7.1.5 Discussie	174
7.2 Experiment 2: Een replicatie bij een heterogener samengestelde steekproef	179
7.2.1 Doelstelling, hypothesen en vraagstellingen	179
7.2.2 Methode	180
7.2.3 Resultaten van experiment 2	184
7.2.4 Discussie	195
7.3 Experiment 3: Visualizers versus verbalizers	199
7.3.1 Doelstelling, vraagstellingen en hypothesen	199
7.3.2 Visualizers versus verbalizers: Theorie en onderzoek	201
7.3.3 Methode	207
7.3.4 Resultaten van experiment 3	212
7.3.5 Discussie	220
7.4 Samenvatting	225
8. <u>Drie experimenten: Generalisatie naar ander materiaal en andere populaties; verschillen tussen verschillende soorten metaforen</u>	228
8.1 Experiment 4: Concrete metaforen versus abstracte comparative organizers	229
8.1.1 Vraagstellingen en hypothesen	229
8.1.2 Methode	232
8.1.3 Resultaten van experiment 4	242
8.1.4 Discussie	253
8.2 Experiment 5: Samenhangende versus niet-samenhangende metaforen	257
8.2.1 Vraagstellingen en hypothesen	257
8.2.2 Methode	261

	<u>pagina</u>
8.2.3 Resultaten van experiment 5	264
8.2.4 Discussie	269
8.3 Experiment 6: Actieve versus passieve vergelijkingen door basisschool-leerlingen	271
8.3.1 Vraagstellingen en hypothesen	271
8.3.2 Methode	273
8.3.3 Resultaten van experiment 6	277
8.3.4 Discussie	283
8.4 Samenvatting	285
9. <u>Algemene discussie</u>	287
9.1 Theoretische implicaties	287
9.2 Vergelijking van de onderzoeksresultaten met die van voorgaande onderzoeken	290
9.3 Methodologische implicaties	292
9.4 Praktische implicaties	294
9.5 Suggesties voor verder onderzoek	299
<u>English summary</u>	303
<u>Referenties</u>	307
<u>Bijlagen</u>	325

I THEORETISCH DEEL

Hoofdstuk 1 Inleiding

In toenemende mate wordt het onderwijspsychologisch onderzoek de laatste jaren gericht op informatieverwerkingsprocessen en cognitieve structuren (zie Wittrock en Lumsdaine, 1977) in de context van taken en opdrachten die in scholen aan bod komen (zie Glaser en Resnick, 1972). Enerzijds worden door middel van empirische taakanalyses de processen opgespoord waarop schooltaken een beroep doen (bijvoorbeeld Resnick, 1976; Glaser, 1976). Anderzijds wordt door middel van empirische leerling-analyses nagegaan welke verschillen in informatieverwerkingsprocessen samengaan met verschillen tussen leerlingen (Snow, 1978; Hunt, 1977).

In dit proefschrift gaat het niet zozeer om empirische taakanalyses noch om empirische leerling-analyses, maar om rationele taakanalyses en rationele leerling-analyses (cf. Resnick, 1976). In een rationele taakanalyse beredeneert men op welke informatieverwerkingsprocessen de taak naar veronderstelling een beroep doet. In een rationele leerling-analyse beredeneert men welke verschillen in informatieverwerkingsprocessen waarschijnlijk zullen samengaan met verschillen tussen leerlingen. Naast de taakanalyse en de leerling-analyse wordt ook nog een zogenaamde situatie-analyse onderscheiden. In een (rationele) situatie-analyse beredeneert men welke functies onderwijssituaties kunnen vervullen en op welke informatieverwerkingsprocessen die situaties een beroep doen. Door de resultaten van deze drie soorten van procesanalyse met elkaar in verband te brengen, kunnen hypothesen worden afgeleid over aptitude-treatment-interacties (ATI's). Aptitude-treatment-interacties zijn onderzoeksresultaten waarin de relaties tussen leerlingkenmerken en onderwijsuitkomsten bij verschillende treatments (onderwijssituaties) verschillen (zie Cronbach en Snow, 1977).

In deel III van dit proefschrift wordt een zestal experimenten gerapporteerd waarin de hypothesen met betrekking tot aptitude-treatment-interacties zijn gebaseerd op dergelijke rationele procesanalyses. In deze experimenten is nagegaan welke effecten een bepaalde onderwijsmaatregel (i.c. het toevoegen van metaforen) kan hebben op de leerresultaten van verschillende typen leerlingen. De ATI-hypothesen die in die experimenten worden getoetst, zijn gebaseerd op veronderstellingen over de invloeden die metaforen zouden kunnen uitoefenen op de wijze waarop verschillende typen leerlingen informatie verwerken.

De experimenten zijn uitgevoerd in het kader van een onderwijspsychologisch onderzoeksproject: "De betekenis van metaforen in het onderwijs". Dit project is een onderdeel van een omvangrijker onderzoeksproject: "Informatieverwerkingsprocessen en onderwijs", dat door medewerkers en studenten van de vakgroep onderwijspsychologie van de Katholieke Hogeschool Tilburg wordt uitgevoerd. Alvorens wordt ingegaan op het deelproject worden enkele algemene uitgangspunten van het vakgroepsproject nader toegelicht.

1.1 Het project "informatieverwerkingsprocessen en onderwijs"

Reeds in 1966 schreef Cronbach dat generalisaties van de volgende vorm het resultaat zouden moeten zijn van (onderwijspsychologisch) onderzoek: "With subject matter of this nature inductive experience of this type in this amount produces this pattern of responses in pupils at this level of development."

In het handboek van Cronbach en Snow (1977) wordt de oude stelling weer van stal gehaald: "That is, investigators (are) advised to look for a fourth order interaction: Ability x Age x Subject matter x Treatment x Outcome." (p. 517).

In het project "informatieverwerkingsprocessen en onderwijs" wordt gestreefd naar generalisaties die vergelijkbaar zijn met die van Cronbach. In algemene zin worden in het project experimenten opgezet die kunnen leiden tot generalisaties als: "Bij dit type van leerstof levert dit soort onderwijsmaatregelen dit patroon van resultaten op bij dit type leerlingen."

Geven we de onderwijsmaatregelen weer met de letter O, het type leerlingen met de letter L en de resultaten van het onderwijs met de letter R, dan kan het voorgaande ook als volgt worden uitgedrukt. In het project wordt gestreefd naar kennis over interacties tussen R, O en L behorend bij een bepaald type leerstof (zie De Klerk, 1979). De hypothesen over de interacties tussen R, O en L worden gebaseerd op een analyse van R-, O- en L-variabelen in termen van de informatieverwerkingsprocessen (bijvoorbeeld: aandacht, korte termijngeheugen, codering) waarop zij een beroep doen, en van de aard van de cognitieve structuren. (Zie verder hoofdstuk 2.)

Hoewel de methodologische problemen bij onderzoek naar interacties van het type R-O-L al groot zijn wanneer hierin slechts één leerlingkenmerk, één kenmerk van onderwijsmaatregelen en één onderwijsuitkomst worden betrokken, is het uiteindelijk doel van de onderzoeken in het project te komen tot generalisaties gebaseerd op multi-O, multi-L, multi-R-interacties. Dat dergelijke complexe interacties de werkelijkheid waarschijnlijk het dichtst benaderen is uitvoerig bevestigd door Cronbach (1975) en empirisch aangetoond door Snow (1977). Ook De Klerk (1979) breekt een lans voor het multi-O, multi-L, multi-R-paradigma.

Om te voorkomen dat bij een dergelijke multi-O, multi-L, multi-R-benadering door de bomen het bos niet meer wordt gezien, is in het project een aantal beperkende maatregelen getroffen.

1. In de eerste plaats worden volgens een bepaalde strategie (zie correspondentie-analyse in hoofdstuk 2) hypothesen en alternatieve hypothesen opgesteld met betrekking tot de interacties tussen de O-kenmerken, de L-kenmerken en de R-variabelen. De hypothesen worden gebaseerd op een analyse van O-, L- en R-variabelen in termen van psychologisch inzichtelijke correspondenties ertussen (Lodewijks en Simons, 1977).
2. Alleen die interacties worden getoetst waarvoor a priori volgens de correspondentie-analysestrategie een theoretische rationale kan worden bedacht (zie het principe van theore-

tische encadrering in Lodewijks en Simons, 1979). Het is immers zinloos allerlei ingewikkelde, maar niet interpreteerbare interacties te vinden. In interactie-onderzoek moet - net zo goed als in ander onderzoek - van tevoren worden bedacht wat mogelijke uitkomsten van het onderzoek zijn en hoe deze kunnen worden geïnterpreteerd.

3. In de derde plaats worden ten aanzien van de drie componenten O, L en R beperkingen gehanteerd: Als O-variabelen worden in het project alleen de zogenaamde directe omgevingsfactoren in het onderzoek betrokken. Kenmerkend voor directe factoren is dat zij het verloop van het leerproces rechtstreeks beïnvloeden, dus zonder tussenkomst van andere variabelen (De Klerk, 1979a, p. 192; zie ook het begrip eerste orde variabelen in Van Parreren, 1978). Als L-variabelen worden in het project alleen de zogenaamde psychologische leerlingkenmerken betrokken. Kenmerkend voor psychologische leerlingkenmerken is dat zij een psychologisch inzichtelijke relatie tot leerprocessen hebben. Voorbeelden van psychologische leerlingkenmerken zijn intelligentie, cognitieve stijl en faalangst, terwijl geslacht, milieu en ras voorbeelden van niet-psychologische leerlingkenmerken vormen (De Klerk, 1979a). Als R-variabelen worden in het project vier soorten afhankelijke variabelen opgenomen: kennistoetsen, begripstoetsen, transfertoetsen en attitudemetingen. Bij kennistoetsen wordt nagegaan in hoeverre de leerstof in het lange-termijn-geheugen is opgeslagen. Bij begripstoetsen wordt nagegaan in hoeverre de leerstof is verwerkt, dat wil zeggen is geïntegreerd in het netwerk van bestaande kennis op het betreffende gebied. Bij transfertoetsen wordt nagegaan in hoeverre de leerling in staat is de verwerkte informatie te gebruiken of toe te passen. Met behulp van attitude-vragenlijsten wordt de mening van de leerling gevraagd over het gevolgde onderwijs.
4. Een vierde beperking betreft vooral de O-kenmerken. In plaats van in elk experiment diverse onderwijskenmerken te variëren, wordt steeds slechts één onderwijskenmerk gevariëerd.

De experimenten vormen derhalve eerder mono-O, multi-L, multi-R-onderzoeken, waarbij één onderwijskenmerk wordt gecombineerd met verschillende leerlingkenmerken en verschillende onderwijsuitkomsten. Pas in latere instantie, wanneer vanuit de diverse mono-O-experimenten meer informatie beschikbaar is gekomen, kunnen ook multi-O, multi-L, multi-R-experimenten worden opgezet. Om toch variatie in de O-kenmerken te krijgen, is het totale project gesplitst in drie deelprojecten, met als indelingscriterium de aard van het onderzochte onderwijskenmerk. Er worden drie soorten van directe onderwijskenmerken onderscheiden: pre-instructiekenmerken, instructiekenmerken en post-instructiekenmerken. In tabel 1.1 zijn voorbeelden van deze kenmerken opgenomen.

Tabel 1.1 Drie soorten van onderwijskenmerken

pre-instructiekenmerken	instructiekenmerken	post-instructiekenmerken
- advance organizers	- structuur van de leerstof	- feedback
- doelstellingen	- sequentie van aanbidding	- post-questions
- pre-questions	- mate van vrijheid	- samenvattingen achteraf
- samenvattingen vooraf	- herhaling	
- metaforen	- plaatjes bij tekst	- opdrachten

In één deelproject wordt nagegaan welke invloed post-instructiekenmerken kunnen hebben op informatieverwerkingsprocessen en cognitieve structuren en via deze op onderwijsuitkomsten. Meer specifiek wordt in dit deelproject de invloed van diverse feedbackprocedures op genoemde variabelen onderzocht (zie De Klerk, 1979b; De Klerk en De Klerk, 1978). In het tweede deelproject wordt de invloed onderzocht van bepaalde instructievariabelen op informatieverwerkingsprocessen en cognitieve structuren en via deze op onderwijsuitkomsten. Meer specifiek wordt hier met name de invloed van verschillende typen van structurering en sequentiëring

van leerstof op de afhankelijke variabelen onder de loupe genomen (zie onder andere Lodewijks, 1978).

In het derde deelproject worden pre-instructievariabelen onderzocht. Hier gaat het derhalve om de invloed van dergelijke pre-instructievariabelen - met name van metaforen - op informatieverwerkingsprocessen en cognitieve structuren en via deze op onderwijsuitkomsten. Dit proefschrift gaat over dit laatste deelproject.

1.2 Enkele algemene uitgangspunten met betrekking tot onderwijspsychologisch onderzoek

In het voorafgaande werden impliciet reeds enkele algemene uitgangspunten geformuleerd met betrekking tot onderwijspsychologisch onderzoek zoals die gestalte kregen in het project informatieverwerkingsprocessen en onderwijs:

1. Het onderzoek moet gericht zijn op multi-O, multi-L en multi-R-relaties.
2. Dit onderzoek moet steeds vertrekken vanuit hypothesen met betrekking tot deze relaties die zijn gebaseerd op een analyse van onderwijskenmerken, leerlingkenmerken en onderwijsuitkomsten in termen van onderliggende informatieverwerkingsprocessen en cognitieve structuren.
3. Er werd een voorkeur getoond voor het beperken van het onderwijspsychologisch onderzoek tot directe omgevingsvariabelen en psychologische leerlingkenmerken.

Naast deze drie liggen aan het deelproject "de betekenis van metaforen voor het onderwijs" (en ook aan de andere deelprojecten) nog de volgende uitgangspunten met betrekking tot onderwijspsychologisch onderzoek ten grondslag:

4. Onderwijspsychologisch onderzoek moet zoveel mogelijk ecologisch valide (De Klerk et al., 1978; De Klerk, 1979a; Bronfenbrenner, 1976) en representatief zijn (Snow, 1974). De Klerk (1979a, p. 232) omschreef "ecologisch onderzoek" als volgt: "...onder ecologisch onderzoek verstaan wij hier empirisch onderzoek dat plaatsvindt in natuurlijke onder-

wijssituaties. Kenmerkend voor deze situaties is de complexiteit. Gewoonlijk zijn er in een realistische situatie veel factoren aanwezig. Sommige versterken elkaar, anderen werken elkaar tegen. Het is meestal onbekend wat de invloed van alle factoren te zamen is op het verloop van het leerproces. In een onderzoek worden vaak twee of slechts enkele factoren tegelijkertijd onderzocht. Als men de invloed van alle overige factoren zou uitschakelen, dan is het niet zeker of dezelfde uitkomsten ook in reële onderwijssituaties verkregen zouden zijn. In ecologisch onderzoek laat men alle factoren zoveel mogelijk intact. Ecologisch onderzoek verhoogt de externe validiteit van het onderzoek...". In een ecologisch onderzoek heeft men niet te maken met een proefleider, proefpersonen, condities en stimulusmateriaal, maar met een docent, leerlingen, klassen en lesmateriaal.

5. Onderwijspsychologisch onderzoek moet zoveel mogelijk intern valide zijn (Campbell en Stanley, 1963). De interne validiteit heeft te maken met de vraag, in hoeverre de uitkomsten van een experiment inderdaad kunnen worden toegeschreven aan de opzettelijk gemanipuleerde (onafhankelijke) variabelen. Interne validiteit - "the basic minimum without which any experiment is uninterpretable" - kan men bereiken door Kerlinger's maxmincon-principe aan te houden:
- a. maximalisatie van de variantie die het gevolg is van de voor de hypothesen relevante variabelen;
 - b. minimalisatie van toevallige individuele verschillen tussen proefpersonen en van meetfouten en
 - c. controlering van allerlei toevallige, buiten het experiment gelegen factoren (zie De Klerk et al., 1978).

Om de interne validiteit van een experiment te verhogen, moet men als onderzoeker soms maatregelen treffen die de ecologische validiteit iets verlagen. Steeds zal men moeten zoeken naar een evenwicht tussen de ecologische en de interne validiteit. Om de interne validiteit van het onderzoek dat in deel III van dit proefschrift wordt gerapporteerd te verhogen, werden de volgende maatregelen genomen:

- a. De onderzoeken vonden meestal plaats in localiteiten van de hogeschool en niet op de scholen. Dit bood de onderzoeker de gelegenheid meer controle uit te oefenen over de gang van zaken tijdens het experiment.
 - b. De rol van de leerkracht werd geminimaliseerd door de instructies op papier te zetten. Bovendien vervulden getrainde proefleiders de rol van leerkracht. Ook dit maakte een betere controle over toevallige factoren mogelijk.
 - c. De schoolklassen werden in twee delen gesplitst, waarbij de leerlingen aselekt aan de verschillende condities werden toegewezen.
6. Een laatste uitgangspunt betreft de criteria die worden gehanteerd bij het bepalen van eventuele effecten van onderwijs. Te vaak worden in onderwijspsychologisch onderzoek (cf. Faw en Waller, 1976) slechts de resultaten betrokken van toetsen die vrijwel onmiddellijk na afloop van de instructie zijn afgenomen. Naar onze mening is het echter niet zozeer van belang wat een leerling onmiddellijk na de instructie weet of kan als wel wat hij/zij na verloop van enkele dagen of weken nog heeft onthouden. Daarom zijn in de onderzoeken - enkele weken na de instructie - retentietoetsen afgenomen.

1.3 Het deelproject "de betekenis van metaforen in het onderwijs"

Metaforen zijn van oudsher beproefde onderwijskundige hulpmiddelen om het begrijpen en onthouden van informatie te faciliteren. Het zijn echter ook verguisde hulpmiddelen. In de praktijk van het onderwijs en in de literatuur treffen we dan ook fervente voorstanders maar ook fervente tegenstanders van het gebruik van metaforen aan. Voorstanders wijzen op allerlei nuttige functies die metaforen kunnen vervullen, terwijl tegenstanders wijzen op de verwarring die metaforen kunnen zaaien, of op hun overbodigheid.

Uit een literatuuronderzoek bleek dat het hier in feite een woordenstrijd betrof, een academische discussie. Argumenten pro en contra werden ontleend aan persoonlijke ervaringen, vooroordelen, theoretische noties en dergelijke. Empirisch onderzoek ontbrak (zie Simons, 1978).

Het doel van deze studie is dan ook om op grond van empirisch onderzoek inzicht te krijgen in de functies van metaforen bij het leren en onthouden van lesstof. Hierbij wordt er van uitgegaan, dat het effect van metaforen afhankelijk is van bepaalde leerlingkenmerken. De algemene vraagstelling van het deelproject werd dan ook als volgt omschreven:

"Voor welke typen van leerlingen vergemakkelijken welke soorten van metaforen het begrijpen en onthouden (op korte en lange termijn) van lesstof."

Om antwoorden op deze vraag te verkrijgen, werd aptitude-treatment-interaction-onderzoek opgezet. Bij dit type onderzoek doen zich methodologische en theoretische problemen voor, waarop in hoofdstuk 2 van dit proefschrift wordt ingegaan. Aan het eind van dat hoofdstuk wordt een door Lodewijks en Simons (1979) ontwikkelde strategie (correspondentie-analyse) als oplossing voor een aantal van die problemen naar voren gebracht.

In hoofdstuk 3 wordt ingegaan op de argumenten van voor- en tegenstanders van metaforen in het onderwijs. Tevens wordt de rol die metaforen spelen of zouden kunnen spelen in een aantal onderwijspsychologische theorieën (onder andere van Pask, Mayer en Ausubel) geanalyseerd. Deze analyse leidt tot een aantal potentiële functies die metaforen in een onderwijsleerproces kunnen vervullen. Deze functies spelen later een belangrijke rol bij het opstellen van de ATI-hypothesen volgens de principes van de correspondentie-analyse.

In hoofdstuk 4 wordt aandacht besteed aan empirisch onderzoek naar de betekenis van metaforen in het onderwijs. Zoals reeds werd vermeld, was bij de start van het project nauwelijks enige empirische evidentie voorhanden. Recentelijk zijn er echter diverse onderzoeken gepubliceerd en deze worden in hoofdstuk 4 besproken.

Hoofdstuk 5 gaat over methodologische aanbevelingen van enkele schrijvers van overzichtsartikelen over onderzoek naar pre-instructievariabelen. Uitvoerig wordt ingegaan op de statistische verwerking van ATI-gegevens door middel van regressie-analyse.

In hoofdstuk 6 wordt ingegaan op het probleem van de studietijdcontrole. Immers alle pre-instructievariabelen vragen extra studietijd. Deze extra tijd levert niet alleen onderzoekstechnische, maar ook theoretische problemen op. Een eigen onderzoek, dat beknopt wordt gerapporteerd, werpt een nieuw licht op de studietijdproblematiek. Een theoretische analyse resulteert in een aantal nieuwe methoden om het studietijdprobleem op te lossen.

In hoofdstuk 7 worden drie (veld)experimenten gerapporteerd waarin werd nagegaan welke invloed metaforen hebben op het begrijpen en onthouden van electriciteitsleerbegrippen door middelbare scholieren. In deze experimenten werden verschillende van de nieuwe studietijdcontrolemethoden beproefd en met elkaar vergeleken. Bij de leerlingen werden onder andere leerstijltests en coderingsstijltests afgenomen.

In hoofdstuk 8 wordt eerst een correlationele studie gerapporteerd, waarin is nagegaan of de in de eerste onderzoeken gerezen twijfels met betrekking tot de coderingsstijltest gegrond waren. Meer specifiek werd nagegaan of de coderingsstijltest een voorkeur voor een bepaalde manier van coderen meet of een vaardigheid in die coderingswijze. Vervolgens worden drie experimenten gerapporteerd. Enerzijds hadden deze tot doel na te gaan of de in de eerste experimenten gevonden resultaten konden worden gegeneraliseerd naar andere lesstof (biologie, computerprogrammeren, ontwikkelingspsychologie). Anderzijds werd in deze experimenten informatie verzameld over effecten van verschillende soorten metaforen bij verschillende typen leerlingen. Zo werden vergeleken:

1. samenhangende versus niet-samenhangende metaforen;
2. actieve verwerking van metaforen versus passieve verwerking en
3. metaforen versus "comparative organizers".

In hoofdstuk 9, tenslotte, volgt een algemene discussie, waarin de verschillende resultaten in verband worden gebracht met elkaar en met andere onderzoeken.

Hoofdstuk 2 Aptitude-treatment-interaction: Op zoek naar een afleidingsstrategie voor zinvolle interacties¹⁾

In dit hoofdstuk wordt aandacht besteed aan de theorievorming rond aptitude-treatment-interactie (ATI). ATI is (in de onderwijspsychologie) het algemene idee dat onderwijsuitkomsten het gemeenschappelijk product zijn van leerlingkenmerken ("any characteristic of a person that forecasts his probability of success under a given treatment", Cronbach en Snow, 1977, p. 6) en omgevingskenmerken ("any manipulable variable" (op.cit. p. 6)).

Bij dit algemene idee behoren zogenaamde ATI-onderzoeksmethoden via welke men kan nagaan of interacties tussen leerlingkenmerken en onderwijsmethoden variaties in onderwijsuitkomsten verklaren. Zij vormen een combinatie van experimentele en correlatieve methoden (zie De Klerk, 1978, 1979c). De resultaten van het tot nu toe verrichte ATI-onderzoek zijn over het algemeen teleurstellend (zie bijvoorbeeld Crombag, 1979; Span, 1977). Er zijn recentelijk dan ook geluiden te horen, dat voortgaan met de ATI-conceptie weinig vruchtbaar is (Crombag, 1979; Merrill, 1975). Op de argumenten van de tegenstanders van ATI-onderzoek en op enkele misvattingen wordt in 2.1 nader ingegaan. In 2.2 wordt beargumenteerd waarom het toch nog zinvol is om door te gaan met ATI-onderzoek. De discussie daaromtrent mondt uit in een pleidooi voor een meer theoriegerichte-conceptuele benadering van ATI. In 2.3 worden enkele in de literatuur aangetroffen eerste stappen in de richting van een theoriegerichte ATI-benadering besproken. In 2.4 worden deze ideeën geïntegreerd in en uitgebreid tot een achttal beginselen. In 2.5 volgt een toepassing van deze beginselen in een heuristische strategie die door Lodewijks en Simons (1979)

1) Dit hoofdstuk is een gemeenschappelijk product van de auteur van deze dissertatie en Drs. J.G.L.C. Lodewijks. In de dissertatie van Lodewijks komt dit hoofdstuk in vrijwel dezelfde bewoordingen voor als hoofdstuk 7. Het in dit hoofdstuk besprokene is een verdere uitwerking van gedachten die eerder zijn gepubliceerd door Lodewijks (1979) en - ook als coproductie - door Lodewijks en Simons (1979).

werd ontwikkeld: correspondentie-analyse. In 2.6 wordt vervolgens verder ingegaan op de toepassing van de correspondentie-analysestrategie in deze dissertatie.

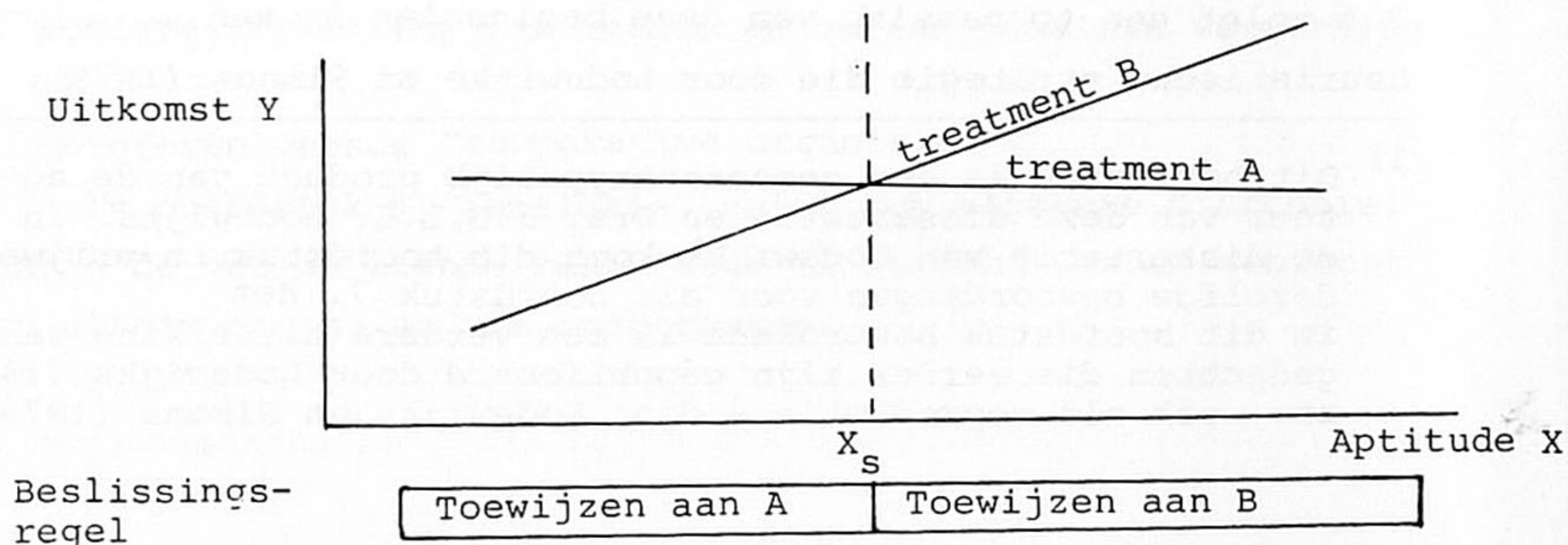
2.1 Enkele wijd verbreide misvattingen met betrekking tot ATI

In de loop der jaren hebben zich binnen en buiten de kring van ATI-onderzoekers misvattingen gevormd die de discussie over ATI vertroebelen. In deze paragraaf worden vijf van dergelijke misvattingen behandeld.

a. Resultaten van ATI-onderzoek bepalen hoe differentiatie moet plaatsvinden

Dit misverstand is mede in de hand gewerkt door een te groot enthousiasme van de ATI-onderzoekers van het eerste uur. Zij meenden namelijk, dat uit de gevonden interacties min of meer rechtstreeks zou volgen voor welke leerlingen welke methode zou moeten worden gebruikt. Dat ook Cronbach en Snow (1977) in hun vroegere werk deze overdreven verwachting ten aanzien van de resultaten van ATI-onderzoek mede in de hand hebben gewerkt, moge blijken uit bijvoorbeeld een tekening als figuur 2.1, waarin een interactie tussen twee "treatments" en een "aptitude" is getekend met onder de figuur direct de beslissingsregel. Personen met een score boven X_s zouden moeten worden toegewezen aan treatment B en personen met een score beneden X_s zouden moeten worden toegewezen aan treatment A.

Figuur 2.1 Een aptitude-treatment-interactie plus een beslissingsregel



De directe stap van de interactie (descriptie) naar de beslissingsregel (prescriptie) lijkt simpel en voor de hand te liggen. Zo'n stap mag echter niet zonder meer worden gemaakt. In de eerste plaats doet zich hierbij een methodologisch probleem voor. Er is immers een gebied op de X-as waar de treatments niet significant van elkaar verschillen: een gebied rondom X_s . Alleen voor de zogenaamde differentiatiegebieden van X, waar de treatments significant van elkaar verschillen (Plomp, 1977), zou een beslissing ten gunste van A of B kunnen worden genomen. In het gebied van X (rondom X_s) waar geen significante verschillen tussen A. en B bestaan, is toewijzing aan A even goed als toewijzing aan B.

In de tweede plaats is het onverstandig om toewijzingsbeslissingen uitsluitend te baseren op prestatiecriteria. Er zijn kosten en baten verbonden aan de realisering van onderwijsmethoden in het onderwijs. In hun latere werk adviseren Cronbach en Snow (1977, p. 40) dan ook toewijzingsbeslissingen te baseren op utiliteiten, waarbij op grond van verschillende criteria (prestaties, kosten, studietijd en dergelijke) volgens bepaalde wegingsfactoren wordt bepaald welke leerlingen aan welke methode moeten worden toegewezen.

In de derde plaats kunnen met betrekking tot verschillende prestatiecriteria verschillende interacties optreden. Op welke interactie moet dan de beslissingsregel worden gebaseerd? Zo bleek uit verschillende studies dat (faal)angstige leerlingen meer profiteren van gestructureerde methoden (Flammer, 1975; De Leeuw, 1979). Het is echter waarschijnlijk dat, wanneer (faal)angstige leerlingen op grond van deze resultaten alleen nog maar gestructureerde onderwijsmethoden zouden worden aangeboden, de (faal)angst in andere (ongestructureerde) situaties zich zou bestendigen of zou toenemen. Het is derhalve maar zeer de vraag of het wel zo gunstig is de (faal)angstige leerlingen aan gestructureerde methoden toe te wijzen.

Concluderend kan worden gesteld, dat kennis van aptitude-treatment-interacties slechts via allerlei tussenstappen een

bijdrage kan leveren aan prescripties met betrekking tot het toewijzen van leerlingen aan onderwijsmethoden. De oorspronkelijke misvatting dat dit verband tussen descriptieve gegevens en prescripties rechtstreeks zou kunnen worden gelegd, heeft de overige misvattingen b, c, d en e mede in de hand gewerkt (vergelijk Lodewijks, 1979).

b. Alleen disordinale interacties zijn interessant en van belang

"At one time, writers stressed the value of disordinal interactions and tended to dismiss ordinal interactions. The ordinal interaction was regarded as a mere artifact of the choice of the measuring scale for the dependent variable" (Cronbach en Snow, 1977, p. 31). Vooral Bracht (1970) heeft sterk de nadruk gelegd op het belang van disordinaliteit van interacties. Hij vond ordinale interacties even onbelangrijk als niet-significante interacties. Ook Borich et al. (1976) hechtten weinig waarde aan ordinale interacties. Deze opvattingen over het uiteenlopend belang van ordinale en disordinale interacties hangen nauw samen met de eerder besproken misvatting met betrekking tot de relaties tussen ATI-resultaten en toewijzingsbeslissingen. In een opvatting volgens welke het kruispunt van twee regressielijnen rechtstreeks aangeeft welke leerlingen nog wel aan methode A moeten worden toegewezen en welke niet meer, passen alleen disordinale interacties. Bij ordinale interacties zouden immers alle leerlingen aan dezelfde methode moeten worden toegewezen. Wanneer men zich echter realiseert, dat nog vele andere factoren mede bepalend zijn bij toewijzingsbeslissingen en dat niet het kruispunt van de regressielijnen maar de differentiatiegebieden de relevante informatie vormen, worden naast disordinale ook ordinale interacties interessant. Cronbach en Snow (1977) geven het volgende voorbeeld ter verduidelijking: "If the treatment that yields the greater outcome is much costlier than the other, the ordinal interaction effect on outcome becomes a disordinal effect on payoff. The more costly method should be applied only to

those who will find it so advantageous that its extra cost is repaid in benefits." (p. 33). Soortgelijke voorbeelden zou men kunnen geven met betrekking tot studietijden of combinaties van verschillende criteria. Met Cronbach en Snow concluderen we dan ook: "We take ordinal interaction seriously." (p. 495).

c. Er zijn nauwelijks belangwekkende ATI's gevonden

In veel publicaties over ATI worden de volgende woorden van Glass (1970) aangehaald: "There is no evidence for an interaction of curriculum treatments and personological variables. I don't know of another statement that has been confirmed so many times by so many people..." (p. 210). Deze uitspraak van Glass is gebaseerd op het overzichtsartikel van Bracht (1970), die, zoals boven vermeld, alleen was geïnteresseerd in disordinale interacties. Wanneer we het overzicht van ATI-studies van Cronbach en Snow (1977) bezien, inclusief de ordinale interacties, dan is de conclusie van Glass niet gerechtvaardigd. Er zijn zeker belangwekkende ATI's gevonden. Snow (1977) schrijft: "...the existence of ATI as phenomena has been clearly established" (p. 50). Het probleem zit dus niet in het feit dat er geen ATI's zouden zijn gevonden. Eerder vormen de tegenstrijdige, vaak moeilijk verklaarbare resultaten van interactie-onderzoek een probleem. In de woorden van Snow: "But while some ATI findings are plausible and some are replicable, few are well understood and none are yet applicable to instructional practice." (p. 50).

d. Mastery Learning, Personalized Systems of Instruction en Learner Control vormen aantrekkelijke alternatieven voor ATI-onderzoek

De Koning (1974) en Crombag (1979) stelden voor de aandacht meer te richten op het verder ontwikkelen en verbreiden van Mastery Learning (beheersings-leren) en Personalized Systems of Instruction (PSI) en af te zien van verder ATI-

onderzoek, omdat de resultaten daarvan teleurstellend waren. Merrill (1975) deed een vergelijkbaar voorstel, maar dan met betrekking tot Learner Control (laat leerlingen zelf kiezen welke vorm van onderwijs zij wensen). Uit deze voorstellen spreekt een misvatting omtrent ATI, namelijk dat ATI betrekking zou hebben op een onderwijsmethode waarbij wordt gedifferentieerd naar psychologische leerlingkenmerken (De Klerk, 1979c).

Zo'n onderwijsmethode bestaat echter (nog) niet en de status van de ATI-resultaten is niet van dien aard dat er een basis is voor zo'n methode. ATI is geen onderwijsmethode, maar een onderzoeksparadigma en onderwijsmethoden als beheersings-leren, PSI en Learner Control kunnen geen alternatieven zijn voor een onderzoeksmethode (zie De Klerk, 1979c). Bovendien - en dat wordt al te gemakkelijk vergeten - liggen ook aan genoemde onderwijsmethoden (merendeels ongetoetste) ATI-hypothesen ten grondslag. Zo hoort bij de Mastery Learning-methode bijvoorbeeld de (ordinale) ATI-hypothese, dat een onderwijsmethode die gebreken in de voorkennis van leerlingen opheft, alvorens tot de volgende stap wordt overgegaan, tot betere resultaten leidt bij leerlingen met voorkennisgebreken, dan een methode die de voorkennisgebreken laat bestaan.

Het is ons inziens van belang de resultaten van ATI-onderzoek niet in de eerste plaats aan praktische criteria te toetsen (zoals "Zijn er ATI-gegevens die de basis kunnen vormen voor differentiatiemaatregelen?", of "Is de differentiatiemaatregel die uit deze ATI-hypothese volgt uitvoerbaar?"), maar aan theoretische criteria (bijvoorbeeld "Levert dit ATI-resultaat een bijdrage aan een instructietheorie?" of "Vergroot dit ATI-resultaat het inzicht in leerprocessen?").

e. Alleen stabiele, onveranderlijke, algemene kenmerken van mensen mogen als leerlingkenmerk worden opgenomen in een ATI-onderzoek

Crombag (1979) en Merrill (1975) maakten bezwaar tegen de soort van leerlingkenmerken die gewoonlijk in ATI-studies worden opgenomen. Crombag (op.cit.) gelooft niet zo in vermogens: "Abilities do not predict performance any better than performance predicts performance" (p. 181). Merrill (op.cit.) gelooft dat "...the multidimensional dynamic state aptitudes which change from moment to moment are more likely to determine the treatment that is most appropriate for a given individual at a given moment in time" (p. 220). Beide auteurs verwijten Cronbach en Snow dat ze alleen stabiele, onveranderlijke, algemene persoonskenmerken willen opnemen in ATI-onderzoek.

Nu is het inderdaad zo dat Cronbach en Snow (1977) op diverse plaatsen in hun boek een voorkeur tonen voor de meer stabiele, onveranderlijke, algemene kenmerken (zie bijvoorbeeld p. 160). Deze voorkeur komt waarschijnlijk voort uit hun wens praktisch toepasbare ATI-resultaten te kunnen leveren (zie misvatting a), waarvoor het noodzakelijk is algemeen geldige uitspraken te kunnen doen. Op andere plaatsen blijkt deze voorkeur echter minder sterk: "aptitude" wordt bijvoorbeeld zeer ruim gedefinieerd als "any characteristic of a person that forecasts his probability of success under a given treatment" (Cronbach en Snow, 1977, p. 6). Verder wordt ook uitgebreid aandacht gegeven aan één van de meest veranderlijke kenmerken van mensen: de voorkennis.

Zowel de vermelde ideeën van Crombag en Merrill als die van Cronbach en Snow moeten ons inziens worden gezien als persoonlijke opvattingen die in onvoldoende mate empirisch kunnen worden gerechtvaardigd. Wanneer men het idee verlaat dat de resultaten van ATI-onderzoek onmiddellijk toepasbaar moeten zijn in de vorm van differentiatie modellen (misvatting a), dan zijn alle soorten kenmerken even geschikt.

2.2 Waarom ATI-onderzoek?

In het voorgaande werden enkele misvattingen behandeld die er mede toe hebben geleid, dat sommigen het vertrouwen in ATI-onderzoek verloren (Crombag, 1979; Merrill, 1975). Er zijn echter ook valide redenen voor een geschokt vertrouwen in ATI-onderzoek: Het aantal interacties waaruit praktische toepassingen kunnen volgen, is inderdaad uitermate beperkt (Snow, 1977). Vele van de gevonden interacties worden in replicatie-onderzoek niet teruggevonden. Andere gevonden interacties zijn onverenigbaar, onverwacht en niet interpreteerbaar. En dan zijn er nog talloze studies waar verwachte interacties niet bleken op te treden.

Voorts blijkt dat gevonden ATI-effecten vaak weinig stabiel zijn. Verschillende interacties blijken vakspecifiek, veranderen, respectievelijk verdwijnen door kleine veranderingen in de instructiemaatregelen en blijven vaak, ook in verschillende fasen van één onderwijsleereenheid, niet stabiel (vergelijk Kanselaar, 1979). De teleurstellende stand van zaken met betrekking tot ATI-onderzoeksresultaten tot nu toe geeft ons echter, om de volgende redenen, geen aanleiding om het ATI-onderzoek op te geven.

In de eerste plaats zijn er legio methodologisch-statistische oorzaken aan te geven die voor de teleurstellende stand van zaken verantwoordelijk kunnen worden gesteld. Cronbach en Snow (1977) hebben in nogal wat studies methodologische fouten aangetoond (verkeerde transformaties, verkeerde toetsen, enzovoorts). Bovendien bleek het onderscheidingsvermogen van de statistische toetsen waarmee de interacties moesten worden aangetoond in de meeste gevallen te gering door te kleine aantallen proefpersonen, waardoor de kans dat een interactie ten onrechte niet significant werd bevonden zeer groot was. Vaak werden de data op een wijze geanalyseerd die de kans op het vinden van interacties er niet groter op maakte. Zo werd dikwijls gebruik gemaakt van variantie-analytische proefopzetten, waarbij een leerlingkenmerk als attribuutvariabele werd opgenomen. De leerlingengroep werd daarbij gedichotomiseerd of ten hoog-

ste getrichotomiseerd. Een regressie-analyse zou echter meer op zijn plaats zijn geweest, omdat daarbij meer informatie van het leerlingkenmerk kan worden gebruikt (met als gevolg een grotere kans op het vinden van interacties).

In de tweede plaats is er ook inhoudelijk gezien nogal wat kritiek te leveren op het eerder verrichte onderzoek. Zo is het merendeel niet opgezet als een ATI-studie, maar werd pas achteraf, na toetsing van de hoofdeffecten, besloten ook naar interacties te kijken (Bracht, 1970). Veel onderzoek had betrekking op geprogrammeerde instructie, waarbij studietijd in plaats van studieprestatie als afhankelijke variabele fungeerde (zie Cronbach en Snow, 1977). Onderwijsvariaties hadden in veel onderzoekingen ofwel betrekking op te grote verschillen, ofwel op miniscuul kleine variaties (zie Cronbach en Snow, 1977, p. 509). Als leerlingkenmerken koos men vaak die kenmerken van leerlingen die in de psychologie in andere contexten (selectie en dergelijke) relevant zijn (intelligentie, angst). De test-instrumenten hiervoor waren echter zodanig geconstrueerd, dat ze zoveel mogelijk het gedrag van individuen in zoveel mogelijk situaties kunnen voorspellen. In ATI-onderzoek heeft men echter juist behoefte aan tests die voorspellen of iemand in een bepaalde soort van situatie goed functioneert (zie onder andere Glaser, 1972). Veel ATI-onderzoek tenslotte, was laboratoriumonderzoek in plaats van veldonderzoek (Rosenkrants-Shapiro, 1975).

In de derde plaats kan worden gesteld, dat het toevoegen van een aantal zorgvuldig gekozen tests aan een experimenteel onderzoek vaak een kleine moeite is, terwijl de baten in termen van vergroot inzicht in leerprocessen, en reductie van de foutenvariantie zeer groot kunnen zijn (overigens is in dergelijke gevallen ook het niet vinden van interacties belangrijke informatie).

In de vierde plaats kunnen tegenstrijdigheden in ATI-resultaten zijn veroorzaakt doordat men onvoldoende oog had voor leertaken. Rhetts (1974) drukte het zo uit: "Analysis of published studies (e.g. Bracht, 1970) reveals that aptitude-

treatment-interaction-researchers have generally conceptualized the aptitude-treatment-interaction-problem as two dimensional: learner variables and a diffuse combination of task and treatment variables. However, both treatment and task characteristics can be expected potentially to enter (singly and jointly) into interaction with learner characteristics, thus strongly suggesting that a three-dimensional reference scheme should be used." (p. 340). Tegenstrijdige ATI-resultaten (zoals bijvoorbeeld die met betrekking tot geprogrammeerde instructie (PI), kunnen soms begrijpelijk worden wanneer men de aard van de leertaken erbij betreft. Het is bijvoorbeeld plausibel te veronderstellen, dat PI bij moeilijke leertaken gunstig is voor de betere studenten, en bij gemakkelijke leertaken juist voor de slechtere studenten.

In de vijfde, maar niet in de laatste plaats, kunnen de teleurstellende resultaten van veel ATI-studies worden geweten aan een gebrek aan theorievorming. Veel ATI-studies werden gekenmerkt door een a-theoretisch karakter (zie Flammer, 1975; Rosenkranz-Shapiro, 1975), waarbij de combinatie van leerlingkenmerken, methoden en taken werd bepaald door allerlei toevallige, praktische zaken. Sternberg en Weil (1980, p. 227) geven het volgende voorbeeld: "For example, studies of interaction with spatial ability often do not conceptualize adequately what is meant by spatial ability and do not use measures of spatial ability that assess the kind of spatial ability for which an interaction is likely to obtain."

Veel auteurs hebben Cronbach en Snow verweten, dat ze te weinig oog hebben voor de ATI-theorievorming en te veel voor de ATI-methodologie (zie bijvoorbeeld Rothkopf, 1978; Bond en Glaser, 1979; Sternberg en Weil, 1980; Hunt en Sullivan, 1974; Hunt, 1979; Crombag, 1979). De theoretische problemen voorafgaande aan ATI-onderzoek zijn talrijk (zie voor overzichten Hunt en Sullivan, 1974; Rosenkrantz-Shapiro, 1975; Cronbach en Snow, 1977). Veel van die problemen hangen samen met het volgende algemene probleem: Hoe kunnen uit de veelheid van leertaken en taakdimensies, de veelheid van leerlingkenmerken en de veelheid van omgevingskenmerken de zinvolle combinaties

worden geselecteerd. Het is ons inziens noodzakelijk dat deze selectie of hypothesevorming niet afhankelijk wordt gemaakt van allerlei toevallige factoren, maar geschiedt vanuit theoretisch goed doordachte uitgangspunten (Lodewijks, 1979; Nijhof, 1980). Enkele aanzetten tot ATI-theorievorming komen in 2.3 en 2.4 aan bod.

2.3 Startpunten voor een ATI-theorie

De bekendste aanzet voor een ATI-theorie vormt de bijdrage van Salomon (1972). In die bijdrage heeft Salomon eerdere ideeën van Cronbach en Snow over heuristische strategieën nader uitgewerkt en geformaliseerd. Salomon's uitgangspunt was dat de wijzen waarop onderwijsmethoden en leerlingkenmerken op elkaar worden c.q. kunnen worden afgestemd (matching-concepties) antwoorden kunnen suggereren op de volgende vragen:

1. What do treatments do to learners that can lead to ATI's?
2. What kinds of aptitude measures will interact with what kinds of treatments?
3. What kinds of predictions follow from each of the suggested heuristic models?"

Salomon onderscheidde drie matchingmodellen: remediatie, compensatie en preferentie.

Bij matching volgens een remediatiemodel worden veranderbare verschillen tussen leerlingen opgeheven alvorens met de eigenlijke instructie wordt begonnen. Bij mastery learning worden verschillen in beheersing van voorafgaande kennis opgeheven voordat een volgende eenheid mag worden bestudeerd. Ook een vorm van matching waarbij leerlingen met bepaalde specifieke tekorten worden toegewezen aan een methode die ervoor zorgt dat deze tekorten verdwijnen, terwijl de overige leerlingen aan een methode worden toegewezen die dit effect niet heeft, wordt door Salomon tot het remediatiemodel gerekend.

Bij matching volgens een compensatiemodel wordt bij leerlingen met bepaalde tekorten in vaardigheden een instructiewijze toegepast die ervoor zorgt dat deze tekorten een waarloosbare invloed hebben op resulterende prestaties, zonder dat die tekorten daardoor teniet worden gedaan. Leerlingen die

slecht zijn in het structureren van leerstof krijgen bijvoorbeeld een instructiewijze waarin ze worden geholpen bij het structureren door middel van begrippenschema's, vragen, aanwijzingen en dergelijke. De overige leerlingen krijgen deze extra hulp niet.

Bij matching volgens een preferentiemodel krijgen leerlingen die goed zijn in bepaalde vaardigheden of een bepaalde wijze van informatieverwerking prefereren een instructiewijze die aansluit bij deze vaardigheden of stijlen. Ook wanneer leerlingen op grond van een voorkeur voor een bepaalde instructiewijze aan deze instructiewijze worden toegewezen, spreekt Salomon van preferentiematching. Ons inziens is de term preferentiematching bij Salomon te ruim en is het beter deze twee vormen van matching van elkaar te onderscheiden. In navolging van Cronbach en Snow (1977) zullen we de eerste vorm capitalisatiematching noemen. De tweede vorm blijven we preferentiematching noemen. Een voorbeeld van capitalisatiematching is het toewijzen van leerlingen met een globale stijl van informatieverwerking aan een globaliserende methode en het toewijzen van leerlingen met een analyserende stijl van informatieverwerking aan een analyserende methode. Een voorbeeld van preferentiematching is het toewijzen van leerlingen met een voorkeur voor discussiegroepen aan een discussiegericht onderwijssysteem en leerlingen met een voorkeur voor lezingen aan een frontaal onderwijssysteem.

Deze matching-concepties vervullen - aldus Salomon - een aantal heuristische functies:

- a. Ze geven een indruk van de wijze waarop omgevingskenmerken kunnen inwerken op leerlingkenmerken: onderwijs kan tekorten van leerlingen opheffen (remediatiemodel), omzeilen (compensatiemodel), of juist een beroep doen op de sterke kanten van leerlingen (capitalisatiemodel), danwel aansluiten bij hun voorkeuren (preferentiemodel).
- b. Ze laten zien welke leerlingkenmerken met welke soort van methodekenmerken zullen interacteren en in welke richting: De hoeveelheden extra studie- of instructietijd zullen interacteren met taak-specifieke vermogens (remediatiemodel);

kenmerken van methoden die voor de leerling doen wat hij zelf niet kan (organisatie van het materiaal, stimulusmodaliteit), zullen interacteren met algemene vaardigheden, cognitieve stijlen en algemene "traits or states" (compensatiemodel); onderwijskenmerken die een beroep doen op bepaalde vaardigheden (verschillen in inhoud, structuur, presentatiemodaliteit) zullen interacteren met algemene vaardigheden, cognitieve stijlen of motivatie (capitalisatiemodel).

c. Ze leiden tot een drietal voorspellingen:

- "1. Proficient learners experience interference or boredom with excessive remediation; low proficiency learners benefit since they attain necessary subordinate objectives.
2. High aptituders experience interference when given treatments which provide them with mediators, they can provide on their own. Low aptituders benefit when mediators they are lacking are provided overtly.
3. Each learner learns best when an aptitude in which he is proficient is called upon." (Salomon, 1972, p. 340).

Meer recent heeft Salomon (1979)), parallel aan de bovenbeschreven modellen, een onderscheid gemaakt tussen drie wijzen waarop onderwijs kan inspelen op leerlingkenmerken. De drie nieuwe termen, die hij daarbij gebruikt, (supplantatie, activatie en cultivatie) zijn conceptueel helderder dan de termen compensatie, preferentie en remediatie. Van supplantatie van vaardigheden en processen is sprake als een onderwijsmethode een externe simulatie van interne processen en operaties bevat. "...Supplantation is the function accomplished by an explicit presentation of what would otherwise have to be done covertly by the learner himself." (Salomon, geciteerd in Heidt, 1977). Van activatie van vaardigheden en processen is sprake als een methode een speciaal beroep doet op bepaalde vaardigheden of processen (Salomon, 1979). Van cultivatie van vaardigheden en processen is sprake als een methode ertoe bijdraagt dat vaardigheden verbeteren of processen beter gaan worden uitgevoerd. Salomon (1979) toonde aan dat cultivatie zowel via supplantatie (namelijk door internalisatie van de door de methode

overgenomen operaties) als door activatie kan plaats vinden.

Een tweetal andere bijdragen aan de ATI-theorie werd geleverd door Rhetts (1974). De eerste werd reeds vermeld in 2.2: Bij het denken over ATI moeten de kenmerken van de taak meer in ogenschouw worden genomen dan in voorgaand onderzoek het geval was. De tweede bijdrage van Rhetts betreft de aanbeveling bij het denken over ATI altijd leertaken centraal te stellen (task-first-sequence). Volgens Rhetts (1974) is de optimale volgorde: "First concentrate on an analysis of the characteristics of the task (and its demands on memory, knowledge, motivation, abilities, etcetera); second, identify plausible individual difference characteristics related to the performance demands of the task and third, develop different treatments or modes of presentation (different instructions, cues, amount or distribution of practice, available information or materials etcetera.)" (p. 340). Twee redenen worden voor zo'n denk- en constructievolgorde aangevoerd:

- a. "One reason for adopting a "task-first" sequence is the fact that one of the outstanding features of the human organism is its adaptive capability. This means that the individual can and does modify his behaviour contingent upon the situation at hand, by tailoring his own responses to the nature of the task environment confronting him." (p. 340).
- b. "A second reason for attacking the aptitude-treatment-interaction in a task-first sequence is that documented aptitude-treatment-interactions could significantly influence and inform the process of instructional design. Such problems begin with the question "what is to be taught?" or "what task is the learner to master?" (p. 341).

Diverse auteurs hebben gewezen op het belang van het analyseren van individuele verschillen voor de ATI-theorievorming (bijvoorbeeld Glaser, 1972; Di Vesta, 1975; Snow, 1977). Zij en anderen stelden dat vooruitgang in het denken over ATI vooral afhankelijk is van kennis omtrent de informatieverwerkingsprocessen die zijn gecorreleerd met c.q. ten grondslag liggen aan individuele verschillen. Snow (1977) schreef daarover bijvoorbeeld: "Laboratory analysis of generally important

aptitudes is one way to circumvent the pitfalls of intermediate instructional experiments. Both aptitudes and real school instructional treatments deserve such analysis. But aptitude task analysis is the more malleable task. Process theories of a few major aptitudes brought to educational studies by appropriately revised measures, and examined with improved methodology, could bring a level of understanding of some ATI to the point where practical as well as theoretical cornerstones might be laid." (p. 99). Bij het denken over ATI is het volgens Snow en anderen van groot belang dat men de leerlingkenmerken die men in het onderzoek wil meten kan vertalen in achterliggende processen en mechanismen (deze kennis kan worden verkregen via laboratoriumstudies waarin de leerlingkenmerken worden "vertaald" in cognitieve processen).

Een andere methode die informatie kan opleveren over processen, is een recent door Sternberg (1977) ontwikkelde vorm van taakanalyse: componentanalyse. Componentanalyse "seeks to identify the component mental operations underlying a series of related information-processing tasks and to discover the organization of these component operations in terms of their relationships both to each other and to higher-order constellations of mental abilities." (Sternberg, 1977, p. 65). De uitkomst van een componentanalyse is niet alleen kennis over een serie informatieverwerkingsprocessen die bij de geanalyseerde taken achtereenvolgens spelen, maar ook inzicht in de vraag welke leerlingkenmerken in een bepaalde fase van de taakuitvoering van belang zullen zijn. Componentanalyse toegepast op leertaken kan een belangrijke bijdrage zijn bij het opzetten van ATI-studies waarin leertaken centraal worden gesteld. Een zeer interessante toepassing van componentanalyse in een ATI-onderzoek werd onlangs beschreven door Sternberg en Weil (1980).

Een andere bijdrage aan de ATI-theorievorming betreft het onderscheid tussen een structurele en een functionele beschrijving. Het onderscheid werd in het kader van theorievorming met betrekking tot media (bijvoorbeeld film, televisie, enzovoorts) gemaakt door Heidt (1977), naar aanleiding van eerdere publicaties van Salomon, en Salomon en Snow. Een structurele beschrij-

ving is een beschrijving waarin de uiterlijke verschijningsvorm, de oppervlakte-eigenschappen, de direct in het oog springende kenmerken van het te beschrijven object worden beschreven.

Heidt geeft de volgende voorbeelden van structurele mediakenmerken: "While some media can show steady transformations without a break, others can only present finite situations or steps of a transformation. Some media can present only information of an oral kind, while others have potential for visual and iconic information." (Heidt, 1977, p. 14). Een functionele beschrijving is een beschrijving waarin alleen die kenmerken centraal staan die invloed hebben op of worden beïnvloed door bepaalde kenmerken van lerenden, of in de woorden van Salomon en Snow: "Only those attributes having a unique psychological effect on the viewer." (in Heidt, 1977, p. 14). Kortom een functionele beschrijving van een medium specificceert wat het medium en leerlingkenmerken bij elkaar induceren, de processen die ze bij elkaar oproepen of de effecten waar ze in combinatie met elkaar toe leiden.

Bij het denken over ATI moeten - aldus Heidt - niet de structurele kenmerken van media centraal staan maar de functionele, want juist die bepalen hoe media inwerken op leerlingen/studenten. Heidt stelt voor in ATI-projecten de volgende stappen te nemen:

- "a. Learner characteristics must be classified in a system of internal operations or operational abilities.
- b. Learning tasks must be specified with reference to the internal operations of the learner which they require.
- c. Media must be described with respect to the supplantation they can provide for these operations in various ways because of their different specific structural attributes." (Heidt, 1977, p. 14).

Ook Hunt en Sullivan (1974) hebben een aantal suggesties gedaan om het conceptualiseren van ATI's te vergemakkelijken. Eén van die suggesties betreft de zogenaamde "accessibility channels". Zij verstonden hieronder die leerlingkenmerken die kunnen worden vertaald naar bepaalde onderwijsomgevingen die naar verwachting effectief zullen blijken. "Accessibility

channels" zijn die kenmerken van leerlingen (op een bepaald moment) die aangeven in hoeverre zij beïnvloedbaar zijn door bepaalde omgevingsfactoren. Het voornaamste voordeel ervan is dat de ATI-hypothesen dan als het ware voor het oprapen zouden liggen: "Such information can be immediately related to different ways of teaching" (Hunt en Sullivan, 1974, p. 48). Een voorbeeld van een "accessibility channel" is de motivationele oriëntatie van een leerling (hoge affiliatiebehoefte versus hoge prestatiemotivatie). Deze motivationele oriëntatie voorspelt welke soort van omgeving tot de beste resultaten zal leiden. Affectief-georiënteerde feedback zal tot hogere prestaties leiden bij leerlingen met een hoge affiliatiebehoefte; op de taak gerichte feedback daarentegen zal prestatie-bevorderend werken bij leerlingen met een hoge prestatiemotivatie. Men zou de accessibility channels in de termen van Heidt (1977) functionele beschrijvingen van leerlingen kunnen noemen. Zij geven aan welke inwerkingen leerlingkenmerken hebben op instructiekenmerken.

De laatste suggestie in de richting van een ATI-theorie is afkomstig van Cronbach (1967). Deze heeft betrekking op het ontwerpen van onderwijsmaatregelen: Cronbach stelt voor om het ATI-denken niet te beperken tot bestaande - in het onderwijs en de psychologie gebruikelijke - procedures, maar ook nieuwe elementen aan te dragen: "Until the present time, the differential psychologist has let the institution tell him the treatments for which he is predicting success, and he has designed tests or batteries to make that prediction. I suggest that we now let the institution specify only the criterion - not the treatment - and that the psychologist select an aptitude variable and design treatments expected to interact with it." (Cronbach, 1967, pp.32-33). Hunt (1979) deed een soortgelijke suggestie: "Such a presentation emphasizes environment adaptation to the student rather than student placement into an environment." (p. 187).

2.4 Acht beginselen voor ATI-onderzoek

De in de literatuur aangetroffen suggesties ter verbetering van de ATI-theorie worden in deze paragraaf samengevat en geïntegreerd in een aantal beginselen voor ATI-onderzoek, ten einde een gerichte afleidingsstrategie voor ATI-hypothesen te vinden. Een aantal principes vormt een uitwerking of een uitbreiding van de in 2.3 beschreven suggesties. De eerste vier beginselen werden eerder beschreven in Lodewijks en Simons (1979). De overige vier kwamen in laatst genoemde publicatie niet afzonderlijk en niet expliciet aan bod.

Principe 1: Theoretische encadrering

Dit eerste principe stelt dat in elk ATI-onderzoek van tevoren moet worden aangegeven welke interacties kunnen worden verwacht, in welke richting deze interacties zullen gaan en hoe de eventueel te vinden interacties kunnen worden verklaard. Alleen interacties die men kan verklaren zijn interessant. Het is beter a priori te bedenken welke interacties kunnen optreden en hoe elk van deze zou kunnen worden verklaard. Het beginsel van theoretische encadrering vloeit rechtstreeks voort uit het pleidooi voor een meer theoriegerichte benadering van ATI waarmee 2.2 werd beëindigd. Waar het om gaat, is niet alleen dat kan worden vastgesteld dat er een interactie optreedt maar vooral ook dat wordt beschreven volgens welke redenering een bepaald leerlingkenmerk x in interactie met een bepaalde omgevingsfactor y een bepaalde onderwijsuitkomst z kan bepalen. Het principe van theoretische encadrering krijgt met name gestalte via hetgeen onder principe 5 (interactiemodellen) is beschreven.

Principe 2: Restrictiviteit

Dit beginsel houdt in dat men zich bij het zoeken naar ATI-effecten, en meer in het bijzonder bij de selectie van variabelen, beperkingen oplegt. Deze beperkingen dienen zowel de onderwijs- als de leerlingvariabelen te betreffen. Restricties zijn noodzakelijk ter voorkoming van (interpretatie)moeilijkheden die zich bij hogere-orde-interacties kunnen voordoen en

ter verhoging van de kans op het vinden van werkelijk relevante A * T-interacties. Het gaat bij dit principe dus om twee zaken: Relevantie en Interpretatie. Over het interpretatieprobleem kunnen wij kort zijn: Wat niet interpreteerbaar blijkt, wordt daarmee ook irrelevant. De kans op interpreteerbaarheid wordt groter indien het beginsel van theoretische encadrering in praktijk wordt gebracht. Moeilijker ligt het met het relevantieprobleem. Een ATI kan bijvoorbeeld theoretisch relevant, maar naar praktische en didactische hanteerbaarheid gemeten irrelevant blijken. Het omgekeerde kan zich uiteraard ook voordoen. Hoe het probleem in iedere concrete situatie kan worden opgelost, valt daarom niet aan te geven. Wel kan worden gesteld, dat een leerlingvariabele relevanter wordt naarmate deze voor de realisering van bepaalde onderwijsuitkomsten noodzakelijker is en in de te onderzoeken groep veel verschillende waarden heeft.

Het verdient, anders gesteld, aanbeveling om juist die leerlingvariabelen in ATI-studies op te nemen waarvan bekend is, of kan worden verondersteld, dat zij voor de realisering van bepaalde onderwijsuitkomsten tot op een zeker niveau van de leerling worden vereist. De relevantie van omgevingsvariatiaties kan volgens dezelfde lijn worden beredeneerd. (Natuurlijke) variaties in omgevingskenmerken of experimentele manipulaties ervan zijn relevant voor zover deze variaties (en manipulaties), inhoudelijk en psychologisch inzichtelijk, verband houden met verschillen tussen leerlingen en met leertaakvereisten. Aangezien hierboven de relevantie van verschillen tussen leerlingen afhankelijk is gesteld van leertaakvereisten is de relevantie van omgevingskarakteristieken derhalve van beide variabelen afhankelijk. Het principe van restrictiviteit sluit het meest aan bij de ideeën van Heidt (1977) en Rhetts (1974), zoals die in 2.3 zijn besproken.

Principe 3: Constructiviteit

Volgens Cronbach (1967) dient ATI-onderzoek zich niet alleen te richten op toevallig voorhanden zijnde omgevingsvariatiaties, maar ook op het ontwerpen van nieuwe, bij onderscheiden leerlingkenmerken passende, omgevingscondities.

Analoog aan de suggestie van Cronbach kan het constructiviteitsbeginsel ook worden toegepast op leerlingkenmerken. Het is van belang dat ATI-onderzoek zich niet alleen richt op in de psychologie bekende verschillen tussen leerlingen, waarvoor meetinstrumenten voorhanden zijn, maar dat ook - waar mogelijk - nieuwe meetinstrumenten worden ontwikkeld om nieuwe relevante geachte verschillen te kunnen meten (vergelijk de "new aptitudes" van Glaser, 1972). Men kan daarbij denken aan verschillen tussen leerlingen die samenhangen met leertaakvereisten, leertaakkenmerken of leertaakprocessen, of aan verschillen die samenhangen met gevoeligheden voor bepaalde onderwijsmethoden (vergelijk de "accessibility channels" van Hunt en Sullivan (1974) in 2.3).

Principe 4: Functionaliteit

Het functionaliteitsbeginsel houdt in, dat specificaties van leerlingkenmerken, omgevingsfactoren en uitkomstvariabelen

- a. in dezelfde termen en
- b. in functionele termen moeten worden beschreven.

ad a

Het is nodig om leerlingkenmerken, omgevingsfactoren en uitkomstvariabelen in dezelfde termen te beschrijven, omdat alleen dan "inhoudelijke, psychologisch inzichtelijke" combinaties van de drie soorten variabelen mogelijk zijn (zie Lodewijks en Simons, 1977). Stel dat een onderzoeker een ATI-effect constateert tussen het leerlingkenmerk veldafhankelijkheid en de variabele leerstofsequentie. Hij vindt bijvoorbeeld, dat veldafhankelijke proefpersonen de beste studiestatistieken leveren onder de werking van een gestructureerde methode, terwijl de veldonafhankelijke proefpersonen het meest zijn gebaat bij een ongestructureerde benadering. Indien de onderzoeker het zou laten bij deze constatering voldoet hij niet aan het functionaliteitsbeginsel. Er ontbreekt een specificatie van het interactieve verband in dezelfde termen. In het voorbeeld zou dat onder andere kunnen door de veldafhankelijkheidsdimensie te beschrijven als een structureringsvermogen: Veldonafhan-

kelijke proefpersonen zijn beter in staat stimulusmateriaal te voorzien van een structuur, terwijl veldafhankelijke proefpersonen daartoe niet of minder in staat zijn.

Het principe houdt echter ook in, dat de criteriumvariabele in soortgelijke termen dient te worden beschreven. Het vaststellen van een interactief verband tussen veldafhankelijkheid en leerstofsequentie bij gebruik van een criterium als studieprestaties in het algemeen is nog onvoldoende. Het gaat er bijvoorbeeld om dat veldafhankelijke proefpersonen betere prestaties leveren onder een gestructureerde methode, omdat inzicht in de structuur van de lesstof een belangrijke voorwaarde vormt voor het leveren van goede studieprestaties. Inzicht in de structuur van de lesstof vormt als het ware een intermediaire variabele via welke de interactie tot stand komt. Het eerste gedeelte van het voorbeeld (veldafhankelijkheid beschrijven als structureringsvermogen) maakt duidelijk, dat het beschrijven van leerlingkenmerken en onderwijsmethoden in dezelfde termen overeenkomt met het beschrijven ervan in termen van "accessibility channels" zoals Hunt en Sullivan (1974) dat voorstellen. Het tweede gedeelte, de criteriabeschrijving betreffende, is een ons inziens noodzakelijke uitbreiding van het idee van de "accessibility channels". Het betreft een (tussen)stap die meestal in ATI-redeneringen impliciet blijft, maar wel degelijk wordt gedaan.

ad b

Het tweede gedeelte van het functionaliteitsbeginsel betreft de aard van de termen waarin leerlingen, onderwijsmethoden en criteriumvariabelen moeten worden beschreven. Dat dient namelijk niet alleen in de gebruikelijke structurele termen, maar ook in functionele termen te gebeuren. Het onderscheid tussen structurele en functionele beschrijving vormt een uitbreiding van de in ^{2,3}1.3 beschreven ideeën van Heidt (1977). Deze auteur pleitte voor het functioneel beschrijven van media en hij beoogde daarmee dat media zodanig moeten worden beschreven, dat duidelijk wordt welke invloeden mediakenmerken kunnen hebben op welke leerlingkenmerken ("only those attributes having a

unique psychological effect on the viewer"). Naar onze mening dienen niet alleen media, maar alle omgevingsvariabelen, leerlingkenmerken en leertaakkenmerken in functionele termen te worden beschreven. Een functionele beschrijving van een variabele in algemene zin is een zodanige specificatie van kenmerken dat duidelijk wordt hoe een variabele inwerkt op andere variabelen. Bij ATI's gaat het om drie soorten variabelen: leerlingkenmerken, omgevingsfactoren en uitkomstvariabelen. Bij functionele beschrijvingen van deze variabelen specificceert men de variabelen naar datgene wat ze bij elkaar induceren, de processen die ze bij elkaar oproepen, of de effecten waar ze in combinatie met elkaar toe leiden. In principe zijn er negen soorten van functionele beschrijvingen mogelijk. Omdat volgens principe 2 (restrictiviteit) en het nog te beschrijven principe 7 de volgorde van functionele beschrijvingen bij voorkeur van leertaak via leerlingkenmerken naar omgevingsfactoren moet verlopen, zijn slechts vier functionele beschrijvingen belangrijk:

- a. Invloeden van de leertaak op leerlingkenmerken: Op welke processen doet deze leertaak een beroep; wat zijn de karakteristieke moeilijkheden van deze taak (cf. Elshout, 1976). De verschillende manieren waarop leertaken invloed hebben op variaties in leerlingkenmerken is in Lodewijks (1980) uitgebreid beschreven.
- b. Invloeden van leerlingkenmerken op de leertaakkenmerken: Over welke van deze processen beschikken de leerlingen in verschillende mate; welke leerlingen ondervinden deze moeilijkheden?
- c. Invloeden van leerlingkenmerken op omgevingskenmerken: Welke functies moeten onderwijsmethoden bij verschillende typen leerlingen vervullen?
- d. Invloeden van onderwijsvariaties op de combinaties van leertaak en leerlingkenmerken: Welke functies kunnen omgevingskenmerken vervullen bij de moeilijkheden die verschillende typen leerlingen ondervinden bij bepaalde leertaken?

De wenselijkheid om de drie ATI-componenten in functionele termen te beschrijven sluit wat betreft de leertaak aan

bij de ideeën van Rhetts (1974), wat betreft de leerlingkenmerken bij de ideeën van Snow (1977), Di Vesta (1975), Hunt en Sullivan (1974) en anderen die individuele verschillen willen vertalen in procestermen, en wat betreft de omgevingsfactoren bij Heidt (1977) en Salomon (1979) (zie 2.3). Nieuw is dat in onze benadering elk van de ATI-componenten functioneel moet worden beschreven. Nieuw is ook, dat vertaling in procestermen niet noodzakelijk behoeft te wachten op empirische taakanalyses, zoals Snow en anderen bepleitten, maar dat ook rationele analyses van leertaken, leerlingkenmerken en situatiekenmerken een uitgangspunt kunnen vormen. Waar empirische analyses voorhanden zijn, moeten deze natuurlijk worden gebruikt. Wanneer ze ontbreken, moet toch zoveel mogelijk langs rationeel-intuïtieve weg een vertaling in procestermen plaatsvinden ("every psychologist is a person" (Hunt, 1979, p. 188)).

Principe 5: Vier interactiemodellen

Impliciet in het artikel van Salomon (1972) zijn vier mogelijke oorzaken van ATI's beschreven. Volgens principe 5 moeten bij het denken over ATI telkens deze vier mogelijke oorzaken van ATI's worden betrokken. Het zijn ons inziens de volgende vier:

- a. De ene methode activeert bepaalde leerlingkenmerken meer dan de andere;
- b. De ene methode supplanteert bepaalde leerlingkenmerken meer dan de andere;
- c. De ene methode interfereert meer met bepaalde leerlingkenmerken dan de andere;
- d. De ene methode wordt door bepaalde leerlingen meer geprefeerd dan de andere.

De termen activatie en supplantatie komen uit het artikel van Salomon (1979) en de termen interferentie en preferentie uit zijn publicatie van 1972. Hoewel Salomon deze vier termen gebruikte, merkte hij ze niet expliciet aan als mogelijke oorzaken van interacties. Waarschijnlijk omdat men in die tijd te veel was gericht op het vinden van disordinale interacties (zie misvatting b in 2.1), presenteerde Salomon (1972) in plaats daarvan zijn drie mat-

chingmodellen: remediatie, compensatie en capitalisatie op zodanige wijze, dat ze alleen bruikbaar waren om mogelijke disordinaire interacties te ontdekken. Een disordinaire interactie die past in het compensatiemodel is een gevolg van het samengaan van supplantatie en interferentie. En ook een disordinaire interactie die past in het remediatiemodel ontstaat, wanneer er sprake is van een combinatie van supplantatie of activatie met interferentie.

Bovendien is het onderscheid tussen het remediatiemodel en het compensatiemodel voornamelijk van belang in verband met praktische toepassingen, niet met betrekking tot mogelijke oorzaken van interacties. Een remediatiematching (dat wil zeggen het remediëren van de tekorten van leerlingen) past men toe op basis van een gevonden interactie die zowel kan zijn veroorzaakt door activatie als door supplantatie (zie voor evidentie Salomon, 1979). Remediatie (en cultivatie) zijn dan ook van een andere orde dan compensatie en preferentie (c.q. capitalisatie). Natuurlijk is het wel degelijk van belang of men leerlingkenmerken gaat remediëren of cultiveren of niet. Dat is echter een praktische, onderwijskundige beslissing die mede kan worden genomen op basis van ATI-resultaten (zie verder principe 8), die echter geen heuristische functie vervult bij het hypothetiseren van ATI's.

Onze bezwaren tegen de drie matchingmodellen van Salomon (1972) kunnen als volgt worden samengevat:

1. Matchingmodellen zijn te veel gericht op het vinden van alleen disordinaire interacties.
2. Het remediatiemodel is conceptueel niet verschillend van het compensatie- of het preferentie (capitalisatie)-model (alleen in praktische zin verschillend).
3. Interferentie wordt onvoldoende als afzonderlijke oorzaak van ATI-interacties erkend.
4. Preferentie en capitalisatie worden niet van elkaar gescheiden.

Als alternatief voor de modellen van Salomon stellen wij derhalve voor als interactiemodellen: activatie, supplantatie, interferentie en preferentie. Bij het hypothetiseren van ATI's

dient men zich ons inziens af te vragen, welke van deze oorzaken van interactie in welke mate kunnen optreden. Waar slechts één van de vier modellen een rol kan spelen, kan men doorgaans alleen ordinale interacties verwachten. Disordinale interacties kan men verwachten als:

- a. één van de modellen (activatie, supplantatie, preferentie) samengaat met het interferentiemodel;
- b. als twee vormen van activatie samengaan (de ene methode activeert bijvoorbeeld cognitieve stijl A en de andere cognitieve stijl B);
- c. als er twee vormen van preferentie samengaan (de ene methode wordt geprefereerd door de ene groep van leerlingen en de andere door een andere groep).

Principe 6: Eerst de leertaak

Dit principe is reeds beschreven in 2.3 bij de bijdragen van Rhetts (1974). Het houdt in dat leertaken bij het denken over ATI centraal moeten staan. Naast de in 2.3 vermelde argumenten van Rhetts, kunnen daarvoor ook nog de volgende argumenten worden genoemd.

Alleen wanneer de leertaak als eerste wordt geanalyseerd, kan worden voldaan aan het constructiviteitsbeginsel met betrekking tot de onderwijsmethoden (principe 3). Ook uit het restrictiviteitsbeginsel vloeit rechtstreeks voort, dat de leertaak centraal wordt gesteld (zie onder principe 2). Uit de aard van de leertaak (en/of het criterium) wordt afgeleid welke minimale procesmatige eisen aan de leerlingen dienen te worden gesteld of welke de karakteristieke moeilijkheden van de taak zijn.

Dat leerlingkenmerk x instructiekenmerk - interacties leertaakspecifiek zijn kan gemakkelijk worden geïllustreerd met allerlei voorbeelden. Zo zullen maatregelen om het geheugen van de leerling te ontlasten zinvoller zijn voor leerlingen met een "slecht" geheugen dan voor leerlingen met een "goed" geheugen - echter alleen dan wanneer de leertaak in kwestie in belangrijke mate een beroep doet op het geheugen. En evenzo zullen onderwijsmaatregelen die leerlingen ertoe moeten brengen met meer inzicht te studeren (bijvoorbeeld door in

de tekst inzichtvragen op te nemen) zinvoller blijken voor leerlingen die gewend zijn alles uit het hoofd te leren dan voor leerlingen die uit zichzelf al met inzicht studeren. Ook hier geldt dat het voor de leertaak in kwestie van belang moet zijn de leerstof met inzicht te verwerken. Vaak denkt men nog aan algemene A-T-interacties als: Visueel ingestelde mensen profiteren meer van plaatjes bij een tekst dan verbaal ingestelde mensen. Het zal inmiddels duidelijk zijn dat dergelijke algemene ATI-hypothesen ons inziens meestal niet erg zinvol zijn. Het hangt er maar helemaal vanaf wat voor soort lesstof in de tekst is opgenomen: Kan men zich gemakkelijk zelf een mentale voorstelling maken van de lesstof of is dat moeilijk?

Principe 7: Optimale volgorde

Met Rhetts (1974) zijn wij van mening, dat er een optimale volgorde is in het analyseren van de drie ATI-componenten (zie 2.3). Volgens principe 6 moet eerst de leertaak worden geanalyseerd. Als tweede stap komen de leerlingkenmerken in aanmerking en dan pas de omgevingskenmerken, omdat de relevantie van de laatste afhankelijk is van de leerlingkenmerken (zie onder principe 2: restrictiviteit). Vanuit een constructief theoretisch standpunt moeten we methoden zoeken bij verschillen tussen leerlingen en niet leerlingen toewijzen aan bestaande methoden (Hunt, 1979). Dus ook het constructiviteitsbeginsel leidt tot de volgorde leerlingkenmerken-omgevingskenmerken.

Een tweede volgordeprobleem betreft het onderscheid tussen "structurele en functionele beschrijving" waarop principe 4 (functionaliteit) berust. Moeten eerst de structurele of eerst de functionele beschrijvingen worden gemaakt? Uit het feit dat de lijst van functionele kenmerken een deelverzameling vormt van de lijst van structurele kenmerken (Heidt, 1977) zou men kunnen afleiden dat de structurele beschrijvingen moeten voorafgaan aan de functionele. In eerdere publicaties (Lodewijks en Simons, 1979) hebben we dan ook de structurele beschrijvingen consequent vooraf laten gaan aan de functionele. In de praktijk blijkt echter een zo stringente prescriptie tot problemen te kunnen leiden. Uit het constructivi-

teitsbeginsel volgt bijvoorbeeld, dat een functionele beschrijving van omgevingsfactoren die op basis van de leertaak- en leerlinganalyse is verkregen, moet worden vertaald in een nieuwe structurele beschrijving. In zo'n geval ligt het meer voor de hand om onmiddellijk aan te sluiten bij hetgeen uit voorgenoemde analyse reeds is verkregen en pas in latere fase "oppervlakte"kenmerken (structurele) in te vullen.

Principe 8: Cultivatie of bestendiging van leerlingkenmerken

In 2.3 werd cultivatie van vaardigheden of processen gedefinieerd als het toenemen van vaardigheden of het beter gaan verlopen van processen. Onder principe 5 (interactiemodellen) werd betoogd, dat remediatie en cultivatie van een andere orde zijn als compensatie (en supplantatie) en ook van een andere orde als preferentie-capitalisatie (en activatie). De vraag of een methode leerlingkenmerken cultiveert of bestendigt, mag dan wel niet zozeer van belang zijn bij het afleiden van hypothesen over ATI's (zie onder principe 5), maar is wel degelijk van praktisch belang. Zie bijvoorbeeld de volgende citaten: "Sometimes the treatment "works back" upon the aptitude, so that after a time the person's aptitude has changed. If so, it might then be advisable to modify his treatment." (Cronbach en Snow, 1977, p. 23).

"Consider the likelihood, however, that in our efforts to optimize the learning of subject matter we may solidify the global child's cognitive style that he may never learn to discover anything in his entire school career. This possibility suggests that teaching to produce maximal learning of subject matter is not enough. We should also be concerned with the student's manner of thinking. One possibility here is that we should attempt to foster alternative modes of cognition and multiple stylistic approaches to problem solving." (Messick, geciteerd in Hunt, 1979).

Daarom is het van belang in ATI-onderzoek informatie te verzamelen over de vraag welke methode het meest de leerlingkenmerken cultiveert, zodat deze informatie bij eventuele latere beslissingen kan worden meegewogen. Bijvoorbeeld door,

zoals Heidt (1977) dat voorstelt, de mate van supplantatie die een medium een leerling biedt, successief en geleidelijk te reduceren. Zie voor uitstekend werk op dit gebied Salomon (1979).

2.5 De Correspondentie-analysestrategie

De correspondentie-analysestrategie is een stap-voor-stap-procedure, die de onderzoeker kan helpen bij het zoeken naar correspondenties tussen een leertaak, een leerlingkenmerk en omgevingsfactoren, die de basis kunnen vormen voor een ATI-hypothese. Met correspondenties worden psychologisch-inzichtelijke, zinvolle en relevante combinaties van de drie ATI-componenten bedoeld. De strategie omvat tien stappen die bij voorkeur in de aangegeven volgorde moeten worden uitgevoerd (zie echter de opmerkingen in het vorige hoofdstuk onder principe 7). De strategie vormt een concrete toepassing van de in 2.4 beschreven principes (met uitzondering van principe 8).

De tien stappen zijn achtereenvolgens:

1. Leertaakbeschrijving (structurele beschrijving van de leertaak)

Een leertaakbeschrijving omvat het in kaart brengen van de uitvoeringsstructuur van een leertaak. Met een leertaakbeschrijving poogt men, met andere woorden, het organisatiepatroon van alle handelingen en processen die tot een bepaalde leertaak behoren af te beelden. Daartoe behoort dus het opsporen van alle deelhandelingen en processen van een leertaak (de zogenaamde leertaakcomponenten) en het aangeven in welke functies de verschillende leertaakcomponenten onderling samenhangen. De centrale vraag bij leertaakbeschrijving is steeds: Wat houdt de betreffende leertaak structureel in? Leertaakbeschrijving kan vanuit verschillende perspectieven worden begonnen.

Men kan de uitgangssituatie beschrijven op basis waarvan een serie handelingen en processen door de leerling moet worden vertoond; men beschrijft dan in structurele zin de objecten waaraan moet worden gehandeld (bijvoorbeeld het analyseren van leerstof).

De uitvoeringsstructuur zelf kan worden beschreven: Welke handelingen en processen worden in welke volgorde en afhankelijk van welke omstandigheden in het kader van een leertaak uitgevoerd?

Tenslotte kan ook een bedoelde eindsituatie worden beschreven; dit houdt in dat men een nadere omschrijving geeft van hetgeen als "terminal behavior", criteriumgedrag of als eindhandelingsstructuur in een doelstelling is aangegeven.

2. Leertaakanalyse (functionele beschrijving van de leertaak)
Met een leertaakanalyse probeert men inzicht te verschaffen in de vraag wat aan kennis, inzicht en psychologische operaties van de leerling wordt gevraagd bij het juist uitvoeren van een bepaalde leertaak. Meer in het bijzonder wordt in een leertaakanalyse de vraag gesteld welke leerprocessen, psychologische leerlingkenmerken en operaties naar veronderstelling ten grondslag liggen aan het correct uitvoeren van alle leertaakcomponenten die in een leertaakbeschrijving zijn gevonden. Bij een leertaakbeschrijving gaat het dus om het in kaart brengen van de leertaakcomponenten, terwijl het bij een leertaakanalyse gaat om het vinden van de leertaakvereisten, de leertaak-specifieke leerlingkenmerken, of de karakteristieke processen, handelingen of moeilijkheden. Het kan gaan om moeilijkheden bij de uitvoering van het geleerde; het kan ook gaan om moeilijkheden die zich tijdens verschillende fasen van het leerproces zelf kunnen manifesteren (zie de "stages of competence" van Resnick, 1976).

Naast de standaardprocedures voor leertaakanalyse (zoals die van Gagné (1977), Landa (1976) of Reigeluth et al. (1978)) komen vooral ook meer intuïtieve, hypothetische methoden van leertaakanalyse in aanmerking. Bij dergelijke intuïtieve methoden kan men zich bijvoorbeeld de volgende vragen stellen:

- Welke zijn de karakteristieke moeilijkheden van deze leertaak (vergelijk Elshout, 1976)?

- Welke zijn de belangrijke processen bij de uitvoering van deze taak (vergelijk Sternberg, 1977)?
- Welke zijn de kenmerkende verschillen tussen een beginnende en een expert (vergelijk Glaser, 1976)?
- Welk beroep doet deze taak op het werkgeheugen (vergelijk Case, 1978)?
- Welk type leerlingen zal (welke) moeilijkheden ondervinden bij het leren van deze taak?

Antwoorden op deze vragen kunnen op drie wijzen verkregen worden:

- a. Vanuit reeds verrichte empirische studies als die van Elshout (1976), Sternberg (1977), Resnick en Glaser (1976) en anderen.
- b. Vanuit nog te verrichten empirische studies (waarbij men kan denken aan analyse van hard-op-denken-protocollen, gesprekken met docenten en leerlingen, componentanalyse (à la Sternberg, 1977)) of literatuuronderzoek.
- c. Vanuit speculatieve intuïties ("the psychologist is a person too." (Hunt, 1979)).

3. Leerlingbeschrijving (structurele beschrijving van de leerlingen)

In een leerlingbeschrijving wordt weergegeven voor welke soort van leerlingen de leertaak is bedoeld, in welke opzichten zij van elkaar verschillen en in welke juist niet. Voornaamste probleem bij de leerlingbeschrijving vormt de selectie: Welke kenmerken van leerlingen zijn wel relevant en welke niet? De volgende heuristieken kunnen behulpzaam zijn bij het opstellen van de leerlingbeschrijving:

- a. Beschrijf de leerlingen in termen van psychologische leerlingkenmerken (zie De Klerk, 1979a).
- b. Beschrijf de leerlingen in termen van die leerlingkenmerken die op enigerlei wijze in verband kunnen worden gebracht met de resultaten van de leertaakanalyse. Bijvoorbeeld door die kenmerken te selecteren die verband lijken te houden met uit de leertaakanalyse gebleken prerequisite vaardigheden of processen.

- c. Beschrijf de leerlingen in termen van die leerlingkenmerken waarvan mag worden verwacht, dat ze een leerlinganalyse mogelijk maken.
- d. Beschrijf de leerlingen in termen van die leerlingkenmerken die verwijzen naar situatiekenmerken. Deze zogenaamde "accessibility characteristics" (Hunt en Sullivan, 1974) specificeren voor welke soort van onderwijsmethoden en/of informatie leerlingen vooral ontvankelijk zijn. Zo maken Hunt en Sullivan bijvoorbeeld onderscheid tussen subjecten met een hoog conceptueel niveau en met een laag conceptueel niveau. Eerstgenoemden zouden vooral open staan voor ongestructureerde onderwijsmethoden en informatie, terwijl de mensen met een laag conceptueel niveau vooral open staan voor gestructureerde informatie. Het zal duidelijk zijn, dat bij dergelijke "accessibility characteristics" de leerlingbeschrijving en de leerlinganalyse in elkaar overlopen.
- e. Beschrijf de leerlingen vooral ook in termen van die leerlingkenmerken waarin ze niet of weinig van elkaar verschillen.
- f. Beschrijf de leerlingen in termen van leerlingkenmerken die verwijzen naar leerprocessen. Een zogenaamde leerstijltest (zoals die van Pask (1976a)) verwijst bijvoorbeeld zodanig rechtstreeks naar leerprocessen dat een leerlinganalyse bijna overbodig wordt.

Soms ligt het voor de hand de leerlingbeschrijving vooraf te laten gaan door de leerlinganalyse (zie principe 7 in 2.4). Dan moet in de leerlingbeschrijving alleen worden aangegeven hoe de functionele leerlingkenmerken kunnen worden vertaald in meetbare verschillen tussen leerlingen.

- 4. Leerlinganalyse (functionele beschrijving van de leerlingen)
Met een leerlinganalyse wordt inzicht beoogd in de vraag op welke wijze(n) beschreven leerlingkenmerken inwerken op de leertaak(kenmerken) en/of op mogelijke situatiekenmerken. In een leerlingbeschrijving gaat het daar-

entegen om een beschrijving op structureel (in tegenstelling tot functioneel) niveau van de leerlingen voor wie het onderwijs is bedoeld. In de functionele leerlinganalyse worden de beschreven leerlingkenmerken op zodanige wijze geanalyseerd dat inzicht wordt verkregen in de relaties leertaak-leerlingkenmerk en/of leerlingkenmerk-situatiekenmerk. Men kan zich bij een leerlinganalyse vier vragen stellen:

a. Welke informatieverwerkingsprocessen liggen ten grondslag aan de beschreven verschillen tussen leerlingen? Antwoord op deze vraag kan men krijgen door resultaten van empirische analyses (cf. Snow, 1977; Hunt, 1977) toe te passen of zelf meer intuïtieve, rationele analyses uit te voeren (zie principe 4 in 2.4).

b. Welke relaties kunnen er bestaan tussen de geanalyseerde leertaakkenmerken en de beschreven leerlingkenmerken: Hoe werken deze leertaakkenmerken en de leerlingkenmerken op elkaar in? - Vraagt men zich bijvoorbeeld af hoe een complexe leertaak en het leerlingkenmerk algemene intelligentie op elkaar inwerken dan vindt men simpel het antwoord: De intelligente leerlingen zullen minder hinder ondervinden van de complexiteit van de leertaak dan de minder intelligente. Een leerlinganalyse kan derhalve eenvoudigweg bestaan uit het afwegen - op grond van empirische bevindingen of van plausibiliteitsredeneringen - welke invloeden de verschillende leerlingkenmerken kunnen hebben op de mate waarin de leertaakkenmerken problemen zullen opleveren.

c. Welke relaties kunnen er bestaan tussen de beschreven leerlingkenmerken en mogelijke situatiekenmerken (bijvoorbeeld onderwijsmethoden of tekstvariabelen)? Ook hier kan men weer trachten op basis van empirische bevindingen, maar ook door te redeneren, te achterhalen welke interacties er zullen kunnen optreden. Het is echter bij deze methode van belang goed in het oog te houden of de veronderstelde interactie tussen leerlingkenmerk en situatiekenmerk wel tot haar recht kan komen binnen de

beperkingen van de leertaak (-beschrijving en -analyse).

d. Zijn er op basis van de leertaakanalyse verschillen tussen leerlingen te bedenken, waarvoor (nog) geen meetinstrumenten bestaan? Stel bijvoorbeeld dat men via een leertaakanalyse op het idee is gekomen, dat een moeilijkheid van het leren van die taak zou kunnen zijn dat men een periode van reflectie op eigen gedrag moet afwisselen met een periode van automatisch handelen zonder te veel na te denken (bijvoorbeeld bij het leren tennissen). Wanneer zou zijn gebleken dat geen der bestaande leerlingkenmerken potentieel van invloed is op dit leertaakkenmerk, dan kan men zich afvragen of het mogelijk is nieuwe verschillen tussen leerlingen te gaan meten die wel inwerken op het betreffende leertaakkenmerk. In het onderhavige geval zou men bijvoorbeeld kunnen denken aan de potenties van leerlingen om af te zien van reflectie op eigen handelen.

5. Situatiebeschrijving (structurele beschrijving van situaties)

In een situatiebeschrijving wordt geïnterpreteerd wat er voor onderwijsmaatregelen en -variëaties mogelijk zijn om de leertaak te leren. Hier speelt - evenals bij de leerlingbeschrijving - het probleem van de selectie: Welke maatregelen en variëaties zijn relevant. Het selectieprobleem kan gedeeltelijk worden opgelost door middel van de volgende heuristieken:

- a. Beschrijf de situaties in termen van directe of eerste orde situatiekenmerken. Dat zijn omgevingskenmerken die naar veronderstelling rechtstreeks - dat is zonder tussenkomst van andere (interveniërende) variabelen - het leerproces beïnvloeden.
- b. Beschrijf de situaties in termen van die kenmerken die op enigerlei wijze in verband kunnen worden gebracht met de resultaten van de leertaakanalyse en van de leerlinganalyse.
- c. Beschrijf de situatie in termen van die kenmerken waarvan mag worden verwacht dat deze de situatie-analyse

mogelijk maken.

- d. Beschrijf de situaties in termen van die kenmerken waarvan door middel van experimenteel didactisch onderzoek enigermate bekend is geworden op welke processen zij een beroep doen. In dat geval gaat de situatiebeschrijving over in een situatie-analyse. Soms kan men de situatie-analyse beter vooraf laten gaan aan de situatiebeschrijving (zie principe 7 in 2.4). Dan moet in de situatiebeschrijving alleen worden aangegeven hoe de functionele situatiekenmerken kunnen worden vertaald in structurele kenmerken.

6. Situatie-analyse (functionele beschrijving van situaties)

In een situatie-analyse worden de uit de situatiebeschrijving resulterende situatiekenmerken (onderwijsmaatregelen en/of onderwijsvariatiën) nader geanalyseerd in termen van onderliggende processen. Het gaat hier om een functionele (en geen structurele) beschrijving van situatiekenmerken. De centrale vraag daarbij is weer op welke wijze de beschreven situatiekenmerken zouden kunnen inwerken op de leerlingkenmerken en de leertaakkenmerken. Vaak komt men dan terecht bij het bepalen van (potentiële) functies van onderwijsmaatregelen en/of -variatiën. Enkele vragen die men zich bij een situatie-analyse kan stellen zijn:

- a. Op welke informatieverwerkingsprocessen doen de beschreven situatieverschillen een beroep?
- b. Welke functies kunnen de beschreven situatiekenmerken vervullen in een onderwijsleerproces?
- c. Op welke wijze kunnen de situatiekenmerken inwerken op de beschreven (en geanalyseerde) leerlingkenmerken?
- d. Op welke wijze kunnen de situatiekenmerken moeilijkheden van de leertaak opheffen of versterken?
- e. Zijn er verschillen tussen onderwijssituaties te bedenken waarvoor nog geen meetinstrumenten bestaan of die nog niet in onderwijsmaatregelen zijn vertaald?

7. Interactie-analyse (eerder correspondentie-analyse genoemd, maar om verwarring met de op dezelfde wijze betitelde totaalstrategie te voorkomen, omgedoopt tot interactie-analyse) Tijdens of na een situatie-analyse komen we voor het probleem te staan, dat voorspellingen moeten worden gedaan over de richting van een beredeneerde interactie: Zal de interactie ordinaal of disordinaal zijn? Zullen hoog scorenden juist meer profiteren van methode A of juist laag scorenden (enzovoorts)?

Soms zal deze voorspelling al tijdens de voorafgaande fasen in de correspondentie-analyse duidelijk worden, soms ook zijn er op basis van een beredeneerde correspondentie nog verschillende voorspellingen mogelijk (zie Simons en Lodewijks, 1979). Dit laatste is met name het geval bij cognitieve stijl-leerlingkenmerken. De vier in principe 5 (2.4) gepresenteerde interactiemodellen (activatie, supplantatie, interferentie en preferentie) en hun combinaties kunnen daarbij een heuristische functie vervullen. Uit het voorafgaande zal duidelijk zijn geworden, dat in het kader van de correspondentie-analysestrategie dergelijke voorspellingen ten eerste in functionele termen en ten tweede ten aanzien van de drie ATI-componenten moeten worden geformuleerd.

8. Realisering van correspondenties

In deze fase gaat het om de praktische realisering van de correspondenties die in de correspondentie-analyse zijn vastgesteld. Anders gezegd: Het gaat hier om het ontwikkelen van (experimentele en onderwijs)situaties waarin de hypothesen die in de vorige fase zijn gegenereerd kunnen worden getoetst.

9. Toetsing van correspondenties

In deze fase wordt het eigenlijke ATI-onderzoek verricht. Dit onderzoek vindt plaats binnen de in de vorige fase geconceptualiseerde onderwijssituaties en is gericht op de beantwoording van twee vraagstellingen:

- a. Blijken de hypothetisch optimale combinaties (correspondenties) gemeten naar in dezelfde termen gespecificeerde criteria effectief?
- b. Welke zijn de effecten bij mismatching?

10. Disseminatie en implementatie

Afhankelijk van de uitkomsten van het ATI-onderzoek dat in fase 9 is uitgevoerd, kunnen verschillende beslissingen worden genomen. Indien voor de veronderstelde hypothetische correspondenties empirische steun is gevonden, kan worden overwogen de inzichten en producten (onderwijsvormen) die het resultaat waren van correspondentie-analyse te verspreiden en in te voeren of te doen invoeren. Dergelijke beslissingen kunnen natuurlijk pas worden genomen wanneer het onderzoeksresultaat tenminste herhaalbaar is gebleken. Bovendien dienen in het beslissingsproces ook andere factoren te worden betrokken (kosten, studietijd, andere effect-criteria, praktische moeilijkheden).

2.6 De correspondentie-analysestrategie in deze dissertatie

Vanuit het perspectief van een ATI-theorie is de in 2.5 gepresenteerde strategie en de daarin gerealiseerde volgorde de meest optimale. Dat wil echter niet zeggen dat andere volgorde altijd ongewenst zijn. Ze zijn alleen niet optimaal. In een project over een bepaald situatiekenmerk (zoals in het onderhavige project over metaforen) kan men niet anders dan beginnen met een situatiebeschrijving en een situatie-analyse (wat zijn metaforen, welke soorten kunnen onderscheiden worden en welke functies kunnen metaforen vervullen in een onderwijsleerproces). Om toch zoveel mogelijk van de principes uit 2.4 te realiseren moet men dan echter snel overgaan naar de leertaak en zich afvragen bij wat voor soort leertaken en leertaakmoeilijkheden metaforen hun functies kunnen vervullen. De aldus beredeneerde correspondenties tussen de leertaak en de omgevingsfactoren vormen de basis voor de leerlingbeschrijving en -analyse. Ook bij deze volgorde worden de acht beginselen voor ATI-onderzoek (zie 2.4) in belangrijke mate gerealiseerd. De situatiebeschrij-

ving en -analyse worden in dit proefschrift gerapporteerd in hoofdstuk 3 waar uit diverse onderwijsleertheorieën een aantal functies van metaforen wordt afgeleid. Deze functies vormen de selectiecriteria voor de leertaken. De leertaken die in de experimenten in deel III centraal stonden, zijn op basis van deze functies geselecteerd. Er werd dus telkens gezocht naar leertaken die zodanige kenmerken c.q. moeilijkheden hadden dat één of meer van de functies van de metaforen daarop van invloed zouden kunnen zijn (zie hiervoor de inleidingen op de experimenten). De leerlingkenmerken werden geselecteerd op grond van de correspondenties tussen de leertaak en de onderwijskenmerken (zie deel III).

Hoofdstuk 3 Theoretische functies van metaforen

Dit hoofdstuk is verdeeld in vier secties. In het eerste deel (3.1) worden metaforen gedefinieerd en wordt een aantal soorten metaforen onderscheiden.

In 3.2 komen diverse psychologische theorieën aan bod. Telkens wordt nagegaan welke voordelen van metaforen uit deze theorieën kunnen worden afgeleid. Deze voordelen worden vertaald in een drietal functies die metaforen in onderwijsleerprocessen kunnen vervullen.

In de volgende paragraaf volgt de keerzijde van de medaille. Uit de literatuur en door extrapolatie wordt een aantal argumenten tegen metaforen afgeleid.

In 3.4, tenslotte, worden conclusies getrokken. Belangrijkste conclusie is, dat er behoefte is aan empirisch onderzoek dat het best volgens een ATI-methode kan worden opgezet.

3.1 Vergelijkend leren: definiëring, categorisering en een procesbeschrijving

In dit proefschrift wordt een theoretische en een empirische analyse gemaakt van "vergelijkend leren". Wat is vergelijkend leren? Welke soorten van vergelijkend leren kunnen worden onderscheiden? Hoe zou vergelijkend leren kunnen verlopen?

Op deze drie vragen worden in paragraaf 3.1 antwoorden geformuleerd.

Onder vergelijkend leren verstaan we het leren (en begrijpen) van begrippen en relaties tussen begrippen door vergelijking met begrippen en relaties die al bekend waren of die gemakkelijker kunnen worden geleerd (of begrepen). Dit kan op verschillende wijzen gebeuren:

1. De lerende kan zelf vergelijkingen trekken tussen nieuwe lesstof en oude begrippenschema's.
2. De lerende kan door de leerkracht worden gewezen op reeds bekende begrippenschema's die hij met de nieuwe stof kan vergelijken.
3. De lerende kan, alvorens tot het leren van de nieuwe lesstof over te gaan, eerst nieuwe begrippenschema's leren die

gemakkelijker te leren (of te begrijpen) zijn dan de lesstof waar het eigenlijk om gaat (bijvoorbeeld doordat deze lesstof concreter of beter voorstelbaar is, beter aansluit bij de voorkennis of leuker is).

Naast de term vergelijkend leren worden ook de termen "leren via metaforen" en "leren via analogieën" gebruikt.

Etymologisch gezien betekent de term "metafoor" overdracht (transfer). De term stamt namelijk af van het Griekse meta en pherein (dragen) (Ortony, 1975). Een metafoor was volgens Aristoteles een middel om twee termen met elkaar te vergelijken (op.cit.). De ene term wordt tegenwoordig "tenor" of "topic" genoemd (namelijk de term waarover iets wordt gezegd), de andere term wordt "vehicle" genoemd (namelijk datgene waarmee wordt vergeleken). Verder worden nog onderscheiden de "ground", datgene wat de twee termen gemeen hebben, en de "tension", het verschil tussen de twee termen (op. cit. p. 45).

In dit proefschrift wordt de term metafoor in een ruime betekenis gehanteerd. Er worden geen onderscheidingen gemaakt tussen metaforen, analogieën, vergelijkingen, gelijkenissen, parabels en dergelijke. Deze verschillende termen worden door elkaar gebruikt en aan de subtiele onderscheidingen ertussen wordt voorbijgegaan (zie voor deze onderscheidingen: Ortony, 1975, 1979a, 1979b). Dus of de vergelijkingen expliciet of impliciet, beperkt of uitgebreid, letterlijk of figuurlijk worden gebruikt, of het daarbij om relaties of objecten gaat, in alle gevallen worden de termen metafoor, analogie en vergelijking (door elkaar) gebruikt. We gebruiken deze termen als er sprake is van een parallelliteit tussen twee of meer verschijnselen of begrippen. "We reason analogically whenever we make a decision about something new in our experience by drawing a parallel to something old. When we buy a new goldfish because we liked our old one, or when we listen to a friend's advice because it was correct once before, we are reasoning analogically" (Sternberg, 1977b, p. 353). Soms wordt de term metafoor in de betekenis van "vehicle" (zie boven) gebruikt. Bij gebrek aan een goede Nederlandse term voor "vehicle" wordt soms datgene waarmee wordt vergeleken de metafoor genoemd, zodat we dan kunnen spreken over de vergelijking van de les en de metafoor.

Naast de onderscheiding in metaforen, analogieën, en twee soorten vergelijkingen ("comparisons" en "similes") worden in de literatuur nog andere onderscheidingen gemaakt. Shibles (1974) onderscheidde bijvoorbeeld 21 verschillende soorten metaforen: onder andere visuele, synaesthetische, "unifying", therapeutische, fundamentele, "mixed", dode, verhelderende (enzovoorts). Hoewel deze onderscheidingen heuristische waarde kunnen hebben bij het bedenken van metaforen en analogieën, is het weinig zinvol over deze onderscheidingen hier verder uit te weiden. Onderscheidingen die in het kader van het gebruik van metaforen in het onderwijs meer van belang lijken, zijn de volgende:

- a. Samenhangende versus niet-samenhangende vergelijkingen. Wordt de lesstof vergeleken met allerlei met elkaar samenhangende verschijnselen (als in een dubbele hiërarchie (Davidson, 1976)) of met verschillende verschijnselen uit verschillende vakgebieden (zie hoofdstuk 8.2)?
- b. Concrete versus abstracte metaforen. Wordt de (abstracte) lesstof vergeleken met meer concrete verschijnselen of met even abstracte?
- c. Oude versus nieuwe vergelijkingen. Wordt de lesstof vergeleken met verschijnselen of begrippen die al bekend zijn bij de leerlingen of met nieuwe stof die gemakkelijker te begrijpen of te leren is dan de lesstof?
- d. Voor de hand liggende versus niet voor de hand liggende vergelijkingen. Zijn de begrippen of verschijnselen waarmee de lesstof wordt vergeleken potentieel verwarrend, omdat de vergelijkingen zo voor de hand liggen dat ze spontaan door leerlingen worden uitgevoerd of betreft het begrippen die minder voor de hand liggen (zie hoofdstuk 8.1)?

Gick en Holyoak (1980) hebben onlangs gewezen op het belang van een conceptueel kader met betrekking tot de rol van analogieën bij het oplossen van problemen: "Hoe representeren mensen analogieën?" en "Hoe verloopt het proces van "analogical problem solving?". Zij bestudeerden de invloed van een beschrijving in verhaalvorm van een probleem met de bijbehorende oplossing (verhaal-probleem) op het oplossen van een ander

probleem (doel-probleem). Gick en Holyoak presenteerden een representatiesysteem dat voortbouwt op de ideeën van macrostructuren van Kintsch en Van Dijk (zie Van Dijk, 1977). Een belangrijke variabele bij zo'n representatiesysteem vormt het abstractieniveau waarop de analogie wordt gerepresenteerd. Hoe lager het abstractieniveau, hoe minder "kloppend" de analogie is. Bij "analogical problem solving" moeten - aldus Gick en Holyoak - tenminste de volgende stappen in het informatieverwerkingsproces worden onderscheiden:

1. Representatie van het verhaal-probleem (de "vehicle") en van het doel-probleem ("topic").
2. "Mapping" (zie Sternberg, 1977) van het verhaal-probleem in het doel-probleem.
3. Gebruikmaking van de "mapping" om de parallele oplossing voor het doel-probleem te genereren, bijvoorbeeld door voor elk van de oplossingen van het verhaal-probleem een corresponderende oplossing voor het doel-probleem te bedenken.

Omdat het begrijpen van een tekst in vele opzichten een proces van "problem solving" is (Frijda, 1978), mag worden verwacht, dat het proces van vergelijkend leren in zekere mate analoog is aan het proces van analogical problem solving. De opmerkingen van Gick en Holyoak met betrekking tot representatie van een analogie en met betrekking tot het abstractieniveau van de macrostructuren, gelden mijns inziens ook in het geval van vergelijkend leren. De stappen in het informatieverwerkingsproces zijn bij vergelijkend leren waarschijnlijk afwijkend van die bij "analogical problem solving", omdat de representatie van de leerstof op zich niet plaats vindt voor de "mapping" van de analogie, maar gelijktijdig.

3.2 Waarom metaforen belangrijk zijn: theoretische functies

3.2.1 Ortony's argumenten

In een artikel met de veelzeggende titel "Why metaphors are necessary and not just nice" voerde Ortony (1975) drie argumenten aan ter beantwoording van die vraag: het compactheidsargument (compactness-thesis), het onbenoembaarheidsargument

(inexpressibility-thesis) en het levendighedsargument (vividness-thesis).

a. Het compactheidsargument

Metaforen maken het mogelijk in één of twee woorden een lange lijst van kenmerken over te dragen, die bij een letterlijk taalgebruik zeer vele woorden zouden vergen. Ortony gebruikte de volgende vergelijking als voorbeeld: "Hij dook het ijskoude water in als een onverschrokken strijder". Wat de vergelijking in feite eigenlijk zegt is: neem al die aspecten die kenmerkend zijn voor onverschrokken strijders die redelijkerwijs kunnen worden toegepast op duikende zwemmers en ken deze toe aan de zwemmer. In concreto wordt dus eigenlijk in de voorbeeld-zin gezegd: "Hij dook het ijskoude water in dapper, sterk, zonder angst, agressief, op een bepaalde manier, gespierd, groot (enzovoorts)".

De redenering die Ortony volgde, is gebaseerd op een reconstructivistische theorie over het begrijpen van taaluitingen. Volgens een dergelijke theorie maakt men bij het begrijpen van taal uitvoerig gebruik van zijn "kennis van de wereld". Tijdens het begrijpen van taaluitingen "reconstrueert" men als het ware de gebeurtenis die wordt beschreven, waarbij men uitvoerig gebruik maakt van wat men al weet van de beschreven gebeurtenis (zowel uit de context als uit algemene kennis van de wereld). Vaak zal een dergelijke reconstructie onder andere bestaan uit een mentale voorstelling. Het invullen van de details die niet expliciet in de tekst worden gegeven en het vormen van een mentale voorstelling van de gebeurtenis(sen) die in de tekst worden beschreven, noemt Ortony "particularisatie". Particularisatie betreft dus het invullen van de details tussen de linguïstische handwijzers ("sign posts") die in de tekst aanwezig zijn. Particularisatie is volgens Ortony niet alleen een essentieel onderdeel van het begrijpen van taal, maar vooral ook van het begrijpen van analogieën: "It is the language-comprehender's digital-to-analog-converter" (p. 47). Empirische evidentie voor het belang van particu-

larisatieprocessen bij het begrijpen van taal en de gevolgen daarvan voor de opgebouwde geheugenrepresentaties is onder andere te vinden bij Anderson en Ortony (1975).

Een aardige metafoor en een voorbeeld van een particularisatieproces worden beschreven in het volgende citaat van Anderson en Bower (1973, p. 460): "Our words spoken to a listener are like the cryptic directions a playwright provides for a play director, from which a competent director is expected to construct an entire setting, an expressive mood, or an action episode. To illustrate, in the course of reading a story, you might read the sentence "James Bond ran to his car and drove to the casino". As you read you can concretize that sentence by bringing to bear all sorts of facts and sensory images about running, about getting into cars, about driving, and so forth... Wat the sentence does is merely mention a couple of signposts (source, instrument, goal) along the way in the description of an event sequence; the listener interpolates or fills in all the interstitial events between the mentioned signposts." Het voordeel van particularisatie is dat taalbegrip mogelijk is zonder dat alle details expliciet behoeven te zijn uitgewerkt.

Metaforen richten en beperken het particularisatieproces en zorgen ervoor, dat met weinig woorden een keten van kenmerken en details wordt overgedragen. Het zijn dus (ook in het onderwijs) bijzonder efficiënte manieren om informatie over te dragen.

b. Het onbenoembaarheidsargument

Allerlei objecten, ideeën en gebeurtenissen hebben (in een bepaalde taal) geen naam: Er hadden wel woorden voor kunnen zijn, maar ze zijn er gewoonweg (nog) niet. Het komt ook voor, dat de ons ter beschikking staande woorden om een of andere reden niet kunnen worden gebruikt om iets uit te drukken of uit te leggen. Dit kunnen redenen van emotionele aard zijn. Denk bijvoorbeeld aan de ouder die zijn/haar kind moet uitleggen hoe de voortplanting in zijn werk gaat en die daarvoor de metafoor van de bijtjes en de bloemetjes han-

teert. Het kunnen ook redenen van meer cognitieve aard zijn. Bijvoorbeeld in het geval, dat ter beschikking staande woorden te moeilijk of te abstract zijn voor de luisteraar die we iets willen vertellen. Het onbenoembaarheidsargument komt onder andere voort uit een door Ortony en anderen aangehangen stelling waarin wordt gepostuleerd, dat wat wij als bewuste waarnemers ervaren continu van karakter is, terwijl de taal een discreet symbolensysteem is. Metaforen kunnen een hulpmiddel zijn om dingen die zich niet direct of moeilijk in taal laten uitdrukken via een indirecte weg naar een lezer/luisteraar te communiceren.

Waar in het onderwijs de letterlijke taal tekort schiet, moet metaforische taal het overnemen of in de woorden van Locke: "Whereof one cannot speak literally, thereof one should speak metaphorically" (Ortony, 1975, p. 49).

c. Het levendighedsargument

Metaforen zijn bijzonder levendig, omdat ze dichters staan bij de "continue ervaring" en bij de emotionele werkelijkheid. Ze brengen ons dichters bij de continue ervaring, doordat het proces van particularisatie wordt gericht en doordat in enkele woorden een hele keten van kenmerken wordt geïmpliceerd zonder dat deze in discrete taal hoeft te worden omgezet. Ze brengen ons om dezelfde redenen ook dichters bij de emotionele werkelijkheid.

Hun onderwijskundige kracht ontleen metaforen aan het feit dat ze zorgen voor levendiger en meer gedetailleerde visuele voorstellingen (images). Paivio (1971) en anderen toonden aan, dat bepaalde eigenschappen van voorstellingen nauw samenhangen met het gemak waarmee bijvoorbeeld woordjes kunnen worden geleerd. Doordat metaforen zorgen voor levendige en gedetailleerde "images" zorgen ze voor beter begrip en betere retentie. Verder ontleen metaforen hun onderwijskundige kracht aan hun grotere beschikbaarheid ("availability") dankzij de emotionele aspecten ervan (vergelijk ook Blumenthal, 1977).

3.2.2 Metaforen en schematheorieën

Recentelijk is er een opkomst waar te nemen van psychologische en psycholinguïstische theorieën, waarin het begrip "schema" of "frame" centraal staat (zie bijvoorbeeld Anderson, Spiro en Montague, 1977). Voorlopers van de meer recente schematheorieën vormden de theorieën van Barlett en Piaget. Barlett (1932) definieerde een schema als een actieve organisatie van vroegere reacties en vroegere ervaringen. Piaget (Flavell, 1963) daarentegen definieerde een schema als een cognitieve structuur die betrekking heeft op een klasse van vergelijkbare handelingssequenties met sterk met elkaar samenhangende gedragselementen (die een sterke eenheid vormen). Netelenbos en Van Rossum (1979) stellen, dat in de opvatting van Bartlett schemata een soort gemiddelde van opgedane ervaringen, een centrale tendentie vormen met een interne, organiserende en controlerende functie, terwijl een schema bij Piaget meer een initiërende functie heeft. In de recentere literatuur kan men nog diverse andere definities van schemata aantreffen. Sommigen stellen schemata gelijk aan stereotypes en prototypes (Hyman, 1976), anderen benadrukken dat schemata kennisstructuren vormen (Anderson, 1978) en weer anderen benadrukken net als Piaget het handelingskarakter van schemata (Neisser, 1976).

Een algemene definitie van schemata werd gegeven door Rummelhart en Ortony (1977): "Schemata are data structures for representing the generic concepts stored in memory. They exist for generalized underlying objects, situations, events, sequences of events, actions and sequences of actions. Schemata are not atomic. A schema contains as part of its specification, the network of interrelations that is believed to generally hold among the constituents of the concept in question. Schemata, in some sense, represent stereotypes of these concepts."

Veel auteurs omschrijven het begrip schema niet, maar volstaan met vergelijkingen met scripts van toneelstukken of formaten ("formats") van computerprogramma's. Kennelijk zijn schema's zo moeilijk te definiëren, dat metaforen te hulp moeten worden geroepen (vergelijk Ortony's onbenoembaarheidsargument in

3.2.1). Ondanks de vaagheid en verscheidenheid van de definieringen, lijkt de opkomst van schematheorieën niet te stuiten.

Vanuit een onderwijspsychologisch perspectief is vooral van belang welke rol schemata spelen bij het leren en hoe schemata tot stand komen en veranderen (zie Brandsford et al., 1977). Volgens Norman (1978) komt kennis op drie verschillende wijzen tot stand, namelijk via "accretion", "restructuring" en "tuning".

Van accretion is sprake als nieuwe kennis wordt verworven binnen een bestaand schema of gestuurd door een bestaand schema. In het eerste geval worden lege plaatsen in een schema gevuld of worden ongespecificeerde argumenten in een schema gespecificeerd. Accretion is dan het invullen van een bestaande structuur met informatie. In het tweede geval wordt een nieuw schema geconstrueerd, waarbij een bestaand schema als uitgangspunt of prototype wordt genomen.

Van restructuring is sprake als er nieuw inzicht in de structuur van een onderwerp ontstaat. Zonder dat nieuwe informatie hoeft te worden toegevoegd, kan een lerende reeds bekende kennis anders gaan zien of beter gaan begrijpen. Vaak zullen dergelijke herstructureringen (aldus Norman) tot stand komen doordat analogieën tussen verschillende schemata worden gezien.

Tuning, tenslotte, heeft betrekking op de laatste fasen van het leerproces, waarin kennis wordt geautomatiseerd en verfijnd. Onnodige berekeningen worden verwijderd, onnodige variabelen worden vervangen door vaste waarden. Voor tuning is het nodig dat kennis steeds maar weer wordt gebruikt in praktische toepassingen en bij het oplossen van problemen.

Metaforen zijn volgens Norman (1978; Norman et al., 1976) in twee fasen van kennisverwerving van belang: tijdens "accretion" en tijdens "restructuring". Bij het tot stand brengen van een nieuw schema (accretion) kan men een oud schema als basismateriaal nemen waarop het nieuwe schema wordt gebouwd. Daarnaast zijn analogieën tussen verschillende kennisgebieden de belangrijkste initiators van herstructurering van een schema. Ook Rummelhart en Ortony (1977) wezen op de mogelijke rol van metaforen en analogieën bij de vorming van nieuwe schema-

ta: "...One might regard metaphors as aids to selecting an old schema, which with relatively little modification, can be used to produce a new one. One might use a "flowing water" schema as the basis for the generation of an "electric current" schema. The former might incorporate knowledge concerning unidirectionality of flow, branching, capacity of the schema, and so on, all of which would have their analogs in an "electric flow" schema. Good instruction would clarify the metaphor of electricity in wires as water in pipes by specifying which variables stay and which go. What makes it a metaphor, after all, is that some of the new information will not fit into the old schema." (p. 132).

3.2.3 Metaforen en Ausubel's subsumptietheorie

3.2.3.1 Ausubel's theorie in een notedop

De theorie van Ausubel is in essentie een consequente uitwerking van één centrale assumptie: de bestaande cognitieve structuur van de leerling vormt de belangrijkste variabele in nieuw leren. Of in de woorden van Ausubel et al. (1978): "If we had to reduce all of educational psychology to just a single principle, we would say this: "Find out what the learner already knows and teach him or her accordingly". (p. 373)

De meeste hypothesen in de theorie zijn uitwerkingen van deze centrale assumptie. Een daarvan is bijvoorbeeld dat de mate en kwaliteit van voorkennis (of in Ausubel's termen: De relatieve stabiliteit van de "ankerideeën" in de cognitieve structuur) bepalend zijn voor nieuw leren. Een andere deductie betreft de onderscheidbaarheid (discriminability) van de ideeën in de bestaande cognitieve structuur en de nieuw te leren ideeën: Nieuw leren wordt bemoeilijkt door een te geringe onderscheidbaarheid. Ook het meest bekende principe in de theorie betreffende de "advance organizers" volgt uit de centrale assumptie: Als de bestaande cognitieve structuur onvoldoende aanknopingspunten biedt om nieuw leren zo optimaal mogelijk te maken, dan kunnen advance organizers de kloof overbruggen tussen wat de leerling al weet en wat hij moet weten om het nieuwe materiaal te leren (Ausubel et al., 1978,

p. 171). Ook de overige principes (zoals progressieve differentiatie en "integrative reconciliation") volgen uit de centrale assumptie.

Ausubel onderscheidde drie wijzen om nieuwe kennis aan reeds bekende concepties te relateren: subordinate subsumption, ^{Super}subordinate en combinatorial learning.

Bij subordinate subsumption worden nieuwe concepten en principes in verband gebracht met abstractere en algemenere concepten en principes in de cognitieve structuur op het betreffende terrein (bijvoorbeeld leren dat (nieuwe kennis) het uithangen van de vlag een daad van patriotisme (oude kennis) is).

Bij ^{Super}subordinate learning wordt precies de omgekeerde weg bewandeld: Het leren van een abstractere en algemenere propositie waaronder verscheidene in de cognitieve structuur aanwezige concepten of proposities kunnen ressorteren (bijvoorbeeld leren dat de bekende begrippen worteltjes, spinazie en bloemkool groenten (nieuwe kennis) zijn).

Tot combinatorial learning moet men vaak zijn toevlucht nemen als er noch bovengeschiede noch ondergeschikte concepten en proposities in de cognitieve structuur aanwezig zijn, die als ankerideeën kunnen fungeren. Nieuwe concepten en proposities worden nu niet gerelateerd aan één bepaald relevant idee in de cognitieve structuur, maar aan een brede achtergrond van algemeen relevante inhouden in de cognitieve structuur (bijvoorbeeld het leren van de relaties tussen massa en energie, hitte en volume, genetische variabiliteit en genetische structuur enzovoorts).

Subordinate subsumption is volgens Ausubel et al. (1978) de meest efficiënte vorm van leren, omdat de algemene ideeën die in dit geval het eerst worden geleerd, als ze eenmaal voldoende worden beheerst, de volgende voordelen hebben:

- "1. They have maximally specific and direct relevance for subsequent learning tasks.
2. They possess enough explanatory power to render otherwise arbitrary factual detail potentially meaningful.

3. They possess sufficient inherent stability to provide the firmest type of anchorage for newly learned detailed meanings.
4. They organize related new facts around a common theme, thereby integrating the component elements of new knowledge, both with each other and with existing knowledge."

3.2.3.2 Advance organizers en hun voordelen volgens Ausubel

De besproken vormen van leren betreffen de manieren waarop leerlingen zelf nieuwe ideeën relateren aan bestaande ideeën. In sommige gevallen - aldus Ausubel - is het nodig, dat leerlingen worden geholpen de kloof te overbruggen tussen wat zij weten en wat zij moeten weten om nieuwe informatie te leren. In dergelijke gevallen moet gebruik worden gemaakt van een advance organizer. Dat is "introductory material presented in advance of and at a higher level of generality, inclusiveness and abstraction than the learning task itself; and explicitness related both to existing relevant ideas in cognitive structure and to the learning task itself..." (Ausubel et al., 1978, p. 628).

Een "expository organizer" bevat bovengeschikte (superordinate) concepten en principes waaraan de begrippen en principes uit de tekst kunnen worden gerelateerd. In een "comparative organizer" wordt expliciet aangegeven, waarin het nieuwe materiaal overeenkomt met en verschillend is van soortgelijke concepten en proposities in de cognitieve structuur. Expository organizers worden vooral gebruikt bij relatief onbekend materiaal en comparative organizers bij relatief bekend materiaal. Comparative organizers hebben vooral tot doel te voorkomen, dat het nieuwe materiaal wordt verward met relevante ankerideeën. Ze vergroten de onderscheidbaarheid (discriminability) van oude en nieuwe ideeën. Ze hebben geen zin als de relevante ankerideeën al helder en stabiel zijn, omdat de onderscheidbaarheid van een nieuwe leertaak in hoge mate een functie is van de helderheid en stabiliteit van de bestaande ankerideeën (Ausubel et al., 1978, p. 183).

De beweegredenen voor het gebruik van advance organizers zijn de volgende drie:

- "1. The importance of having relevant and otherwise appropriate established ideas already available in cognitive structure to make logically meaningful new ideas potentially meaningful and to give them stable anchorage.
2. The advantage of using the more general and inclusive ideas of a discipline as the anchoring ideas of subsumers (namely the aptness and specificity of their relevance, their greater inherent stability, their greater explanatory power and their integrative capacity).
3. The fact that they themselves attempt both to identify already existing relevant content in cognitive structure (and to be explicitly related to it) and to indicate explicitly both the relevance of the latter content and their own relevance for the learning material." (Ausubel et al., 1978, p. 171).

3.2.3.3 De plaats van metaforen in de theorie van Ausubel

Naar een verhandeling over het leren via metaforen zal men in de geschriften van Ausubel tevergeefs zoeken. Toch treft men in studieboeken waarin de theorie van Ausubel wordt samengevat (bijvoorbeeld Klausmeier, 1975) en in artikelen van volgelingen van Ausubel (bijvoorbeeld Mayer, 1979a, 1979b) voorbeelden aan die betrekking hebben op leren met behulp van metaforen. Zo gaf Klausmeier het volgende voorbeeld van combinatorial learning: "As an example, a child may learn how atoms function by observing a physical model that portrays their function. Similarly the concept of equation may be formed by observing a seesaw." (p. 42). Evenzo schreef Mayer (1979b, p. 162): "In the present studies, good organizers have been concrete models or analogies or examples, sets of higher rules and discussions of the main themes in familiar terms..."

Klausmeier maakte geen onderscheid tussen combinatorial learning en koppeling van nieuwe kennis aan analoge kennis ("metaforische subsumptie"). Mayer rekende analogieën en concrete modellen tot de advance organizers. Beide generalisaties

zijn mijns inziens onjuist (zie ook Ausubel, 1978). Metaforische subsumptie past niet bij de omschrijving die Ausubel van combinatorial learning gaf (zie 3.2.3.1).

Bovendien geldt het voornaamste nadeel van combinatorial learning niet voor metaforische subsumptie. Dat nadeel is volgens Ausubel et al. (1978, p. 59), dat slechts verankering aan algemene niet specifieke, relevante inhouden in de cognitieve structuur kan plaats vinden en geen verankering aan bijzondere specifieke, relevante ideeën. Daardoor kunnen proposities minder goed worden gerelateerd aan eerder verworven kennis met als gevolg dat ze moeilijker zijn te leren en te onthouden. Dit nadeel geldt niet voor metaforische subsumptie, omdat daarbij juist wel verankering aan specifieke, relevante ideeën in de cognitieve structuur kan plaats vinden.

Ook het gelijkstellen van analogieën en concrete modellen met advance organizers is problematisch. Ausubel definieerde organizers immers expliciet als algemener, omvattender en abstracter dan de leertaak (zie 3.2.3.2, 8.1.1 en Mayer en Bromage, 1980). De voor- en nadelen van metaforische subsumptie en metaforische organizers kunnen dan ook niet rechtstreeks uit de theorie van Ausubel worden afgeleid.

Via extrapolatie kan echter aannemelijk worden gemaakt, dat metaforische subsumptie drie van de vier functies van subordinate subsumptie (zie 3.2.1) ook kan vervullen: De begrippen en proposities waarmee wordt vergeleken, kunnen ook:

1. voldoende verklarende kracht hebben om anders willekeurig feitenmateriaal potentieel betekenisvol te maken;
2. voldoende inherent stabiel zijn om vaste ankerpunten te bieden voor de nieuw te leren betekenissen;
3. verwante feiten organiseren rond een gemeenschappelijk thema, waarbij ze de constituerende elementen van nieuwe kennis integreren, zowel met elkaar als met bestaande kennis.

Evenzo is het aannemelijk, dat metaforen als concrete advance organizers:

1. kunnen zorgen voor verankering aan heldere, stabiele ankerideeën in de cognitieve structuur;

2. expliciet relevante inhouden in de cognitieve structuur kunnen actualiseren, die de leerlingen bij hun subsumptie kunnen gebruiken.

(Vergelijk de voordelen 1 en 3 van organizers in 3.2.3.2.)

3.2.4 Metaforen en de assimilatie-coderingstheorie van Mayer

Volgens Mayer (1977) is het ontstaan van een "meaningful learning set" (vergelijk Ausubel et al., 1978) afhankelijk van drie condities: a. receptie van het te leren materiaal; b. beschikbaarheid van een betekenisvolle structuur van vertrouwde ideeën die kunnen worden gebruikt om het nieuwe materiaal te organiseren en te assimileren en c. activatie van deze betekenisvolle structuur tijdens het leren.

Receptie heeft te maken met ontvangst van de informatie in het werkgeheugen. Beschikbaarheid betreft de aanwezigheid van bepaalde concepten (ankerideeën) in het lange termijn geheugen en activatie vindt plaats doordat in het werkgeheugen een integratie plaats vindt van de nieuwe informatie en ideeën uit het lange termijn geheugen (Mayer, 1979a).

Advance organizers werden door Mayer ruimer gedefiniëerd dan door Ausubel: "An advance organizer is defined as a stimulus (usually a prose passage) that: a. is presented prior to learning and b. contains a system for logically organizing the incoming information into a unified structure. The advance organizer may be abstract (such as giving a list of general rules or principles) or concrete (such as a physical analogy), but it must serve to logically integrate the new information."

(Mayer en Bromage, 1980, p. 211).

Mayer's theorie over advance organizers heeft derhalve ook betrekking op metaforen en analogieën als organizers. Volgens de assimilatie-coderingstheorie (Mayer, 1979a, 1979b; Mayer en Bromage, 1980) beïnvloeden organizers vooral de manier waarop de nieuwe informatie wordt gecodeerd. Doordat ze een betekenisvolle assimilatieve context beschikbaar maken (vergelijk conditie b in het bovenstaande) en doordat leerlingen actief de nieuwe en de oude informatie integreren (vergelijk conditie c in het bovenstaande) resulteren andere, namelijk meer geïntegreerde leeruitkomsten.

De assimilatie-coderingstheorie werd door Mayer (zie Mayer en Bromage, 1980) afgezet tegen een viertal alternatieve theorieën:

- a. De retrievaltheorie die stelt, dat de organizers in de eerste plaats als een "retrieval-aid" fungeren.
- b. De receptietheorie die voorspelt, dat organizers geen effect hebben omdat leerlingen alleen leren wat wordt gepresenteerd en ontvangen zonder dat er van integratie of assimilatie sprake is.
- c. De additie-coderingstheorie volgens welke organizers het coderen van de informatie vergemakkelijken doordat er meer ankerideeën beschikbaar worden gemaakt om de nieuwe informatie aan te koppelen zonder dat echter de leeruitkomsten door integratie in het werkgeheugen kwalitatief anders worden.
- d. De selectieve additie-coderingstheorie die stoelt op het idee dat de ankerideeën die beschikbaar worden gemaakt alleen het coderen van een gedeelte van de informatie beïnvloeden.

Het voornaamste verschil tussen de assimilatie-coderingstheorie en de beide additietheorieën is gelegen in de aard van het vooronderstelde integratie- en assimilatieproces. Voorstellen de additietheorieën een totaaleffect van organizers, de assimilatie-coderingstheorie voorspelt een "Treatment x Type of test interaction": De organizer leidt volgens deze theorie tot meer geïntegreerde leeruitkomsten (die tot uiting komen in transfer- en toepassingstoetsen) maar niet noodzakelijk tot meer kennis.

Belangrijke functies die metaforen volgens de theorie van Mayer - onder expliciet omschreven condities (zie Mayer 1979b) - kunnen vervullen, zijn:

- a. het beschikbaar maken van een assimilatieve context;
- b. het activeren van die assimilatieve context tijdens het leren met als gevolg meer geïntegreerde leeruitkomsten;
- c. het organiseren en structureren van de binnenkomende informatie in "a unified structure".

3.2.5 Metaforen en de generatieve theorie van Wittrock

Wittrock (1977) stelde een model voor, waarin begrijpend leren wordt opgevat als een generatief cognitief proces: "That is, learning involves the active construction of meaning for stimuli, using verbal processing, imaginal processing (or propositional and appositional processing), and perhaps other types of processing. From this point of view, it is plausible that learning is basically a process of relating stimuli to previous experience, from which one induces and elaborates meanings and representations. According to this model, learning with understanding is the process of transferring previous experience to new events and problems." (pp. 176-177).

Volgens Wittrock is bij begrijpend leren het actief relateren van de nieuwe informatie aan oude informatie de belangrijkste variabele. Begrijpen van informatie hangt af van de cognitieve omvormingen en uitbreidingen die lerenden op stimuli uitvoeren. Evidentie ten gunste van dit generatieve model werd beschreven door Wittrock en Lumsdaine (1977).

Eén van de implicaties van het model is, dat onderwijs zo moet worden opgezet dat de actieve constructie van mentale uitbreidingen door de lerende zo veel mogelijk wordt vergemakkelijkt. Eén van de manieren waarop dit volgens Wittrock (1979) kan gebeuren is via metaforen en analogieën. Want deze didactische hulpmiddelen kunnen bestaande geheugenrepresentaties relateren aan of vertalen naar nieuwe informatie.

Het is volgens Wittrock (1979) echter afhankelijk van leerlingkenmerken, strategieën en specifieke geheugenstructuren van de lerende, hoe de instructie het best kan plaats vinden. Als lerenden - met of zonder instructies - niet in staat zijn of niet spontaan geneigd zijn relevante relaties tussen informatie en ervaring te genereren, dan kunnen deze relaties worden geëxpliciteerd of gesuggereerd door onder andere metaforen en analogieën. Als lerenden echter die relaties zelf kunnen leggen, komen andere technieken (als vragen, discussies, tekeningen) meer in aanmerking.

Impliciet werden in het model van Wittrock twee mogelijke functies van metaforen gesuggereerd:

- a. Metaforen kunnen een actievere constructie van betekenis door de lerende bewerkstelligen, namelijk wanneer deze zelf niet in staat of geneigd is relaties tussen nieuwe informatie en bestaande kennis te leggen.
- b. Metaforen kunnen informatie beschikbaar maken waaraan nieuwe informatie kan worden gerelateerd.

3.2.6 Metaforen en de conversatietheorie van Pask

Volgens Pask is de fundamentele eenheid voor het onderzoek naar complex menselijk leren een "conversatie" waarin communicatie plaats vindt tussen twee participanten in het leerproces (lerende en leraar) (Pask, 1976b). Deze conversaties kunnen plaats vinden tussen twee individuen (lerende en leraar) maar ook binnen één menselijk brein. Sommige functies van de leraar kunnen ook door machines worden uitgevoerd. In theorie en onderzoek van Pask worden bepaalde strikt gedefinieerde soorten conversaties onderzocht. Voor deze conversaties geldt namelijk dat:

1. alle gespreksonderwerpen tot een van te voren vastgesteld domein behoren;
2. het gesprek in episodes verloopt die begrensd worden door "understandings" (Vastenhouw, 1979, p. 241).

Het van te voren vastgestelde domein van gespreksonderwerpen wordt op twee manieren gerepresenteerd. Enerzijds is er een begrippenschema waarin de te leren onderwerpen en relaties in een soort pijlenschema zijn samengevat. Via dit begrippenschema (entailment structure) waarover de lerende en de leraar beschikken, kunnen onderwerpen worden geïdentificeerd en kunnen wensen over volgorde van bestudering en dergelijke kenbaar worden gemaakt.

Anderzijds is er de zogenaamde taakstructuur (task structure) waarin de leerstof op een andere wijze is gerepresenteerd. In de taakstructuur zijn de relaties die in het begrippenschema zijn samengevat inhoudelijk beschreven. Een relatie in de taakstructuur is een constructie van deze relatie in de fysieke wereld (Daniël, 1975, p. 416). In de taakstructuur is ge-

(zie ook hoofdstuk 7.1.2). "Globetrotting" is het verkeerd begrijpen van valide analogierelaties en/of het gebruik maken van onjuiste analogieën. "Improvvidence" is het voorbij gaan aan valide analogieën of het ten onrechte niet gebruik maken van een gemeenschappelijk principe. Dus ook in deze afwijkingen spelen de analogierelaties een belangrijke rol.

Volgens Pask vormt ook het onderkennen en gebruik maken van analogieën in leerstof een belangrijke component van "understanding". Daarnaast toont hij zich recentelijk ook voorstander van het gebruik van minder rigide analogieën als allegorieën, parabels en verhalen. "Further analysis of the entailment graph data reaffirms the outstanding significance of analogies, in general as determinants of effective learning... Moreover the practical significance of allegories and stories deserve emphasis. Recalling the proposition of 2.5 (11) of this report, stories (plays, television dramas), are the most potent means of securing "understanding" in a "one way" communication educational system." (Pask et al., 1978, p. 99).

"...The valid analogies which promote rapid assimilation of meaningful data are often of a much more complex, idiosyncratic type than the material analogies of form so far discussed." (Pask et al., 1977, p. 136).

Het mogelijk maken van "understanding" (echt begrip) kan derhalve volgens Pask een belangrijke functie van metaforen zijn.

3.2.7 Metaforen en de "hypostatization"-hypothese van Davidson
 "Hypostatizations" hebben volgens Davidson (1976) betrekking op cognitieve transformatieprocessen, die in werking kunnen worden gesteld wanneer de lerende wordt geplaatst voor complexe of abstracte leertaken. Het doel van zulke processen is het transformeren van complexe of abstracte informatie in vormen die de informatie begrijpelijker of gemakkelijker te leren maken. Een lerende kan bijvoorbeeld proberen een abstract concept te begrijpen in termen van een vertrouwd concreet concept. "Hypostatization" is een, in essentie, cognitieve activiteit van metaforische aard. Davidson onderscheidde vier vormen: de

representeerd wat er met betrekking tot de begrippen en de relaties daartussen moet worden geleerd.

Het begrippenschema (de "entailment structure") wordt in een aantal stappen geconstrueerd op basis van gesprekken met deskundigen. Eerst wordt een groot aantal onderwerpen en verbindende lijnen daartussen getekend (de "entailmentmesh"). Vervolgens worden vereenvoudigingen aangebracht en wordt het begrippenschema zoveel mogelijk ingedikt. In deze fase gaan analogieën een belangrijke rol spelen: Wanneer men namelijk analoge relatiesystemen kan aangeven, kan het begrippensysteem aanzienlijk worden vereenvoudigd. De taakstructuur wordt vervolgens geconstrueerd op basis van het eenvoudige begrippenschema.

Een van de voorwaarden voor de "strikte" conversaties was, dat het gesprek in episodes verloopt die worden begrensd door echt begrip (understanding). Voor "echt begrip" is het volgens Pask (1976b) niet voldoende dat de lerende een verbale verklaring geeft. De lerende moet - gebruik makend van het begrippenschema - de onderliggende relaties op bepaalde nauwkeurig omschreven wijzen gebruiken. Ook hierbij spelen analogierelaties een belangrijke rol.

Uit het onderzoek van Pask (1976c) is gebleken, dat lerenden van elkaar verschillen in de wijze waarop zij gebruik maken van de analogieën in de leerstof en dat dit bepalend is voor de efficiëntie van het leren. Zie bijvoorbeeld de volgende twee citaten:

" However, scrutiny of the records, either in the multi purpose or the school based experiments, discloses an interesting fact. Students who learn rapidly are students who use (and understand) valid analogy relations." (Pask, 1976a, p. 129).

"...analogy relations are the glue required to make sense of an otherwise chaotic reality and to stick together theories (and, a fortiori, scientific theories) which otherwise are disparate essays bringing order only to small regions of what may be known." (Pask, 1976a, p. 130).

Twee belangrijke afwijkingen die Pask (1976c) in leerprocessen constateerde, waren "globetrotting" en "improvidence"

verbaal-symbolische metafoor, de visuele-"imagery" metafoor en de verbale en de pictoriale analogie. De belangrijkste functie van metaforen is volgens Davidson het concretiseren van abstracte informatie.

3.2.8 Samenvatting

Metaforen zijn om de volgende redenen belangrijke onderwijskundige hulpmiddelen:

1. Het zijn efficiënte manieren van informatie-overdracht (het compactheidsargument van Ortony).
2. Ze maken overdraagbaar wat niet letterlijk in woorden kan worden uitgedrukt (het onbenoembaarheidsargument van Ortony).
3. Ze staan dicht bij de perceptuele en/of emotionele werkelijkheid dan meer letterlijke taal (het levendighedsargument van Ortony).
4. Ze zijn in het bijzonder behulpzaam bij het produceren van nieuwe schemata (schematheorieën).
5. Ze zorgen ervoor, dat verankering van nieuwe informatie aan bestaande stabiele, heldere ankerideeën mogelijk wordt (Ausubel en Mayer).
6. Ze actualiseren bij de leerlingen relevante inhoud in de cognitieve structuur, die ze als ankerideeën kunnen gebruiken (Ausubel, Mayer en Wittrock).
7. Ze organiseren verwante feiten rond een gemeenschappelijk thema, waarbij ze de constituerende elementen van de nieuwe kennis integreren (Ausubel en Mayer).
8. Ze structureren binnenkomende informatie in "a unified structure" (Mayer).
9. Ze maken een actievere constructie van betekenis door de lerende mogelijk (Wittrock).
10. Ze kunnen zorgen voor echt begrijpen (Pask).
11. Ze kunnen abstracte informatie concreet maken (Davidson).

Belangrijke functies die metaforen in het onderwijsleerproces kunnen vervullen zijn op grond van het bovenstaande:

- a. Het structureren van de te leren informatie (Mayer, Ausubel, schematheorieën).

- b. Het vergemakkelijken van mentale voorstellingen c.q. het concreet maken van abstracte informatie (Ortony, Davidson).
- c. Het stimuleren van een actiever verwerkings- en integratieproces door de lerende (Wittrock, Mayer).

3.3 Argumenten tegen metaforen

In de loop van de geschiedenis zijn er door verschillende filosofen argumenten aangevoerd tegen het gebruik van metaforen. Daniel (1975) refereerde bijvoorbeeld aan een boek van een Fransman genaamd Bachelard, die een studie maakte van het onverantwoorde gebruik van analogieën in de "wetenschappelijke" verhandelingen in de 18e eeuw. Bachelard concludeerde op basis van deze studie dat alle wetenschappers zoveel mogelijk analogieën moesten vermijden. Ortony (1975) verwees naar de 19e eeuwse Franse linguïst Bréal, die beweerde dat analogieën niets nieuws leren en ze vergeleek met "...the sayings of some peasant endowed with good sense and honesty, but not without a certain rustic cunning". Davidson (1976) verwees naar een filosoof die waarschuwde voor de "fallacy of misplaced concreteness".

Naast de vermelde argumenten tegen analogieredeneringen in het algemeen, kunnen ook meer specifieke argumenten tegen het gebruik van metaforen in het onderwijs worden onderscheiden. Zo schreef Miller (1976) bijvoorbeeld een artikel met de titel: "The dubious case for metaphors in educational writing", waarin hij aannemelijk trachtte te maken, dat Ortony's drie argumenten ten gunste van metaforen (het onbenoembaarheidsargument, het compactheidsargument en het levendighedsargument) overdreven en in bepaalde opzichten fout zijn. Hij deed dit door aan de hand van praktijkvoorbeelden te laten zien dat metaforen vaak worden gebruikt "to gloss over matters which cannot be well explained or clearly specified" (p. 174) en dat metaforen vaak op een misleidende wijze worden gebruikt, doordat ze op emotie spelen of een argument steun geven via verdraaiing of overdrijving. "In going beyond literal meaning, metaphors become vague. Similarly, they can be suggestive precisely because they are not exact and do not pin meanings down. Thus, if metaphors are used to suggest, or approximate, the

inexpressible, it would seem they do so at the cost of clarity." (p. 176).

Terecht merkte Ortony (1976) in een "reply to my critics" op, dat voorbeelden van misbruik van metaforen nooit een argument tegen de theoretische argumenten kunnen zijn: "...he appears to confuse the notion of metaphors as a (potentially) powerful pedagogical tool, with the notion of metaphor in educational writing."

Ook Haynes (1975) leverde kritiek op het artikel van Ortony. Hoewel ze het met de strekking van Ortony's betoog eens was, had ze bezwaar tegen het levendighedsargument: Aan de hand van voorbeelden toonde ze aan dat metaforen lang niet altijd levendig zijn. Verder was ze van mening, dat Ortony in zijn pleidooi ten gunste van metaforen niet ver genoeg was gegaan, omdat er geen aandacht werd besteed aan de "click of comprehension" of de "creative leap".

Een ander argument tegen analogieredeneringen in het onderwijs werd beschreven door Kooreman (1976). Bij een analogieredenering (in het spellingsonderwijs) is er een onvolledige oriëntering die tot fouten aanleiding kan geven. Er is een niet geëxpliciteerde identificatie, waardoor een transformatie ten onrechte wordt toegepast (p. 266).

Davidson (1976) meende dat het grootste gevaar van metaforen in het onderwijs schuilt in ongewenste bij-effecten of cognitieve verdraaiingen: "For example, the tendency to imbue an abstraction with substance (c.q. the reification of "intelligence" or "id"), is a particularly virulent form of cognitive distortion." (p. 135). Wanneer men abstracte ideeën bijvoorbeeld met behulp van metaforen concreet maakt, loopt men het gevaar dat leerlingen in dit concrete niveau blijven steken zonder ooit tot echt begrip van de abstracties te komen. Ook in gesprekken met natuurkundedocenten hier te lande bleek deze angst voor "misplaced concreteness" een belangrijke reden te zijn tegen het gebruik van metaforen. Wittrock (1977) zou deze angst waarschijnlijk in verband brengen met de neo-platonische conceptie "the primacy of abstract ideas", die tot gevolg had dat de rol van "imagery" in leren en begrijpen min-

der werd benadrukt dan in de meer aristotelische concepties.

Een argument dat nauw verband houdt met het vorige en dat in de literatuur niet werd aangetroffen, heeft betrekking op de bestendiging van leerlingkenmerken (zie hoofdstuk 2.4 principe 8): Wanneer men concreet ingestelde leerlingen voortdurend verwent met concretisering zou er wel eens bestendiging van die gerichtheid op het concrete kunnen optreden in plaats van de cultivatie van een (gewenste?) gerichtheid op het abstracte. Met andere woorden we moeten leerlingen leren abstract te denken en dan moeten we de leerstof niet te veel concretiseren.

Ook de controverse tussen realistische en "relevant-cue"-theorieën (Parkhurst, 1975) zou men kunnen vertalen in een argument tegen metaforen. Volgens de realistische theorieën, die in de jaren veertig en vijftig opgang deden, gold dat hoe realistischer een onderwijsmethode is hoe efficiënter deze de leerprocessen van de leerlingen zal faciliteren. De "relevant-cue"-theorieën uit de jaren zestig daarentegen voorspelden dat realisme vaak overbodig is. Realisme zorgt - volgens deze theorieën - voor te veel irrelevante, afleidende cues. Parallel aan het bezwaar van de "relevant-cue"-theorieën tegen realisme in leermateriaal, zou men ook kunnen stellen dat metaforen irrelevante, afleidende cues bevatten en daardoor wellicht beter kunnen worden weggelaten.

Peeck (1972) beschreef een aantal problemen die het gebruik van plaatjes in leermateriaal zou kunnen opleveren:

- a. Plaatjes kunnen onjuist geïnterpreteerd worden of kunnen aanleiding geven tot verkeerde gevolgtrekkingen.
- b. Bepaalde inhoud uit de leerstof krijgen door illustraties ongewenst grote aandacht. Deze aandachtsvermeerdering zou ten koste kunnen gaan van andere inhoud.
- c. Het leerresultaat kan negatief worden beïnvloed, doordat het leerproces voortdurend onderbroken en derhalve verstoord wordt telkens als de leerling zich van de tekst tot het plaatje wendt.
- d. Het bekijken van illustraties kost tijd, wat tot slechtere tekstkennis kan leiden, bijvoorbeeld wanneer plaatjes geen relevante informatie bevatten, de leestijd beperkt is of

de leesvaardigheid van de lezer gering is.

Deze problemen met illustraties kunnen mutatis mutandis ook gelden voor metaforen.

Het voorgaande kan worden samengevat in de volgende (mogelijke) argumenten tegen het gebruik van metaforen in het onderwijs:

1. Metaforen kunnen leerlingen in de war of op een dwaalspoor brengen en ze kunnen leerstof onhelder en vaag maken (Miller, Davidson, Kooreman, Peeck, "relevant-cue"-theorieën).
2. Metaforen zijn overbodig ("relevant-cue"-theorieën, Miller).
3. Metaforen kunnen abstracte leerstof te veel concretiseren waardoor leerlingen de abstracte ideeën niet echt gaan begrijpen en bestendiging van een gerichtheid op het concrete en een concrete wijze van denken kan optreden (Davidson; hoofdstuk 2.4).
4. Metaforen kosten tijd die beter anders kan worden besteed (Peeck).
5. Bepaalde inhouden uit de leerstof krijgen door metaforen ongewenst grote aandacht (Peeck).
6. Metaforen kunnen het leerproces onderbreken (Peeck).

3.4 Conclusie

In dit hoofdstuk werden zowel argumenten voor als argumenten tegen het gebruik van metaforen behandeld. Om te kunnen beslissen of het nu wel of niet verstandig is metaforen te gebruiken in het onderwijs (en wanneer wel en wanneer niet) heeft men meer informatie nodig dan deze theoretische, subjectieve argumenten alleen. De discussie tussen voor- en tegenstanders zal alleen een stap verder kunnen komen als er meer empirische informatie beschikbaar komt. Belangrijke vragen voor empirisch onderzoek zouden, met inachtneming van de pro- en contra-argumenten, kunnen zijn:

- Vergroten metaforen het begrip van de leerstof?
- Hoeveel extra tijd kost deze eventuele begripsvergroting?
- Vervullen metaforen de toegedachte functies (structureren, concretiseren, activeren) die in 2.2 werden afgeleid?

- In hoeverre brengen metaforen leerlingen in de war en op een dwaalspoor?
- Onder welke omstandigheden is het gebruik van metaforen wel aan te raden, onder welke niet?

Een andere belangrijke vraag, die voortvloeit uit 3.3, betreft het argument dat metaforen te veel een concrete wijze van denken en een gerichtheid op het concrete zouden bestendigen. Empirisch onderzoek zou moeten uitwijzen of die bestending door het gebruik van metaforen inderdaad meer optreedt dan bij meer abstracte instructiewijzen. Dergelijk onderzoek heeft te kampen met grote methodologische problemen: Hoe kan men dit concreet denken en die gerichtheid op het concrete operationaliseren? Bovendien is het onderzoek dat zich over een relatief lange termijn moet uitstrekken. We kunnen niet verwachten, dat veranderingen in denkwijzen in korte tijd tot stand kunnen worden gebracht. Naar mijn mening is het toch van belang dat dergelijk onderzoek wordt uitgevoerd. Alvorens daartoe wordt overgegaan, verdient het echter aanbeveling eenvoudiger en goedkoper onderzoek te doen naar de eerder vermelde vraagstellingen.

Een belangrijke conclusie die uit dit hoofdstuk kan worden getrokken, is verder nog dat er aptitude-treatment-interacties kunnen worden verwacht. Sommige leerlingen zullen meer in de war worden gebracht door metaforen dan andere. Sommige leerlingen zullen meer behoefte hebben aan concretisering van leerstof dan andere. Sommige leerlingen zullen meer in staat zijn zelf leerstof te structureren dan andere. Sommige leerlingen zullen meer geactiveerd moeten worden om de leerstof te integreren dan andere (enzovoorts). In navolging van Parkhurst (1975) concluderen we dan ook, dat juist waar tegengestelde argumenten bestaan AT-interacties kunnen worden verwacht. De in dit hoofdstuk vermelde potentiële functies van metaforen, kunnen bij het opstellen van hypothesen over ATI's in correspondentie-analyses een belangrijke rol vervullen.

Hoofdstuk 4 Empirisch onderzoek naar vergelijkend leren

In dit hoofdstuk wordt een overzicht gegeven van empirische onderzoeken naar vergelijkend leren.

In 4.1 worden onderwijspsychologische studies besproken, die een zekere ecologische validiteit hadden. Enerzijds is dit onderzoek dat doorgaans tot het terrein van de advance organizers wordt gerekend (zie 3.2.3³ en ~~3.2.3~~), anderzijds betreft het onderzoeken die specifiek op vergelijkend leren waren gericht. In 4.2 worden deze studies geëvalueerd.

4.1 Onderzoek naar vergelijkend leren

Er kunnen zes soorten studies met betrekking tot vergelijkend leren worden onderscheiden: a) onderzoek naar comparative organizers, b) onderzoek naar concrete modellen, analogieën en spellen (games), c) onderzoek naar verbale analogieën in teksten, d) onderzoek naar de invloed van het bestuderen van analoge begrippen op doelbegrippen, e) onderzoek naar "analogical problem solving" en f) onderzoek naar spontaan gebruik van analogieën door lerenden.

4.1.1 Comparative organizers

In 3.2.3.3 werd betoogd dat analogieën en concrete modellen door Ausubel zelf niet tot de advance organizers zouden worden gerekend, omdat deze niet abstracter, algemener en omvattender dan de leertaak zelf zijn. Een categorie van advance organizers, namelijk de comparative organizers (zie 3.2.3.2) kan echter wel tot vergelijkend leren worden gerekend (zie ook 8.1.1). Bij comparative organizers gaat het erom, dat relatief bekende leerstof wordt gerelateerd aan potentieel verwarrende, analoge leerstof. In een comparative organizer worden expliciet - op een abstracte en algemene wijze - de belangrijkste overeenkomsten en verschillen tussen de nieuwe ideeën en de eerder geleerde ideeën benadrukt (Ausubel, 1978). Een analyse van de overzichtsartikelen met betrekking tot advance organizers (Barnes en Clawson, 1975; Hartley en Davis, 1976; Faw en Waller, 1976; Lawton en Wanska, 1977; Mayer, 1979a,

1979b) wees uit dat in slechts vier van de zeer vele studies een comparative organizer werd gebruikt:

Ausubel en Fitzgerald (1961) vergeleken drie condities. Eén groep bestudeerde voorafgaande aan de studietekst over de idealen van het Boedhisme (2500 woorden) een comparative organizer (500 woorden), waarin de belangrijkste verschillen en overeenkomsten tussen het Boedhisme en het Christendom werden benadrukt. Een tweede groep bestudeerde voorafgaande aan de studietekst een expository organizer, waarin de principes van het Boedhisme op een abstracter en algemener manier werden behandeld dan in de tekst zelf. De derde groep was een controlegroep die een historische introductie bestudeerde, voordat de eigenlijke tekst aan bod kwam.

Er deden 155 collegestudenten aan het onderzoek mee. Op een toets die onmiddellijk na het bestuderen van de tekst werd afgenomen, presteerden de proefpersonen in de comparative organizer conditie significant beter dan de proefpersonen in de beide andere condities. Ook op een retentietoets die tien dagen later werd afgenomen, presteerden de proefpersonen in de comparative organizer conditie beter dan de controleproefpersonen. De verschillen tussen de condities waren vooral groot voor de proefpersonen die over het Christendom weinig voor-kennis hadden.

Ook in een onderzoek van Ausubel en Youssef (1963) met 162 collegestudenten, bleek een comparative organizer tot betere prestaties te leiden dan een historische introductie. Ook in dit onderzoek ging de studietekst over het Boedhisme en de comparative organizer over het Christendom. De tekst over het Boedhisme werd uitgebreid met materiaal over "Zen". In een covariantie-analyse met "verbal ability" als covariaat werd een significant hoofdeffect aangetoond. Met name de studenten met een lage "verbal ability" profiteerden van de organizer.

De derde studie naar comparative organizers werd verricht door Fitzgerald en Ausubel (1963) met 264 middelbare scholieren (high school students). Het lesmateriaal bestond uit een tekst van 2900 woorden, die betrekking had op hoe mensen uit

de zuidelijke staten van Amerika de oorzaken van de Amerikaanse burgeroorlog interpreteren. De comparative organizer (450 woorden) bestond uit een overzicht van de belangrijkste verschillen tussen de zuidelijke en de noordelijke standpunten. Zowel op een onmiddellijk na de bestudering van de tekst afgenomen toets als op een retentietoets waren er significante verschillen tussen de proefpersonen die een comparative organizer en proefpersonen die een historische introductie (450 woorden) bestudeerden. Met name waren de verschillen groot bij die scholieren die een "pro-noordelijke" attitude hadden.

Het enige onderzoek naar comparative organizers dat niet door Ausubel zelf werd verricht, is het onderzoek (N=252) van West en Fensham (1976). Deze onderzoekers ontwierpen een organizer die gedeeltelijk "expository" en gedeeltelijk "comparative" was. Het "comparative"-gedeelte van de organizer, waarin scheikundige principes uit de cursus werden vergeleken met natuurkundige weegschaalprincipes, voldoet mijns inziens slechts zeer gedeeltelijk aan de eisen die volgens Ausubel aan een dergelijke organizer moeten worden gesteld. Deze studie hoort dan ook eigenlijk meer in categorie d (4.1.4) thuis. De organizers vergemakkelijkten het leren van die middelbare scholieren (uit de twaalfde klas) die weinig voorkennis hadden. In een tweede experiment bleek dat het effect van de organizers achterwege bleef, wanneer de voorkennis eerst werd aangevuld door middel van een extra programma.

4.1.2 Concrete modellen, analogieën en spelen

Sommige onderzoekers ontwikkelden concrete simulatiespelen om het leren van abstracte, onbekende informatie te vergemakkelijken. Scandura en Wels (1967) presenteerden aan collegestudenten twee simulatiespelen: "Followed-by" en "Play like". Het lesmateriaal was wiskundig van aard en betrof verzamelingenleer en topologie. In het spel "Followed-by" werden binaire operaties en mathematische groepen concreet uitgebeeld door middel van een viertal commando's die gegeven konden worden: naar links, naar rechts, rechtuit en blijf op de plaats. In het spel moesten steeds twee commando's elkaar opvolgen.

Een commando kon dus zijn "naar links gevolgd door naar rechts". Het resultaat van dit commando is "blijf op de plaats". Het commando "gevolgd door" (followed-by) representeerde de binaire operatie. Door het spelen van verschillende combinaties van commando's konden de eigenschappen van wiskundige groepen worden geconcretiseerd.

In het spel "Play like" werden topologische feiten op een simpele wijze begrijpelijk gemaakt, doordat de proefpersonen zich moesten voorstellen een mier te zijn. De proefpersonen die een spel als introductie deden (met name het laatste spel) leverden daarna op een toets significant betere prestaties dan proefpersonen die de historische introductie bestudeerden. Niet alleen waren de prestaties beter, maar ook waren de tijden korter die aan de bestudering van de tekst werden besteed.

Livingston (geciteerd in Barnes en Clawson, 1975) verrichtte een drietal experimenten waarin het effect van een simulatiespel op het leren van economische geografie werd onderzocht. In deze experimenten leverden de experimentele proefpersonen ("eighth-grade en highschool students") geen betere prestaties dan de controleproefpersonen die het simulatiespel niet speelden.

Lesh (geciteerd in Mayer, 1979b) verrichtte een drietal onderzoeken waarin concrete voorbeelden en modellen als advance organizer of als post organizer bij lesmateriaal over meetkunde werden gepresenteerd. In twee studies fungeerden 48 collegestudenten als proefpersonen, in de derde studie vierde en zevende klassers. In alle drie de onderzoeken leverden de proefpersonen die advance organizers kregen betere prestaties dan de proefpersonen die post organizers kregen. In één van de onderzoeken was het verschil tussen deze condities groter bij hiërarchisch dan bij "spiraalsgewijs" georganiseerd lesmateriaal. Het is natuurlijk wel de vraag of de resultaten van Lesh moeten worden toegeschreven aan de concrete voorbeelden, aan de concrete modellen of aan beide.

Een indrukwekkende serie experimenten met betrekking tot concrete modellen is verricht door Mayer (zie voor een overzicht Mayer 1979a, 1979b). De meeste experimenten hadden be-

trekking op het leren programmeren in een computertaal (basic of fortran) door collegestudenten. Een concreet model van een computer, waarvan de verschillende onderdelen werden getekend en beschreven als een postloket, een schrijfmachine, een uitwisbaar schoolbord en een boodschappenlijst, fungeerde als organizer. In de studietekst over fortran werd telkens terugverwezen naar het concrete model.

In de eerste drie experimenten (Mayer, 1975) met in totaal 176 collegestudenten bleek een tekst die was gebaseerd op het concrete model tot hogere far-transfer resultaten te leiden dan een tekst zonder het model, terwijl de controletekst tot hogere near-transfer resultaten leidde. Bovendien werd een aptitude-treatment-interactie gevonden: het concrete model vergemakkelijkte alleen het leren van die proefpersonen die laag scoorden op een intelligentietest.

In de tweede serie van twee experimenten (Mayer, 1976) met 160 collegestudenten, werd het concrete model nog aangevuld met een "activator". Dat is een model van de computer, dat door de lerende kan worden gemanipuleerd. In het eerste experiment bleek dat de activator, wanneer deze vooraf en tijdens het leren kon worden gemanipuleerd, tot hogere transferresultaten leidde dan wanneer deze alleen na de bestudering van de tekst beschikbaar was. Op een kennistoets (near-transfer) waren er geen verschillen, of verschillen in de omgekeerde richting. De manipuleerbare activator bleek belangrijker dan de genoemde afbeelding van de computer in de vorm van een typemachine, boodschappenlijst en dergelijke. De tekst met dit model bleek namelijk niet tot significant betere prestaties op een transfertoets te leiden dan een tekst zonder dit concrete model.

In het tweede experiment vormde dan ook alleen de activator de experimentele variabele (naast een experimentele manipulatie van de volgorde van frames die hier niet relevant is). De proefpersonen die de fortran-cursus, in de vorm van geprogrammeerde instructie, met behulp van de activator bestudeerden, leverden significant betere prestaties dan de proefpersonen die studeerden zonder de activator. Dit verschil was groter voor far-transfer items dan voor near-transfer items.

De derde serie van twee experimenten (Mayer, 1978) had de organisatie van teksten en de compenserende werking van organizers bij slecht georganiseerde teksten tot onderwerp. Mayer gebruikte in het eerste experiment twee versies van geprogrammeerde instructie over het leren programmeren in fortran: één met een logische en één met een willekeurige volgorde van de frames.

De helft van de proefpersonen kreeg weer een beschrijving van de computer in bekende termen (boodschappenlijst, typemachine en dergelijke), de andere helft kreeg deze organizer niet. Als proefpersonen fungeerden 56 psychologiestudenten. Mayer vond geen hoofdeffecten voor de twee experimentele manipulaties: noch het concrete model, noch de tekstorganisatie leidden tot totaaleffecten. De interactie van de beide variabelen was echter wel significant: het concrete model vergemakkelijkte het leren van de willekeurige tekst (41 procent tegenover 31 procent goede antwoorden), maar werkte nadelig op het leren van de logische tekst (36 procent goed met het concrete model tegen 44 procent zonder het model).

De proefpersonen die studeerden met behulp van het concrete model leverden betere prestaties op de far-transfer-toets (58 procent versus 28 procent juiste antwoorden) en de proefpersonen die zonder het model studeerden juist op de near-transfer toets (46 procent versus 39 procent juiste antwoorden). De AT-interactie met de intelligentietest werd nu niet significant bevonden. In het tweede experiment werden soortgelijke resultaten - inclusief de ATI - gevonden als in het eerste experiment, nu echter met een "expository organizer" en met andere leerstof.

In de volgende twee experimenten (Mayer en Bromage, 1980) werd nader ingegaan op de aard van de leerresultaten, die onder verschillende condities worden verkregen. Er werd nu gebruik gemaakt van een gesimplificeerde tekst over de computertaal Basic. De experimentele manipulatie betrof het tijdstip waarop het concrete model (beschreven in een tekst en getekend als een verbinding van een postloket met een schoolbord, een boodschappenlijst en een typemachine) beschikbaar werd gesteld:

voor of na het bestuderen van de basic-tekst. In het eerste experiment werd de 60 psychologiestudenten gevraagd zich na het bestuderen van de teksten bepaalde gedeelten ervan in herinnering te roepen (free recall). In het tweede experiment met 48 psychologiestudenten was de enige verandering dat na één week een retentietoets bij de helft van de proefpersonen werd afgenomen.

De resultaten van beide experimenten waren gelijklopend: het totaal aantal geproduceerde ideeën verschilde niet tussen de groepen, maar in kwalitatief opzicht verschilden de leerresultaten, zowel op korte als op lange termijn. De proefpersonen die het concrete model vooraf en tijdens het bestuderen van de tekst beschikbaar hadden, herinnerden zich meer algemene ideeën en leidden meer verwante ideeën af ("inferences and intrusions"). De proefpersonen die het concrete model pas achteraf beschikbaar hadden, herinnerden zich de technische eenheden beter en voegden meer vage samenvattingen en vage irrelevante verbindingen toe.

Een laatste experimentele studie naar de invloed van concrete modellen is (nog) niet gepubliceerd, maar werd samengevat in Mayer (1979a). Er werd nu een andere computertaal geleerd (een taal voor een data management system) en het concrete model werd ook iets veranderd (vergelijking met een archief en een sorteermant). Ook in dit experiment leverden de proefpersonen (N=40) die het concrete model eerst bestudeerden betere prestaties op een far-transfer-toets dan de proefpersonen die het model niet kregen. Op de near-transfer-toets was er nu geen verschil tussen de condities.

Ook Rigney en Lutz (1976) deden onderzoek naar de invloeden van een concretisering van leerstof door middel van analogieën op het leren van 40 undergraduates. Deze onderzoekers maakten twee versies van een door een computer gestuurde les over electrochemie: één met een getekende uitbreiding en één met een verbale uitbreiding (die inhoudelijk dezelfde informatie verstrekke). De versie met de tekeningen leidde tot significant hogere prestaties op een drietal afhankelijke variabelen. Bovendien werd deze versie significant gunstiger beoor-

deeld op een semantische differentiaalschaal en op een algemene attitudeschaal. De eerste groep van proefpersonen rapporteerde meer mentale voorstellingen met betrekking tot de lesstof dan de proefpersonen van de tweede groep. Rigney en Lutz vonden in hun onderzoek geen ATI's.

4.1.3 Verbale analogieën

Bell en Gagné (1979) deden onderzoek naar de invloed van analogieën in een tekst over elektrische geleiding en over de factoren die daarbij een rol spelen. Zij gebruikten daarvoor de tekst van Royer en Cable (1975) (zie 4.1.4).

In de analogieën werden elektrische geleidingsverschijnselen vergeleken met een danspartij waar mensen tegenelkaar opbotsten en dergelijke (eenzelfde analogie werd door ons ook gebruikt in experiment twee, zonder dat we op de hoogte waren van het onderzoek van Bell en Gagné!). De analogieën waren opgenomen in de tekst voor de experimentele groep en bleven achterwege in de versie voor de controlegroep. Er waren twee afhankelijke variabelen: "free recall" van de tekstgedeelten waar de analogieën betrekking op hadden, en "cued recall" van de overige tekstgedeelten (invultest). Als proefpersonen fungeerden 114 psychologiestudenten.

Er was geen significant verschil tussen de experimentele en de controlegroep in de "free recall" van de tekstgedeelten waar de analogie betrekking op had. Wel was er een significant verschil tussen deze groepen op de "cued recall" test: de controleconditie bracht het er beter af. Verder werden er enkele aptitude-treatment-interacties gevonden: Bij de free-recall test profiteerden alleen de proefpersonen die hoog op verbale en laag op rekenkundige intelligentie scoorden van de analogieën. Bij de "cued-recall" test ondervonden de proefpersonen met een grote rekenkundige intelligentie nadeel van de analogieën en was er een trend in de richting van een gunstig verschil voor de proefpersonen met een lage rekenkundige intelligentie. Bell en Gagné vertaalden deze ATI's in twee interessante processtrategieën die opmerkelijk overeenkomen met de strategieverschillen tussen comprehension- en operation-learners (zie 7.1.2).

Pearson, Raphael, Te Paske en Hyser (1979) lieten zich inspireren door een niet gepubliceerde dissertatie van Arter, die onderzoek deed naar het effect van metaforen op begrip en onthouden van teksten. Arter slaagde er niet in dit bij leerlingen uit de zesde klas (basisschool) aan te tonen. Pearson et al. (op.cit.) meenden dat wel een effect van metaforen zou kunnen worden aangetoond bij gebruik van andere metaforen en andere studieteksten. De metaforen die Arter had gebruikt, waren wellicht bij de leerlingen niet bekend genoeg.

In een eerste experiment, met twintig zesde klassers en twintig undergraduates als proefpersonen, werden tien metaforen verweven in een studietekst over de pyramiden van Egypte. In een pilot-study werd nagegaan of proefpersonen uit dezelfde populatie met de metaforen bekend waren. De tien "targetstructures" waarop de metaforen betrekking hadden, werden significant beter onthouden door de experimentele dan door de controle-groep (42 procent versus 18 procent goede antwoorden bij de zesde klassers). Er was geen interactie tussen leeftijd en de experimentele variabele.

In een tweede experiment gebruikten Pearson et al. een studietekst over watervervuiling. Als proefpersonen fungeerden leerlingen uit de derde klas (basisschool). In dit experiment werden geen verschillen gevonden tussen groepen die een tekst met en een tekst zonder metaforen bestudeerden.

Om na te gaan of het ontbreken van verschillen in dit experiment te wijten was aan de studietekst, die volgens Pearson et al. vertrouwde en bekende stof bevatte voor de leerlingen, werd een derde experiment uitgevoerd. Hieraan deden zowel derde (N=10) als zesde klassers (N=20) deel. Voor de derde klassers was de tekst over watervervuiling de "vertrouwde" tekst en een tekst over "deep sea exploring vessels" de "niet-vertrouwde" tekst. Voor zesde klassers was de pyramidetekst de niet-vertrouwde en een tekst over cowboys de vertrouwde tekst. Inderdaad was er zowel voor derde als voor zesde klassers alleen een verschil tussen de groep die de versie met metaforen en de groep die de versie zonder metaforen bestudeerde bij de teksten waarmee men niet vertrouwd was.

Bielsky (1974) tenslotte, deed onderzoek naar het effect van analogieën op de lange termijn. Uit de beschrijving in de "Dissertation Abstracts" is de aard van de analogieën helaas niet erg duidelijk geworden. Over een periode van negen maanden kregen de experimentele proefpersonen ("ninth-grade" natuurkunde klassen) eens per week een half uur de tijd om 25 analogie-items te beantwoorden. Vervolgens werden deze items klassikaal besproken. Er waren in totaal 155 subjecten, waarvan er 119 de toets invulden. Er werden geen verschillen gevonden tussen proefpersonen uit de experimentele groep en proefpersonen uit de controlegroep (die het normale onderwijs volgden).

4.1.4 De invloed van een relatief eenvoudige tekst op het leren van een analoge abstracte tekst

In de secties 4.1.2 en 4.1.3 werden studies besproken waarin het leren van abstracte teksten werd vergemakkelijkt door middel van vergelijkingen met concrete, dagelijkse verschijnselen in getekende en/of gespeelde vorm. In deze paragraaf komen studies ter sprake waarin, in plaats van vergelijkingen met concrete alledaagse verschijnselen, vergelijkingen met andere, simpeler leerstof worden aangeboden. Die andere leerstof is in een aantal opzichten analoog aan de te leren leerstof, maar is gemakkelijker te leren, of al bekend. Deze leerstof is (naar wordt verondersteld) gemakkelijker te leren doordat zij beter aansluit bij de voorkennis van de leerling (bijvoorbeeld doordat zij al gedeeltelijk bekend is), of doordat zij concreter en beter voorstelbaar is. Meestal moeten de precieze overeenkomsten en verschillen tussen de eenvoudigere leerstof en de doel-leerstof door de lerenden zelf worden gevonden (dit vormt misschien zelfs de kern van het (vergelijkend) leerproces). De verschillen tussen de concrete modellen uit 4.1.2 en de in deze paragraaf te bespreken metaforen zijn overigens klein.

Royer en Cable (1975) maakten de eerst te bestuderen tekst gemakkelijker te begrijpen en te leren door er zoveel mogelijk verwijzingen naar alledaagse verschijnselen (fysische analogieën) in op te nemen (vergelijk de studies in 4.1.2). Zowel van een tekst over warmtegeleiding als over elektrische ge-

leiding werd een concrete versie met analogieën en een abstracte versie zonder analogieën gemaakt. De tekst over warmtegeleiding en die over elektrische geleiding waren in een groot aantal opzichten analoog. Er werd ook nog een controle-introductie (CI) geschreven over mythes en legendes die met geen van de beide natuurkundige teksten iets van doen had. Een combinatie van alle mogelijkheden leverde twaalf (!) verschillende condities op: zes met de warmtegeleidingstekst als tweede tekst en zes met de elektrische geleidingstekst als tweede tekst. De helft van deze tweede teksten was concreet (c) en de andere helft abstract (a). Alle proefpersonen bestudeerden twee teksten. Als tweede teksten waren dat de teksten over warmtegeleiding (sommige in abstracte, andere in concrete vorm) en die over elektrische geleiding (ook in abstracte en concrete vorm). Als eerste teksten fungeerden de andere tekst of de controletekst. Er werden door Royer en Cable twee experimenten uitgevoerd met elk 240 proefpersonen. Alleen de resultaten van het tweede experiment zijn hier relevant.

De proefpersonen die eerst een concreet gemaakte passage bestudeerden en daarna een abstracte, leverden significant betere prestaties dan de proefpersonen die als eerste een abstracte of een controlepassage bestudeerden en als tweede een abstracte. Geen verschillen kwamen er naar voren wanneer een concrete tekst als tweede werd bestudeerd. Royer en Cable interpretererden dit resultaat als volgt: Leren verloopt het best als nieuw materiaal kan worden opgenomen (geassimileerd) in een bestaande cognitieve structuur. Bij de concrete teksten is dit het geval, bij de abstracte niet. Daarom treedt de transfer van de ene tekst naar de andere alleen op als de eerste tekst concreter is dan de tweede, en de tweede tekst bovendien niet kan worden geassimileerd in een bestaande cognitieve structuur.

Een tweede experiment van Royer en Cable (1976) was in eerste instantie bedoeld om een alternatieve verklaring voor de resultaten van hun eerste experiment te weerleggen. Volgens deze verklaring zou het effect niet bewerkt zijn door een context die assimilatie mogelijk maakt, maar door de stijl van de

concrete tekst, die een ander (mathemageen) aandachtsproces bij de bestudering van de tweede tekst zou hebben geïnduceerd. Om deze verklaring te kunnen weerleggen creëerden zij een conditie waarin dit "stijleffect" niet kon optreden: een tekst die abstract van stijl was, maar toch de fysische analogieën bevatte. Daarnaast werd ook nog als conditie ingevoerd, dat een abstracte tekst aangevuld met illustraties als eerste tekst fungeerde.

De resultaten die werden verkregen in een onderzoek met 80 collegestudenten, waren in overeenstemming met de oorspronkelijke verklaring van Royer en Cable (1975). De proefpersonen die als eerste de abstracte tekst met de analogieën, en de proefpersonen die de abstracte tekst met illustraties bestudeerden, leverden even goede prestaties als de proefpersonen die de concrete tekst als eerste bestudeerden. Onder deze drie condities werden significant betere resultaten behaald dan wanneer eerst een controletekst of een abstracte tekst werd bestudeerd. Conclusie: Transfer van een tekst naar een tweede analoge tekst treedt op als de eerste tekst gemakkelijker te leren is dan de tweede (dankzij illustraties, analogieën of concreetheid), en als assimilatieve context kan fungeren voor de tweede tekst.

Ook Devine-Hawkins (geciteerd in Davidson, 1976) verrichtte onderzoek naar het effect dat het bestuderen van een tekst had op het bestuderen van een tweede tekst (een dubbele hiërarchie). De doeltekst ging over begrippen als biosfeer, bioom ecosysteem en de eenvoudiger te begrijpen tekst ging over begrippen als activiteit, werk, beroep (enzovoorts). De teksten waren extreem kort (tien regels).

De experimentele proefpersonen bestudeerden gedurende vijf minuten de tekst over activiteit en werk, een eerste controlegroep bestudeerde gedurende deze tijd een tekst over een ziekenhuis waarin alle woorden uit de experimentele tekst voorkwamen, echter zonder dat de relaties tussen de begrippen aan bod kwamen. Een tweede controlegroep mocht gedurende deze vijf minuten "voor zich zelf werken".

Vervolgens werd de tekst over ecologie gedurende twee en een halve minuut bestudeerd, waarna de proefpersonen alles wat

ze nog wisten van de ecologietekst moesten opschrijven. Daarna werd de ecologietekst nog eens twee en een halve minuut bestudeerd. Tot slot volgde nog een "free recall"-meting, alsmede een woordassociatie-test en "similarity ratings". Proefpersonen waren 96 leerlingen uit de zesde klas (basisschool).

Er werden geen significante verschillen gevonden tussen de drie groepen. In het algemeen waren de prestaties bovendien erg laag. Davidson (1976) vermeldde twee mogelijke oorzaken voor deze teleurstellende resultaten: a) het feit dat de relaties tussen de twee hiërarchieën van begrippen onvoldoende expliciet waren en b) de mogelijkheid dat het kunnen werken met zo'n dubbele hiërarchie meer cognitieve ontwikkeling vereist dan de proefpersonen hadden. Andere - mijns inziens meer voor de hand liggende - verklaringen zijn: c) de studieteksten waren te kort; d) de studieteksten waren te moeilijk voor de proefpersonen; e) de specifieke, dubbele hiërarchie was niet erg verhelderend.

4.1.5 De rol van analogieën bij het oplossen van problemen

De onderzoeken die tot nu werden behandeld, gingen alle over de invloed van (verschillende soorten) analogieën op het begrijpen en onthouden van teksten. Twee groepen onderzoekers (Reed et al., 1974 en Gick en Holyoak, 1980) deden onderzoek naar de invloed die het oplossen van een probleem (of kennis van de oplossing) hebben op het oplossen van een analoog probleem.

Reed et al. (1974) onderzochten hoe het oplossen van twee - in de cognitieve psychologie veel gebruikte - problemen (het missionaris-kannibalen probleem (MK) en het jaloerse echtgenoten probleem (JE)) het vinden van de oplossing van het andere probleem beïnvloedde. In hun eerste experiment werden de formeel analoge problemen (MK en JE) in twee volgordes (MK-JE en JE-MK) opgelost door psychologiestudenten (N=97). Er waren geen verschillen wat betreft de totale aantallen zetten en de oplossingstijden tussen de als eerste en de als tweede opgeloste problemen. Alleen het aantal foute zetten was bij de eerste problemen significant groter dan bij de tweede.

In het tweede experiment gingen Reed et al. na of er meer transfer zou optreden als eenzelfde probleem twee maal achter elkaar werd opgelost. Dit bleek inderdaad het geval.

In het derde experiment (N=75) kregen de proefpersonen (weer psychologiestudenten) instructies over de formele gelijkheid van de problemen. Nu bleek er wel transfer op te treden van het ene probleem naar het tweede, echter uitsluitend in de volgorde JE-MK en niet in de volgorde MK-JE.

Gick en Holyoak (1980) construeerden zogenaamde "verhaalproblemen", die het oplossen van het doelprobleem moesten vergemakkelijken. Als probleem diende Duncker's bekende stralingsprobleem (Hoe kun je een kwalijk gezwel bestralen zonder het andere weefsel te beschadigen?). Analoge problemen gingen bijvoorbeeld over een land met een fort in het midden dat moest worden veroverd door een vreemde mogendheid. In de teksten waarin de analoge problemen werden beschreven, gaven Gick en Holyoak ook verschillende oplossingen. Doordat bij verschillende condities verschillende oplossingen werden gegeven, kon worden nagegaan in welke mate de oplossingsprotocollen van de problemen waren beïnvloed door de verhaalproblemen. In vijf experimenten met psychologiestudenten werd aangetoond, dat:

- 1) Transfer optrad van de verhaalproblemen naar het doelprobleem en dat de soort van oplossingen ten dele werd bepaald door de soort van oplossing die vooraf werd beschreven; 2) dat een zoveel mogelijk analoog probleem tot meer transfer leidde dan een minder analoog probleem, ondanks identieke oplossingen in de verhaalproblemen; 3) dat ook wanneer de verhaalproblemen door de proefpersonen zelf werden opgelost transfer optrad;
- 4) dat de transfer echter in de volgende gevallen minder werd als:
 - a) Geen aanwijzingen werden gegeven betreffende het gebruik van de verhaalproblemen bij het oplossen; b) de verhaalproblemen te midden van afleidende verhalen werden geplaatst zonder aanwijzing ze te gebruiken; c) de verhaalproblemen in een andere context werden gememoriseerd zonder dat een aanwijzing werd gegeven ze te gebruiken; d) de verhaalproblemen tussen twee pogingen om het doelprobleem op te lossen werden gepresenteerd.

4.1.6 Spontaan gebruik van analogieën

Bij het oplossen van problemen vergelijken mensen waarschijnlijk vaak het probleem dat ze willen oplossen met eerder opgeloste problemen. Davidson (1976) vermeldde bijvoorbeeld een aantal voorbeelden van de rol die analogieën gespeeld hadden bij grote ontdekkingen. Gordon ontwikkelde zelfs een creativiteitstraining rondom analogierelaties (zie Davis, 1976).

Maar ook bij het leren en begrijpen van nieuwe informatie (bijvoorbeeld uit studieteksten) kunnen spontane analogieën een belangrijke rol spelen. Bijvoorbeeld: Een ijskastmonteur die zich laat omscholen tot verwarmingsmonteur zal, als hij verstandig is, vele principes van de verwarmingspomp vergelijken met de hem bekende principes van de ijskast; en iemand die veel weet van brommers, zal de onderdelen en functies van een automotor vergelijken met die van een brommer.

In informele gesprekken over leerprocessen kan men regelmatig horen dat sommige mensen veelvuldig nieuwe kennis proberen te vergelijken met wat ook maar ter vergelijking in aanmerking komt. Onderzoek naar het spontaan trekken van vergelijkingen is er echter nauwelijks geweest.

Pask is één van de weinigen die zich ermee bezig houdt (zie 3.2.6 en 7.1.2). Zijn onderzoek (zie Pask 1976b, 1976c; Entwistle, 1978) heeft duidelijk gemaakt dat spontane analogieën een belangrijke rol spelen bij het begrijpen van informatie en dat er grote individuele verschillen zijn in de mate waarin mensen geneigd zijn spontaan vergelijkingen te trekken (zie verder 7.1.2).

4.1.7 Overig onderzoek naar metaforen en analogieën

In 4.1 werden slechts die onderzoeken naar metaforen en analogieën behandeld, die betrekking hadden op onderwijsleersituaties en die enigszins ecologisch valide waren. Er is ook minder ecologisch valide onderzoek naar de rol van metaforen en analogieën in onderwijsleersituaties verricht. Davidson (1976) heeft daarvan een overzicht gegeven. Over het algemeen kan men stellen, dat ook bij het leren van niet zo rechtstreeks voor het onderwijs relevant materiaal (als woordparen, setrela-

ties en proza) metaforen en analogieën een significante bijdrage kunnen leveren.

Voorts is er ontwikkelingspsychologisch onderzoek naar het begrijpen van metaforen en analogieën door kinderen verricht. Deze studies werden samengevat door Billow (1977) en door Ortony et al. (1978). Sommigen (onder andere Billow) legden de nadruk op moeilijkheden die kinderen - door onvoldoende cognitieve ontwikkeling - zouden ondervinden bij het begrijpen van metaforen, anderen legden de nadruk op studies die lieten zien dat veel jongere kinderen al in staat waren metaforen te begrijpen (zie Ortony et al., 1978; Pearson et al., 1979) en op de methodologische problemen in voorgaand onderzoek (Ortony et al., 1978).

Belangwekkend onderzoek naar processen die ten grondslag liggen aan het begrijpen van analogieën van de vorm $A:B=C:D$ door volwassenen werd geïnitieerd door Sternberg (1977a, 1977b).

4.2 Evaluatie van het tot nu toe verrichte onderzoek naar vergelijkend leren

De in 4.1 samengevatte experimenten vertonen - zeker wanneer men de resultaten ervan vergelijkt met bijvoorbeeld de advance organizer studies in het algemeen (Barnes en Clawson, 1975) - een redelijk consistent beeld. Deze resultaten kunnen als volgt worden samengevat.

- Comparative organizers leiden tot betere prestaties bij collegestudenten en middelbare scholieren dan historische introducties (de onderzoekingen van Ausubel).
- Simulatiespelen leiden bij wiskunde tot hogere studieprestaties bij collegestudenten (Scandura en Wels) dan een historische introductie, maar niet bij geografie wanneer de experimentele groep werd vergeleken met een controlegroep die geen introductie kreeg ("eighth-grade" en "high school" studenten; Livingston).
- Concrete modellen leiden bij collegestudenten en bij vierde en zevende klassers soms tot hogere resultaten, soms alleen tot hogere far-transfer en tot lagere near-transfer resultaten (de onderzoekingen van Lesh, Mayer en Rigney en Lutz).

- In teksten verweven verbale analogieën leiden soms tot hogere prestaties bij basisschool leerlingen (Pearson et al., experimenten een en drie), soms trad dit effect niet op (Pearson et al., experiment twee; Arter). Het trad ook niet op bij oudere proefpersonen (Bell en Gagné; Bielinsky).
- Het vergelijken van twee formeel vergelijkbare begrippensystemen leidde bij collegestudenten en twaalfde klassers tot transfer mits aan de volgende voorwaarden was voldaan: a) de eerst gepresenteerde serie begrippen moest een context voor assimilatie bieden voor de tweede serie; b) de eerste serie begrippen moest gemakkelijker kunnen worden geleerd dan de tweede serie (door concrete analogieën, illustraties of abstracte analogieën) (Royer en Cable; Cable en Royer; West en Fensham). De transfer bleef achterwege bij zesde klas leerlingen (Devine-Hawkins).
- Onder bepaalde omstandigheden trad bij collegestudenten transfer op van het oplossen of bestuderen van een probleem naar een analoog probleem (Reed et al.; Gick en Holyoak).
- Soms bleek het effect van analogieën afhankelijk van lesmateriaalfactoren (Lesh; Mayer; Pearson et al.) en specifiek voor bepaalde afhankelijke variabelen (Mayer).
- Er waren diverse aanwijzingen in de richting van ATI-effecten (Ausubel en Youssef; Ausubel en Fitzgerald; Mayer; Bell en Gagné).

Hoewel de resultaten alles bijeen een redelijk consistent beeld opleverden, moeten de conclusies met de nodige voorzichtigheid worden gezien, omdat de meeste op een beperkt aantal studies zijn gebaseerd. Er is ten aanzien van al deze conclusies nog nader onderzoek vereist. Niet alleen is het van belang dat de besproken onderzoeken worden gerepliceerd, maar ook onderzoek waarin generalisatie naar andere groepen proefpersonen en ander lesmateriaal wordt bestudeerd, kan belangrijke informatie opleveren. Werken comparative organizers en concrete modellen bijvoorbeeld ook bij basisschool-leerlingen? Leiden simulatiespelen wel tot resultaten bij middelbare scholieren bij wiskundelesmateriaal? Werken concrete modellen ook bij andere lesstof als meetkunde en computerprogrammeren? (enzovoorts).

De conclusies moeten ook met voorzichtigheid worden bekeken vanwege het feit dat in ongeveer de helft van de studies controlemateriaal met een geringe ecologische validiteit (zie hoofdstuk 1) werd gehanteerd. In plaats van controlegroepen die de analogieën niet ter beschikking hadden, werden vaak controlegroepen gebruikt die een historische introductie bestudeerden of het analoge materiaal na de lezing van de tekst bestudeerden.

Faw en Waller (1976) hebben erop gewezen, dat een controlegroep met historische introductie ernstige generalisatieproblemen met zich mee brengt. Men mag in het geval dat men een significant verschil heeft aangetoond tussen een experimentele introductie en een historische introductie, op zijn hoogst concluderen dat de experimentele introductie beter is dan een slechte andere introductie. Voor zover deze historische introductie in de ecologie niet voorkomt of niet belangrijk wordt gevonden, is dit een ernstige beperking van de generalisatiemogelijkheden (zie verder hoofdstuk 6).

Ook voor de zogenaamde "post organizers" geldt eenzelfde bezwaar. Deze komen in de ecologie zeker niet voor en de generalisatiemogelijkheden van onderzoeken waarin een post organizer conditie als controleconditie fungeerde, zijn dan ook beperkt. Voor theoretische doeleinden kunnen post organizer condities natuurlijk van groot belang zijn (zie bijvoorbeeld Mayer, 1979a, 1979b). In volgende onderzoeken zijn echter (ook) andere controlegroepen gewenst (zie hoofdstuk 6).

Een andere beperking van de besproken onderzoeken betreft de studietijden. Met uitzondering van het onderzoek van Scandura en Wels, waarin het simulatiespel tot kortere tekstbestuderingstijden leidde, werd in geen der artikelen aandacht besteed aan studietijden. De tijd die aan de metaforen, organizers of analogieën werd besteed, werd ofwel genegeerd ofwel kunstmatig geneutraliseerd door de controlegroep een historische introductie of een post organizer te geven. Ook de eventuele tijdsbesparing of tijdsverlenging die als gevolg van de metaforen kan optreden, werd buiten beschouwing gelaten. Er is dan ook behoefte aan onderzoek naar de tijd die metaforen, organi-

zers, concrete modellen en dergelijke kosten (zie ook de argumenten tegen metaforen in 3.3), en naar de efficiëntie van vergelijkend leren (zie verder hoofdstuk 6).

De onderzoeken van Mayer hebben duidelijk gemaakt, dat het bij onderzoek naar vergelijkend leren van groot belang is verschillende soorten afhankelijke variabelen te hanteren. Soms waren de resultaten van far-transfer toetsen en begripsmetingen immers tegengesteld aan de resultaten van meer conventionele (op kennis gerichte) studietoetsen. Een gebrek van het merendeel van de samengevatte studies was dat slechts korte termijn metingen werden verricht. Zoals reeds in hoofdstuk 1 werd betoogd, zijn echter vooral ook retentiemetingen op langere termijn van belang.

De verschillende studies werden in 4.1 ingedeeld in zes categorieën. De onderscheidingen tussen deze categorieën zijn niet altijd even scherp. Doordat de beschrijvingen van de metaforen in sommige artikelen erg vaag waren, zijn sommige onderzoeken wellicht in een verkeerde categorie terecht gekomen (Rigney en Lutz?; Bielinsky?). Toch is het mijns inziens van belang verschillende soorten van vergelijkend leren van elkaar te onderscheiden (zie ook 3.1).

Kenmerkend voor de eerste categorie vergelijkingen (comparative organizers) is dat vooraf een vergelijking wordt getrokken met potentieel verwarrende eerder geleerde leerstof. Bij de simulatiespelen en de concrete modellen vindt vooraf een concretisering van de lesstof plaats in een spel of concrete tekening met beschrijving, waarnaar in de lesstof voortdurend wordt verwezen. De analogierelaties worden in de lesstof of vooraf geëxpliciteerd. Kenmerkend voor de verbale analogieën (in 4.1.3) is dat de analogieën niet (noodzakelijk) concreter zijn dan de lesstof maar in ieder geval wel bekender. Bovendien worden deze analogieën in de lesstof op strategische plaatsen opgenomen. Het onderzoek in de vierde categorie (4.1.4) wordt gekenmerkt doordat twee systemen van begrippen (een dubbele hiërarchie) naast elkaar worden geplaatst (de ene wordt dus voor de andere geleerd), doordat de relaties tussen de twee systemen (de "ground" en de "tension" van de metafoor)

door de lerende zelf moeten worden gevonden, en door het feit dat het eerste systeem van begrippen eenvoudiger kan worden geleerd dan het tweede. Hierdoor kan het een context vormen voor assimilatie van het tweede systeem van begrippen.

Kenmerkend voor de vijfde categorie onderzoeken was dat het om het oplossen van problemen als afhankelijke variabele ging. In de zesde categorie onderzoeken maakten de lerenden zelf spontaan gebruik van vergelijkingen.

In deel III van deze dissertatie worden zes experimenten gerapporteerd. Vier daarvan vormen een nog niet eerder onderzochte variant van de vierde categorie van onderzoeken: vooraf wordt een tekst bestudeerd, waarin begrippen (en relaties daartussen) worden behandeld die formeel analoog zijn aan de te bestuderen tekst. Net als in de onderzoeken van Royer en Cable zijn deze teksten gemakkelijker te leren dan de doelttekst. Echter niet doordat er analogieën of illustraties in zijn opgenomen, maar doordat deze teksten door hun geaardheid (ze zijn concreet) gemakkelijker te begrijpen zijn. Het zijn derhalve niet voor de hand liggende vergelijkingen met samenhangende, soms bekende oude, soms concrete verschijnselen die vooraf worden gegeven. De mate van explicitering van de relaties tussen de analoge tekst en de doelttekst varieert tussen de experimenten. Experiment vijf past in de derde categorie (verbale analogieën in de tekst opgenomen) en experiment zes sluit het meest aan bij de tweede categorie (concrete modellen).

II METHODOLOGISCH DEEL

Hoofdstuk 5 Onderzoekstechnische problemen en het verwerken van de gegevens van Aptitude-Treatment-Interaction-Research

De belangrijkste problemen bij onderzoek naar tekstbestudering - met name bij een van de meest onderzochte deelgebieden daarbinnen (de pre-instructievariabelen) - hebben betrekking op studietijden en de controle en manipulatie daarvan. Daaraan wordt dan ook het volgende hoofdstuk (6) gewijd. In dit hoofdstuk (5) worden de voornaamste onderzoekstechnische problemen in dergelijk onderzoek geanalyseerd, waarbij ook weer de nadruk ligt op pre-instructie onderzoek, omdat dit het dichtst aansluit bij het onderzoek naar metaforen. In 5.2.1 wordt een serie aanbevelingen opgesomd, die uit diverse overzichtsartikelen gedestilleerd werd en in 5.2.2 wordt de rol van de voortoets in het pretest-posttest-control-group(s)-design belicht. In het laatste deel (5.3) van dit hoofdstuk wordt aandacht besteed aan het verwerken van ATI-gegevens met behulp van regressie-analyse en aan de beslissingen die een onderzoeker daarbij moet nemen.

5.1 Inleiding

Onlangs heeft Van Hout-Wolters (1980) een overzicht gemaakt van het tekstbestuderingsonderzoek. Zij onderscheidde op basis van een 50-tal artikelen twee soorten van onderzoek: onderzoek naar het effect van door de docent of auteur in een tekst aangebrachte wijzigingen, zogenaamde experimenter provided activities (EPA's), en onderzoek naar het effect op de leerresultaten van door de lerende uitgevoerde tekstbewerkingen en bestuderingsactiviteiten, zogenaamde subject generated activities (SGA's). Binnen elke soort onderscheidde zij nog vier categorieën van onderzoek (zie figuur 5.1):

- I Toevoegen van verduidelijkende gedeelten aan de tekst;
- II Toevoegen van sturende en instruerende gedeelten aan de tekst (of zichzelf geven van instructies);

III Wijzigen van de tekst zelf, met name van structurering en taalgebruik;

IV Wijzigen van de vormgeving van de tekst.

In figuur 5.1 zijn van alle acht categorieën van onderzoek de voorbeelden van Van Hout-Wolters (op.cit.) overgenomen. Onderzoek naar metaforen komt in het overzicht niet voor en dus was het nodig zelf te bepalen in welke categorie(ën) onderzoek naar metaforen zou passen.

Figuur 5.1 Een overzicht van de acht categorieën van tekstbestuderingsonderzoek volgens Van Hout-Wolters

Categorie van onderzoek		Voorbeelden van maatregelen
I Toevoegen van verduidelijkende gedeelten aan de tekst	a) E P A	Advance organizers, schema's, samenvattingen, plaatjes toegevoegd door docent/auteur
	b) S G A	Door subject zelf maken van aantekeningen, overzichten, schema's, samenvattingen; zelf visualiseren van de tekstinhoud
II Toevoegen van sturende en instruerende gedeelten aan de tekst of zichzelf geven van instructies	a) E P A	Voortoetsen, doelen, vragen en dergelijke (toegevoegd door docent/auteur)
	b) S G A	(Door subject zelf) stellen van doelen; maken en beantwoorden van vragen; navertellen van de tekstinhoud
III Wijzigen van de tekst zelf (met name van structurering en taalgebruik)	a) E P A	Wijzigen van de opbouw van de tekst door de docent/auteur
	b) S G A	Wijzigen van de opbouw van de tekst door subject (bijvoorbeeld tekstvervangend uittreksel); wijzigen van de volgorde van bestudering van tekstgedeelten
IV Wijzigen van de vormgeving van de tekst	a) E P A	Aanbrengen van onderverdelingen, nummering, dik of cursief drukken van gedeelten, kleuren, onderstrepen (door docent/auteur)
	b) S G A	Idem door lerende zelf

Zoals uit hoofdstuk 4 bleek, werd zowel EPA als SGA-onderzoek naar metaforen verricht; het SGA-onderzoek (bijvoorbeeld Pask, 1976a) was echter zeldzamer dan het EPA-onderzoek. Het onderzoek dat in deze dissertatie in deel III wordt gerapporteerd, is uitsluitend EPA-onderzoek: De onderzoeker voegde metaforen toe aan studieteksten. In hoofdstuk 3 werd betoogd, dat metaforen zowel een verduidelijkende als een sturende functie kunnen vervullen. Ons onderzoek past derhalve zowel in de categorieën Ia als IIa. Het onderzoek dat past in de categorieën Ia en IIa van Van Hout-Wolters wordt in de literatuur tegenwoordig onder één noemer gebracht: Onderzoek naar "pre-instructional strategies" (zie Hartley en Davies, 1976).

Omdat relatief weinig onderzoek is verricht naar de rol van metaforen (zie 4.1) ontbreekt op dit terrein een specifieke researchtraditie. Om toch inzicht te krijgen in potentiële methodologische problemen, werd een analyse gemaakt van aanbevelingen die voor het onderzoeksterrein van de pre-instructievariabelen werden geformuleerd. De resultaten van deze analyse worden weergegeven in 5.2.1. In 5.2.2 wordt aandacht besteed aan de opzet van een onderzoek naar pre-instructievariabelen met speciale nadruk op de rol van de voortoets daarbij.

5.2 Methodologische aanbevelingen en de rol van de voortoets

5.2.1 Methodologische aanbevelingen met betrekking tot pre-instructievariabelen

In overzichten van onderzoek naar pre-instructievariabelen komen impliciet en expliciet allerlei richtlijnen aan de orde die onderzoekers zich ter harte kunnen nemen. Veel aanbevelingen komen overeen met de eerder (zie 1.3) geformuleerde uitgangspunten voor onderwijspsychologisch onderzoek. Zo pleitten veel reviewers voor grotere ecologische validiteit van het lesmateriaal (Faw en Waller, 1976; Van Hout-Wolters, 1980), van de leersituatie (Van Hout-Wolters, 1980), van de soort van leerlingen (Barnes en Clawson, 1975) en van het doel waarmee wordt gelezen (Van Hout-Wolters, 1980). Evenzo pleitte men voor retentiemetingen die na een langere periode worden afge-

nomen (Barnes en Clawson, 1976; Faw en Waller, 1976; Van Hout-Wolters, 1980). Andere aanbevelingen zijn:

- a) Gebruik niet alleen meerkeuze-kennistoetsen, maar ook open toetsen (Barnes en Clawson, 1975), en toetsen die betrekking hebben op hogere vormen van leren (bijvoorbeeld verticale en laterale transfer, probleem oplossen, relaties tussen begrippen) (Lawton en Wanska, 1977; Andre, 1979).
- b) Besteed meer aandacht aan individuele verschillen tussen leerlingen (Andre, 1979; Barnes en Clawson, 1975; Faw en Waller, 1976; Van Hout-Wolters, 1980).
- c) Ga ook na wat het effect is van het pre-instructiemateriaal bij langdurig gebruik (Faw en Waller, 1976). De meeste experimenten besloegen slechts een korte tijdspanne en het is maar de vraag of resultaten die onder dergelijke omstandigheden zijn gevonden generaliseerbaar zijn naar situaties waarin langduriger gebruik wordt gemaakt van extra pre-instructiemateriaal.
- d) Gebruik ook mondeling of visueel materiaal. Barnes en Clawson (1975) constateerden dat het merendeel van de onderzoekers gebruik maakte van geschreven pre-instructiemateriaal. Zij raadden aan meer onderzoek te doen naar de effecten van mondeling gepresenteerd pre-instructiemateriaal of van schema's en tekeningen, en dergelijke.
- e) Laat proefpersonen die te hoog scoren op de voortoets buiten beschouwing. Het is volgens Lawton en Wanska (1977) alleen mogelijk zinvolle natoetsgegevens te verkrijgen als "subjects that pass the pretest" uit de steekproef worden verwijderd.
- f) Besteed meer aandacht aan de constructie van het pre-instructiemateriaal. Lawton en Wanska (1977) en Mayer (1979a) beschreven bijvoorbeeld hoe advance organizers moeten worden geconstrueerd, en Barnes en Clawson (1975) pleitten voor operationeel gedefinieerde organizers (beoordeling door een panel van deskundigen).
- g) Houd meer rekening met de condities waaronder pre-instructiemateriaal wel en niet kan functioneren. Mayer (1979a) toonde bijvoorbeeld aan dat advance organizers alleen kunnen functioneren als:

1. de leerling de organizer heeft begrepen en onthouden;
2. de leerling zelf geen relevante "subsumers" beschikbaar heeft en
3. de leerling de nieuwe informatie actief integreert (zie ook hoofdstuk 3).

Ook volgens Lawton en Wanska (1977) bepalen soortgelijke condities mede het effect van een organizer. Hartley en Davies (1976) beschreven onder welke omstandigheden voortoetsen, gedragsdoelstellingen, samenvattingen en advance organizers waarschijnlijk een functie zouden kunnen hebben. Ook Salomon en Clark (1977) benadrukten - in een andere context - dat meer aandacht voor functies van instructievariabelen is geboden.

- h) Controleer of de onafhankelijke variabele functioneert zoals bedoeld. Barnes en Clawson (1975) pleitten er voor dat getoetst wordt of de advance organizer wordt beheerst (zie ook Mayer, 1979a), en Andre (1979) wees erop dat het van belang is na te gaan of vragen bij een tekst niet te moeilijk zijn (door een controlegroep bijvoorbeeld de vragen met de tekst erbij te laten beantwoorden). In een eerdere publicatie (Simons, 1978) werd voorgesteld bij onderzoek naar metaforen in de onderzoeksopzet toetsen op te nemen, waarmee op de volgende drie punten controle wordt uitgeoefend:
1. Begrijpen de leerlingen de metaforen?
 2. Herinneren ze zich de metaforen na afloop van de bestudering van de tekst?
 3. Kunnen ze relaties leggen tussen de studietekst en de metaforen?
- i) Rapporteer uitvoerig met betrekking tot het materiaal en de gevolgde procedure, zodat replicatie mogelijk is (Andre, 1979). Cronbach en Snow (1977) pleitten om dezelfde reden voor uitvoerige rapportage van descriptieve gegevens van ATI-experimenten.
- j) Zorg voor gelijke hoeveelheden informatie in de verschillende condities (Mayer, 1979a). "A major shortcoming of many advance organizer studies cited by Barnes en Clawson is that they failed to control for the amount of information

presented to subjects in different groups. For example, in most studies a group given an advance organizer is compared to a group not given an organizer or to a group that is given a control organizer. Since subjects do not receive identical information it is possible that any subsequent differences in performance are due to the content included in the organizers." (Mayer, 1979a, p. 372).

Een soortgelijk argument werd naar voren gebracht door Jaspers (1977). In studies waarin effecten van een organizer worden vergeleken met die van een neutrale, biografische of historische introductie, verschillen de hoeveelheden informatie. Over de biografische of historische introductie worden weliswaar geen vragen gesteld, maar dat weten de leerlingen niet.

In totaal moeten zij derhalve meer informatie verwerken dan de leerlingen die alleen de extra informatie uit de organizer moeten bestuderen die de inhoud van de studietekst overlapt. Ook al tellen de organizer en de neutrale introductie evenveel woorden, de hoeveelheden informatie kunnen verschillen. Gelijke hoeveelheden informatie kunnen volgens Mayer (op.cit.) worden bereikt door een groep die met een advance organizer werkt te vergelijken met een groep die met een post organizer werkt.

- k) Ga na waar een pre-instructievariabele vooral invloed uitoefent: Tijdens de codering van de informatie of ten tijde van de "retrieval" (Mayer, 1979a). In de cognitieve psychologie wordt, waar mogelijk, onderscheid gemaakt tussen de verschillende stadia in een leerproces: Aanleren (acquisition), opslag (storage) en retentie (retention) (zie Crowder, 1976; Wickelgren, 1977). Veel leerpsychologisch onderzoek heeft tot doel uit te zoeken op welk onderdeel van het leerproces bepaalde onafhankelijke variabelen hun invloed uitoefenen.

Underwood (1964) toonde bijvoorbeeld aan, dat bij het leren van gepaarde associaties de mate van "meaningfulness" van de associaties vooral invloed had op het gemak waarmee de associaties werden aangeleerd. Hij vond echter geen ver-

schil tussen de retentie van betekenisvolle en betekenisloze associaties als ze tot eenzelfde criterium werden geleerd (dus geen verschillen in het retentiestadium).

Eenzelfde onderscheid tussen aanleren en retentie of tussen "encoding" en "retrieval" kan ook worden gemaakt bij het bestuderen van teksten. Mayer (op.cit.) verwijst in dat verband naar studies van Dooling et al., die aantoonen dat het presenteren van een titel voorafgaande aan een verhaal wel effect had maar achteraf niet. De titel vergemakkelijkte dus de encoding en niet de retrieval.

Een onderscheid dat enigszins analoog is aan het onderscheid tussen aanleren en onthouden, is dat tussen begrijpen en onthouden. Begrijpen kan men bijvoorbeeld operationaliseren als dat wat een leerling op een begripstoets (een volgens de richtlijnen van Anderson (1972) geconstrueerde toets) presteert, nadat hij een tekst één keer heeft doorgelezen. Onthouden kan men op de gebruikelijke manier operationaliseren als dat wat een leerling na verloop van een bepaalde tijd nog van een tekst weet.

- l) Vergelijk ook de verschillende soorten (pre)-instructievariabelen met elkaar (Van Hout-Wolters, 1980; Hartley en Davies, 1976). Veel onderzoeken werden gericht op effecten van één van de pre-instructievariabelen (organizers, voortoetsen, doelstellingen, samenvattingen, metaforen). Het is echter van belang te weten te komen onder welke omstandigheden en voor welke leerlingen welke pre-instructievariabele(n) moeten worden toegepast.
- m) Barnes en Clawson (1975) besloten hun overzicht met de volgende serie aanbevelingen: "Carefully meet all of the required conditions for experimental research including random assignment of subjects to treatment groups, maintain independence of subjects, use the appropriate statistical tests, test for mastery of the organizer prior to presentation of the material to be learned, be certain that the test of the advance organizer and all tests of the material to be learned are different from one another, include a retention measure if possible, follow established procedures

for test construction and maintain high reliability and validity."

5.2.2 De betekenis van de voortoets in het "pretest-posttest-control-group-design".

Hoewel volgens Van Hout-Wolters (1980) weinig onderzoekers een voortoets afnamen in hun onderzoek, zijn er toch goede redenen om het wel te doen. Wanneer het niet mogelijk is de proefpersonen aselekt toe te wijzen aan de groepen die onder verschillende condities zullen werken, is het noodzakelijk een voortoets af te nemen om zo te kunnen controleren of eventuele verschillen (ook) niet al voor het onderwijs bestonden. Maar ook in het geval van aselekte toewijzing aan condities zijn er goede (andere) redenen om een voortoets af te nemen:

- a) Om na te gaan of er ondanks de aselekte toewijzing toch geen verschillen met betrekking tot de voortoetsprestaties zijn ontstaan.
- b) Om na te gaan of er wel iets geleerd is door de leerlingen.
- c) Om na te gaan of er wel voorkennis aanwezig is.
- d) Om (eventueel) proefpersonen met te veel voorkennis uit te selecteren en
- e) om het onderscheidingsvermogen van de statistische toets te vergroten door verkleining van de foutenvariantie (Cohen en Cohen, 1975, p. 378).

In beide gevallen zijn er derhalve goede redenen om een voortoets af te nemen. Wat er vervolgens met die voortoetsgegevens moet worden gedaan is een andere zaak, die ook nauw samenhangt met een algemeen probleem dat met voortoetsen verbonden is. Het betreft de vraag wat een voortoets meet als er geen of weinig specifieke voorkennis is.

Dat is geen denkbeeldig probleem: Men ontwerpt onderwijs vaak juist voor die leerlingen die nog geen of weinig specifieke voorkennis bezitten. In al onze experimenten (zie deel III) werd gekozen voor leerstof die bij de leerlingen relatief onbekend is. Neemt men dan een voortoets af die gelijk is aan, of gelijkenis vertoont met de natoets, dan kan niet anders worden verwacht dan dat de meeste leerlingen op die voortoets op

basis van het toeval zullen scoren. De voortoets meet dan hoe succesvol iemand is in het raden van het juiste antwoord. Een gevolg daarvan is dat - in bijna al onze in deel III gerapporteerde onderzoeken - de homogeniteit van de scores op de voortoets zeer laag is (meestal rond de .20), de scores niet hoger zijn dan op basis van toeval verwacht kan worden en ook de correlaties van de voortoets met de na- en retentietoets laag zijn te noemen. (Dat deze resultaten niet te wijten zijn aan een slechte toetsconstructie, blijkt uit de hoge homogeniteitsindices van de na- en retentietoetsen en acceptabele correlaties tussen de beide toetsen.)

De afname van een algemene in plaats van een specifieke voortoets, zou een oplossing vormen voor het gesignaleerde probleem. Hiermee kan men bijvoorbeeld nagaan of er relevante "subsumers" aanwezig zijn (zie Lawton en Wanska, 1977, p. 240), en of voor de te leren stof relevante voorafgaande kennis beschikbaar is (vergelijk West en Fensham, 1976). De constructie van dergelijke algemene voortoetsen vormt echter tot op heden een onoplosbaar probleem.

In de literatuur werden verschillende adviezen geformuleerd met betrekking tot de verwerking van voortoetsgegevens. In hun klassieke artikel over experimentele en quasi-experimentele proefopzetten adviseerden Campbell en Stanley (1963) om bij de analyse van de data van "pretest-posttest-control-group(s)-designs" winstscores te berekenen (natoetsscore min voortoets-score, zie onder a), ofwel niveau's van voortoetscores te onderscheiden (zie onder b), ofwel een covariantie-analyse uit te voeren met de scores op de voortoets als covariant en de natoets als afhankelijke variabele (zie onder c). In de loop der jaren werden nog diverse andere methoden bedacht om de gegevens van "pretest-posttest-designs" te analyseren. In het onderstaande passeren de verschillende alternatieven de revue.

a) Analyse van ruwe winstscores

Veel auteurs pleitten tegen het gebruik van ruwe winstscores vanwege de onbetrouwbaarheid ervan, vanwege de negatieve correlatie met de voortoets, omdat voortoetsen meestal iets anders meten dan natoetsen (lack of common trait and

scale, zie Linn en Slinde (1977)), en omdat de aanname van een intervallschaal bij verschillcores extra zwaar telt (Cronbach en Snow, 1977, p. 73).

Recentelijk werden echter weer empirische (Richards, 1975) en theoretische (Hendrix et al., 1978) argumenten gelanceerd ten gunste van verschillcores. Toetsingen gebaseerd op verschillcores kunnen onder bepaalde omstandigheden (bijvoorbeeld bij gelijke voor- en natoets betrouwbaarheden) even veel onderscheidingsvermogen hebben als die welke gebaseerd zijn op andere afhankelijke variabelen. Over het algemeen lijkt het gebruik van verschillcores echter nog steeds ongewenst (Linn en Slinde, 1977; Cronbach en Snow, 1977).

- b) Analyse van de natoetsscores in een variantie-analyse met de voortoets als "blocking" variabele (bijvoorbeeld via mediaansplitsing)

Alle geraadpleegde auteurs waren het erover eens, dat deze zeer vaak gebruikte methode door gebrek aan precisie en onderscheidingsvermogen ongeschikt is (zie bijvoorbeeld Cronbach en Snow, 1977). Toch zijn er omstandigheden waarin deze methode van nut kan zijn, bijvoorbeeld wanneer niet aan de assumpties van covariantie-analyse voldaan is.

- c) Analyse van de natoetsscores in een covariantie-analyse met de voortoets als covariant

Cronbach en Furby (1970) adviseerden de voortoets op te nemen als covariant om zo de precisie en het onderscheidingsvermogen van de toetsing te vergroten. Linn en Slinde (1977) bespreken echter een onderzoek van Feldt op basis waarvan zij besluiten, dat opname van de voortoets als predictor alleen zinvol is als de voortoets een correlatie met de natoets heeft van minimaal .60. Deze oplossing is alleen verantwoord als er aselechte toewijzing aan de groepen die onder verschillende condities werken, heeft plaatsgevonden en als de regressiecoëfficiënten voor de verschillende groepen homogeen zijn (zie De Leeuw, 1979, p. 71).

- d) Analyse van de natoetsscores in een regressie-analyse met de voortoets als predictor

Regressie-analyse is een flexibeler analysetechniek dan covariantie-analyse: met gemak kunnen voortoetsscores en an-

dere predictoren gezamenlijk in de analyse worden opgenomen. Alternatief d vormt mijn inziens dan ook - zeker in een ATI-onderzoek (zie 5.3) - een betere oplossing dan alternatief c. Deze methode is ook bruikbaar indien de proefpersonen niet aselekt aan de condities zijn toegewezen (zie Cohen en Cohen, 1975, p. 398).

e) Analyse van de natoetsscores in een variantie-analyse of regressie-analyse

Deze methode wordt door verschillende auteurs aangeraden indien aselekte toewijzing aan condities heeft plaatsgevonden (vergelijk Cronbach en Furby, 1970; De Leeuw, 1979; Linn en Slinde, 1977). De voortoetsscores kunnen in de analyse buiten beschouwing blijven en vervullen in het experiment slechts controlefuncties (zie boven). Deze methode heeft ^{wanneer geen voorbets wordt afgenomen} een grotere interne en externe validiteit, omdat er geen "testing" en "interaction effects of testing" kunnen optreden (Campbell en Stanley, 1963). Wanneer aselekte toewijzing niet mogelijk is, vormt alternatief e echter een slechte oplossing.

f) Analyse van de natoetsscores in een regressie-analyse met voor attenuatie gecorrigeerde voortoetsscore als predictor

Cohen en Cohen (1975, p. 379) toonden aan, dat met name de onbetrouwbaarheid van de voortoetsscore een kritische factor vormt en dat partialisatie van een onbetrouwbare voortoets leidt tot een te geringe correctie voor voortoetsverschillen (terwijl verschillcores leiden tot overcorrectie). Wanneer men niet voor attenuatie corrigeert, werkt men met de assumptie dat de voortoets een betrouwbaarheid van 1.00 heeft (op.cit. p. 374).

Ook Cronbach en Snow (1977, p. 33) bepleitten correctie voor de diverse soorten meetfouten. Deze correctie voor attenuatie wordt - aldus Cohen en Cohen (1975, p. 372) - vooral noodzakelijk als de betrouwbaarheid van de predictor lager is dan .80, daarboven maakt de correctie niet veel verschil. Wanneer de betrouwbaarheid (i.e. homogeniteit) van een voortoets echter te laag wordt (bijvoorbeeld kleiner dan .50), verdient het mijn inziens echter de voorkeur

om - althans bij aselechte toewijzing aan de groepen die onder verschillende condities werken - de voortoets niet in de analyse te betrekken daar deze dan voornamelijk ruis meet (zie boven).

g) Analyse van de voor- en natoets als factor in een variantie-analyse volgens een "repeated measurement-design"

Sommige onderzoekers hebben de problemen van analyses met verschilcores omzeild door in hun variantie-analyse het meettijdscip (voortoets-natoets) als factor op te nemen in het design (bijvoorbeeld De Klerk, 1979). Hendrix et al. (1978) toonden onlangs echter aan dat de resultaten van zo'n proefopzet met herhaalde metingen exact gelijk zijn aan de resultaten van een variantie-analyse op verschilcores, omdat de termen in de beide variantie-analysemodellen tot elkaar herleidbaar zijn. Omdat methode g bovendien meer computertijd, meer ponskaarten en grotere interpretatieproblemen met zich meebrengt, vormt deze methode geen aantrekkelijk alternatief.

h) Analyse van de ruwe verschilcores in een covariantie-analyse met de voortoetsscore als predictor

Deze methode is - volgens Hendrix et al. - volledig equivalent met methode c (covariantie-analyse op de natoets met de voortoets als predictor), maar verdient de voorkeur omdat de verschilcores beter interpreteerbaar zijn dan natoetsscores. In de omstandigheden waarin voor methode c wordt gekozen, zou ook deze methode h in aanmerking komen.

i) Analyse van de residue winstcores

Om de problemen verbonden aan het gebruik van verschilcores te vermijden, werden diverse oplossingen voorgesteld. Eén daarvan was het voorstel om in plaats van de ruwe winstcores residue winstcores te berekenen. Residue winstcores zijn de afstanden van de natoetsscores tot de regressielijn die het verband tussen de voor- en de natoets weergeeft. Daardoor wordt het gedeelte uit de natoets-informatie dat rechtstreeks voorspelbaar is uit de voortoets, statistisch gecontroleerd (Cronbach en Furby, 1970). De residue winstscore geeft aan in hoeverre iemand meer veranderd

is dan op grond van zijn voortoetsscore verwacht kon worden. Cronbach en Furby (1970) toonden aan, dat het doel dat men met behulp van de berekening van residue winstscores wil bereiken efficiënter bereikt kan worden door covariantie-analyse (methode c) of regressie-analyse (methode d).

j) Analyse van geschatte "ware" verschillcores

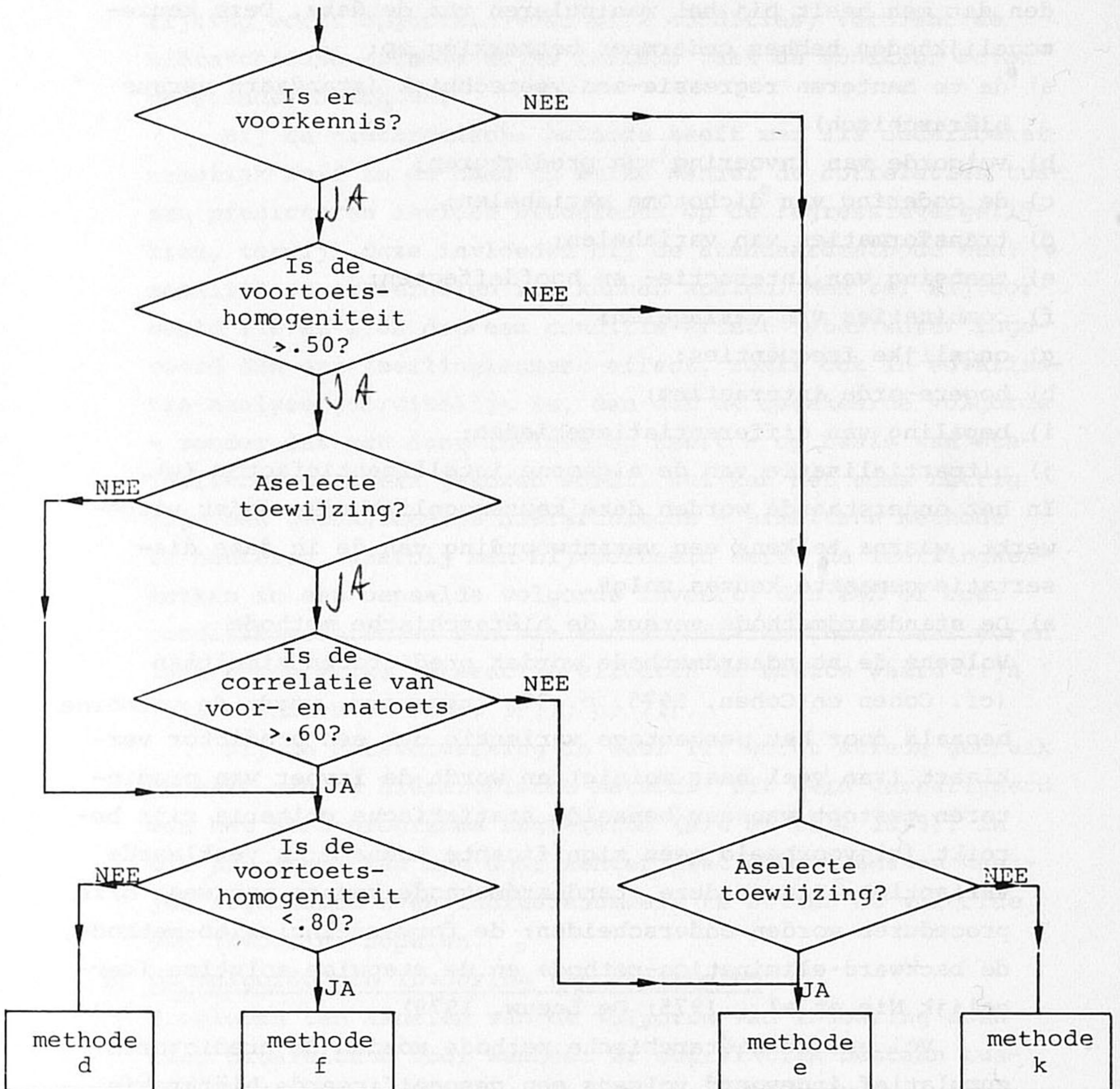
Sommige auteurs stelden voor om de winstscores te corrigeren voor onbetrouwbaarheid van de voor- en de natoets en de aldus verkregen "ware" winstscores als afhankelijke variabele te nemen. Cronbach en Furby (1970) toonden echter aan, dat ook deze oplossing ongeschikt is en dat berekening van de ware winstscores overbodig is, omdat er betere regressie-analytische procedures voorhanden zijn. Cohen en Cohen (1975, p. 379) toonden aan dat het er vooral om gaat te corrigeren voor de onbetrouwbaarheid van de voortoets (zie methode f) en niet voor die van de natoets.

Op grond van de besproken argumenten wordt in figuur 5.2 een samenvattend beslissingsschema gepresenteerd met behulp waarvan een analysemethode kan worden gekozen. Daarin komt ook nog een nieuwe methode k voor die mijns inziens nodig is als de voortoetsgegevens onbruikbaar zijn en de toewijzing van de proefpersonen niet aselekt geschiedde. In dat geval moet op een andere wijze gecorrigeerd worden voor eventueel van tevoren bestaande verschillen tussen de (intacte) groepen, bijvoorbeeld door een algemene voortoets of een algemene intelligentiescore als eerste predictor in een regressie-analyse op te nemen.

5.3 Het verwerken van gegevens van ATI-experimenten door middel van multi-pele regressie-analyse

Voor de verwerking van ATI-gegevens is multi-pele regressie-analyse de meest geschikte dataverwerkingstechniek (zie voor argumenten bijvoorbeeld Cronbach en Snow, 1977; De Leeuw, 1977, 1979). Een probleem bij het gebruik van deze techniek vormt echter de onervarenheid van de sterk op variantie-analyse gerichte psychologische onderzoekers en het ontbreken van ge-

Figuur 5.2 Wat te doen met de voortoets: een beslissingschema



methode d: Analyse van natoetsscores in een regressie-analyse met voortoets als predictor.

methode f: Analyse van natoetsscores in een regressie-analyse met voor attenuatie gecorrigeerde voortoets als predictor.

methode e: Analyse van de natoetsscores zonder de voortoets.

methode k: Analyse van de natoetsscores zonder de voortoets maar met IQ of algemene voorkennis als predictor.

schikte - specifiek op ATI toegespitste - handleidingen. Dit is te meer problematisch gezien het grote aantal keuzemogelijkheden dat men heeft bij het manipuleren van de data. Deze keuzemogelijkheden hebben ondermeer betrekking op:

- a) de te hanteren regressie-analysetechniek (standaard versus hiërarchisch);
- b) volgorde van invoering van predictoren;
- c) de codering van dichotome variabelen;
- d) transformaties van variabelen;
- e) toetsing van interactie- en hoofdeffecten;
- f) combinaties van variabelen;
- g) ongelijke frequenties;
- h) hogere-orde interacties;
- i) bepaling van differentiatiegebieden;
- j) uitpartialisatie van de algemene intelligentiefactor (g).

In het onderstaande worden deze keuzemogelijkheden nader uitgewerkt, waarna telkens een verantwoording van de in deze dissertatie gemaakte keuzes volgt.

a) De standaardmethode versus de hiërarchische methode

Volgens de standaardmethode worden predictoren simultaan (cf. Cohen en Cohen, 1975, p. 97) ingevoerd, wordt de volgorde bepaald door het percentage variantie dat een predictor verklaart (van veel naar weinig) en wordt de invoer van predictoren gestopt wanneer bepaalde statistische criteria zijn bereikt (bijvoorbeeld geen significante toename in verklaarde variantie). Binnen deze standaardmethode kunnen nog weer drie procedures worden onderscheiden: de forward-inclusion-methode, de backward-elimination-methode en de stepwise solution (vergelijk Nie et al., 1975; De Leeuw, 1979).

Volgens de hiërarchische methode worden de predictoren cumulatief ingevoerd volgens een gespecificeerde hiërarchie. Wanneer afzonderlijke regressie-analyses per conditie worden uitgevoerd, is de simultane methode ongeschikt: Als gevolg van kansfluctuaties zullen de regressie-analyses tot gevolg hebben dat het aantal predictoren in de verschillende condities niet gelijk is (zie Cronbach en Snow, 1977, p. 40).

Maar ook bij gebruik van de "generalized regression approach", waarbij via één of meer dummyvariabelen één regressievergelijking wordt opgesteld voor alle condities, verdient de hiërarchische methode mijns inziens vaak de voorkeur boven de standaardmethode.

Bij de hiërarchische methode heeft men als onderzoeker namelijk zelf in de hand op welke manier de correlaties tussen predictoren invloed uitoefenen op de regressievergelijking, terwijl deze invloeden bij de standaardmethode een moeilijk te achterhalen rol kunnen spelen. Men zal bijvoorbeeld liever zien dat een conditie-effect later wordt ingevoerd dan een leerlingkenmerk-effect, zoals ook in covariantie-analyse gebruikelijk is, dan dat de omgekeerde volgorde - zonder dat men daar invloed op heeft - op basis van statistische criteria gekozen wordt. Wel kan het soms nuttig zijn een gecombineerde hiërarchische - simultane methode te hanteren, waarbij men bijvoorbeeld eerst de leerlingkenmerken in een bepaalde volgorde invoert, dan één of meer conditievectoren en dan via een standaardmethode naar voren laat komen welke interactie-effecten de moeite waard zijn (zie Cronbach en Snow, 1977, p. 71).

Bij de dataverwerking in deel III wordt steeds gebruik gemaakt van de hiërarchische methode. Dit werd gerealiseerd via het SPSS-programma Regression (Nie et al., 1975). In dit programma kan men door achter predictoren tussen haakjes zogenaamde even inclusienummers te zetten de volgorde van invoering bepalen.

b) De volgorde van invoering van variabelen

Problemen ten aanzien van de volgorde van invoering doen zich uitsluitend voor wanneer er correlaties bestaan tussen predictoren. Wanneer een variabele B wordt ingevoerd na een ermee gecorreleerde variabele A, kan B alleen nog maar aan de verklaring van de variantie bijdragen met zijn eigen specifieke variantie (die B dus niet gemeenschappelijk met A had). In de volgorde A-B kan B dus vaak minder variantie verklaren dan in de volgorde B-A.

In een groot aantal gevallen maakt de volgorde van invoering geen verschil, simpel omdat er geen correlaties bestaan tussen predictoren. Zo zijn hoofdeffectvectoren en interactievectoren onder bepaalde condities (gelijke aantallen proefpersonen in de cellen, effect- of orthogonale codering) niet met elkaar gecorreleerd (vergelijk Kerlinger en Pedhazur, 1973). Evenmin bestaat er onder bepaalde omstandigheden een correlatie tussen verschillende dummyvectoren die worden gecreëerd om drie of meer condities van elkaar te onderscheiden.

In andere gevallen kan de volgorde wel degelijk problemen opleveren. Er werd reeds gewezen op correlaties tussen leerlingkenmerken en conditievectoren (bijvoorbeeld wanneer in de groep die onder de ene conditie werkt toevallig hoger wordt gescoord op een bepaald leerlingkenmerk dan in de groep die onder de andere conditie werkt). Ook correlaties tussen leerlingkenmerken onderling en tussen verschillende interactie-effecten kunnen aanzienlijk zijn. Bij ongelijke celfrequenties tenslotte (zie ook g) kunnen ook correlaties tussen hoofdeffectvectoren en interactievectoren optreden (cf. Kerlinger en Pedhazur, op.cit.).

Bij de bepaling van de volgorde van invoering van predictoren kan rekening worden gehouden met:

- 1) De correlaties tussen de predictoren.
- 2) Temporele of logische ordening (bijvoorbeeld: algemene intelligentie kan wel de voortoetsscore mee bepalen, maar het omgekeerde kan niet).
- 3) Leerlingkenmerken kunnen - net als in een covariantie-analyse - beter voorafgaan aan experimentele variabelen.
- 4) De voortoets (indien bruikbaar, zie 5.2.2) kan beter voorafgaan aan de experimentele variabelen (vergelijk Cohen en Cohen, 1975, p. 100).
- 5) Variabelen met een hoge onderzoeksrelevantie (bijvoorbeeld die variabelen waarop het onderzoek in de eerste plaats was gericht) kan men een hogere prioriteit geven dan variabelen met een minder hoge onderzoeksrelevantie (variabelen die men ook "meegenomen" heeft) (op.cit.).

p. 100). Waar deze regels geen uitkomst bieden, zullen verschillende volgordes moeten worden beproefd (zie Borich, Godbout en Wunderlich, 1976). Bij de dataverwerking in deel III wordt de volgorde waar nodig bepaald via deze regels.

c) De codering van nominale variabelen

Om nominale variabelen (bijvoorbeeld de conditievariabele of schooltype) in een "generalized regression-analysis" te kunnen opnemen in de regressievergelijking, worden dummyvariabelen gecreëerd. Men heeft daarbij als onderzoeker de keuze tussen dummycodering, effectcodering en orthogonale codering (Kerlinger en Pedhazur, 1973). Bij nominale variabelen met twee niveaus (bijvoorbeeld experimentele versus controlegroep) zijn effectcodering en orthogonale codering identiek: de ene groep kent men de score +1 toe en de andere -1. Bij dummycodering kent men aan de tweede groep in plaats van de score -1 de score 0 toe. In tabel 5.1 zijn voorbeelden opgenomen van de drie typen van codering bij nominale variabelen met drie niveaus (men dient dan twee vectoren te creëren).

Tabel 5.1 De drie typen van codering bij een nominale variabele met drie niveaus

	Dummycodering		Effectcodering		Orthogonale codering	
	vector 1	vector 2	vector 1	vector 2	vector 1	vector 2
niveau 1	1	0	1	0	1	$-\frac{1}{2}$
niveau 2	0	1	0	1	-1	$-\frac{1}{2}$
niveau 3	0	0	-1	-1	0	1

Men noemt een codering - bij gelijke frequenties in de cellen - orthogonaal wanneer de som van de producten van de coëfficiënten gelijk is aan nul. In het voorbeeld in tabel 5.1 is $(1)(-\frac{1}{2}) + (-1)(-\frac{1}{2}) + (0)(1)$ inderdaad gelijk aan nul, terwijl in het geval van de effectcodering de som $(1)(0) + (0)(1) + (-1)(-1)$ ongelijk is aan nul (namelijk +1).

Hoewel de totaalresultaten van de regressie-analyse (in termen van het percentage verklaarde variantie) bij de drie coderingswijzen identiek zijn, treden er verschillen op met betrekking tot de bepaling van de B-gewichten, de significantietoetsing en het hanteren van ongelijke frequenties (cf. Kerlinger en Pedhazur, op.cit.). Een belangrijk voordeel van effectcodering is dat de resulterende regressievergelijking het lineaire model weerspiegelt. Dat betekent dat de regressiecoëfficiënten de effecten van de condities weerspiegelen en de constante term het totaal-gemiddelde (grand mean) vormt. In een opzet met één experimentele en één controlegroep is volgens Cohen en Cohen (1975, p. 211) dummycodering de beste keus, hoewel andere keuzes ook legitiem zijn.

In hoofdstuk 7 is toch gekozen voor effectcodering, omdat de meeste andere onderzoekers (De Leeuw, 1979; Peterson, 1979; Snow, 1977) dat ook deden. Bij meer dan twee condities maakt het wel wat uit: Effectoedering levert gemakkelijker interpreteerbare resultaten dan dummycodering. Orthogonale codering is echter de beste keuze wanneer men a priori geplande contrasten belangrijker vindt dan post-hoc-vergelijkingen (cf. Kerlinger en Pedhazur, 1973, p. 131). Daarom wordt bij de verwerking van de experimenten die in hoofdstuk 8 worden gerapporteerd, de orthogonale codering gebruikt.

d) Transformaties van variabelen

Om verschillende redenen kan het nuttig zijn variabelen te transformeren (bijvoorbeeld om beter te kunnen extrapoleren naar andere steekproeven, om scheve verdelingen te corrigeren, om betere vergelijkingen tussen variabelen te verkrijgen, en dergelijke). De meest gebruikte transformaties in de ATI-research zijn deviatiescores ($X - \bar{X}$) en standaardscores ($z = \frac{X - \bar{X}}{s}$). Cronbach en Snow (1977) pleitten voor het gebruik van deviatie- of standaardscores voor predictoren en tegen transformaties van afhankelijke variabelen. Berekening van predictor standaardscores is vooral geïndiceerd, wanneer composities van leerlingkenmerken worden gevormd

en wanneer hogere-orde interacties worden getoetst. Maar ook wanneer dat niet gebeurt, verdient het volgens Cronbach en Snow (1977; zie ook Peterson, 1979, p. 526) aanbeveling de predictoren te transformeren tot deviatiescores of standaardscores. Deze transformaties dienen natuurlijk wel voor de gehele steekproef te worden uitgevoerd en niet per groep (Cronbach en Snow, 1977, p. 27, p. 514).

In alle dataverwerkingen die in deel III worden gerapporteerd, werden daarom de predictoren gestandaardiseerd (over de hele steekproef). Als b -coëfficiënten worden conform de voorschriften van Cronbach en Snow (1977) de ongestandaardiseerde en niet de gestandaardiseerde gewichten gerapporteerd.

e) Toetsing van interactie- en hoofdeffecten

Wanneer men de hiërarchische methode (of een gecombineerde methode) toepast, zijn de door het SPSS-programma Regression geproduceerde F-waarden in het algemeen ongeschikt: "The researcher should be aware that the standard F-test for a partial regression coëfficiënt gives credit to each variable only for its incremental contribution after all the other independent variables have been introduced in the equation." (Nie et al., 1975, p. 338).

De significantietoetsing die Corno (1979) onlangs rapporteerde, is dan ook onjuist. Ook De Leeuw (1979) maakte ter toetsing van de interactie-effecten gebruik van de door het SPSS-programma geleverde F-waarden (telefonische mededeling). Dat vormt wat betreft de simpele interactie-effecten geen probleem omdat die terecht als laatste predictor fungeren. Wel problematisch is dat echter wanneer De Leeuw zowel de voortoetsinteractie als andere leerlingkenmerkinteracties in de regressievergelijking opneemt. De toetsing van de voortoetsinteractie is in dat geval onjuist.

Wanneer men volgens de hiërarchische methode werkt, heeft men twee mogelijkheden om bijdragen van predictoren aan de verklaring van de variantie op significantie te toetsen (Cohen en Cohen, 1975, pp. 135-144). Volgens model I toetst men elk van de individuele bijdragen tegen

andere foutenvarianties, namelijk steeds tegen die variantie die resteert als de voorgaande predictoren in de hiërarchie en de predictor zelf uitgepartialiseerd zijn, maar de erop volgende predictoren nog niet. De formule voor de F-statistic is daarbij:

$$F = \frac{(R^2_{y.AB} - R^2_{yA}) / k_B}{(1 - R^2_{y.AB}) / (n - k_A - k_B - 1)}$$

waarbij $R^2_{y.AB}$ = de proporties door predictoren A en B verklaarde variantie in y.

$R^2_{y.A}$ = de proporties door predictor A verklaarde variantie in y.

n = het aantal proefpersonen.

k_A = aantal vrijheidsgraden van A.

k_B = aantal vrijheidsgraden van B.

Volgens model II toetst men elk van de individuele bijdragen tegen dezelfde foutenvariantie, namelijk tegen die variantie die resteert na het uitpartialiseren van alle predictoren. De formule voor de F-statistic wordt dan

$$F = \frac{(R^2_{y.AB} - R^2_{y.A}) / k_B}{(1 - R^2_{y.ABC}) / (n - k - 1)}$$

waarbij $R^2_{y.ABC}$ = de proportie door A, B en C verklaarde variantie in y.

k = het totaal van de aantallen vrijheidsgraden van A, B en C (= $k_A + k_B + k_C$).

Het voordeel van model II is dat een zuiverder schatting van de foutenvariantie wordt verkregen. Dat gaat echter ten koste van het aantal vrijheidsgraden in de foutenterm, alsmede van een verhoging van de kritieke F-waarde. Het is dan ook van vele factoren afhankelijk welk model de voorkeur verdient (zoals aantal proefpersonen, aantal predictoren, researchstrategie, relevantie, en dergelijke (zie Cohen en Cohen, 1975)). Op basis van de regels van Cohen en Cohen (1975, p. 303) en in navolging van onder

andere Cronbach en Snow (1977), Snow (1977) en Peterson (1979) is in dit proefschrift gekozen voor model II.

De algemene formule die daarbij gehanteerd wordt, is volgens Cronbach en Snow (1977, p. 67):

$$F = \frac{\Delta SS_{\text{Regr.}} / \text{df for change}}{MS_{\text{Res.}}}$$

waarbij $MS_{\text{Res.}}$ de "meansquare" vormt die is gebaseerd op het totale model en $\Delta SS_{\text{Regr.}}$ de toename in "sum of squares" die het gevolg is van één of meer predictoren (ergens in de hiërarchie). Deze formule is equivalent met de volgende (cf. Kerlinger en Pedhazur, 1973, p. 124):

$$\frac{\Delta R^2 / \text{df for change}}{(1 - R^2_{y.1\dots k}) (n - k - 1)}$$

waarbij: ΔR^2 = De toename in proportie verklaarde variantie onder invloed van één of meer predictoren.

df for change = Het aantal vrijheidsgraden dat met die verandering gepaard gaat.

$R^2_{y.1\dots k}$ = De proportie totaal door alle predictoren verklaarde variantie.

n = Het aantal proefpersonen.

k = Het aantal predictoren.

Een probleem dat zich bij het toetsen van interacties voordoet, vormt het gebrek aan onderscheidingsvermogen van de statistische toetsen. Cronbach en Snow (1977, p. 57) berekenden de kansen op β -fouten onder bepaalde voorwaarden (zoals matige correlaties) bij verschillende aantallen proefpersonen. De kans op een β -fout bleek bij de gebruikelijke aantallen proefpersonen erg groot. Cronbach en Snow pleitten dan ook voor vergroting van de aantallen proefpersonen per conditie (minstens 100), waardoor de kans op een β -fout rond de 10 procent komt te liggen. In de praktijk van het onderzoek is het vaak echter moeilijk om aan deze eis te voldoen. Een andere oplossing - die minder goed

is, maar toch ook wat helpt - vormt de verhoging van de kans op de Type-I-fout (α), waardoor de β -fout wat kleiner wordt. In dit proefschrift worden daarom bij de toetsing van interacties α -waarden van .10 aangehouden.

f) Combinaties van variabelen

"It is hard to give general advice about the handling of multiple aptitude measures. The alternatives are to treat them singly, to carry out a multiple regression within treatments, to enter them in a generalized regression analysis, to form composites or principal components and then apply one of the preceding methods or to make a full multivariate analysis." (Cronbach en Snow, 1977, p. 515). Ondanks hun aarzeling om een advies te geven over het behandelen van diverse leerlingvariabelen (zie citaat), spreekt uit diverse delen van het handboek toch de voorkeur van Cronbach en Snow voor de vorming van combinaties van variabelen.

Regelmatig adviseren zij bijvoorbeeld om in plaats van twee leerlingkenmerken afzonderlijk in te voeren, twee gecombineerde variabelen te vormen: Eén variabele die bestaat uit de som van de twee variabelen en één variabele die bestaat uit het verschil tussen de twee variabelen. De somvariabele representeert de gecombineerde bijdrage van de twee variabelen en via de verschilvariabele kan men nagaan of de variabelen verschillende porties variantie verklaren. Bij een dergelijke benadering zijn - aldus Cronbach en Snow - deze som- en verschilvariabelen - mits gestandaardiseerd - niet met elkaar gecorreleerd wat als voordeel heeft, dat de volgorde van invoering niet meer relevant is. Alleen als de verschilvariabele (in interactie met het conditie-effect) een significante bijdrage levert, is het nodig (gewenst) aan beide leerlingkenmerken aandacht te besteden. Op andere plaatsen adviseren Cronbach en Snow op theoretische gronden a priori combinaties te vormen of via principale componentenanalyse een reductie in het aantal variabelen te bewerkstelligen.

In deze dissertatie is gekozen voor een andere oplossing. Eerst worden diverse leerlingkenmerk-conditie-interacties afzonderlijk getoetst (eventueel onder uitpartialisering van de voortoets (zie 5.2)). Als er meer dan één interactie significant of bijna significant is, wordt via een procedure die beschreven is door Borich et al. (1976) nagegaan of de verschillende interacties dezelfde variantie verklaren of verschillende. Volgens deze procedure moeten er drie regressie-analyses worden uitgevoerd: Eén waarbij de interacties simultaan worden ingevoerd, één waarbij eerst de ene interactie wordt ingevoerd en dan pas de tweede en één met de omgekeerde volgorde. Borich et al. hebben regels beschreven om te bepalen welke interactie(s) in ogenschouw moeten worden genomen bij de verschillende mogelijke uitkomsten van de drie analyses.

g) Ongelijke celfrequenties

Hoewel de problemen die ontstaan door ongelijke celfrequenties niet zo groot zijn als bij variantie-analyse, treedt dan toch ook bij multipele regressie-analyse een aantal problemen op (zie Kerlinger en Pedhazur, 1973):

1. De regressie-coëfficiënten in de "generalized regression" vergelijking komen niet meer overeen met de coëfficiënten die in afzonderlijke regressie-analyses per conditie worden verkregen. Ze vormen namelijk ongewogen gemiddelden van de afzonderlijke coëfficiënten.
2. De hoofdeffecten en interactie-effecten zijn niet meer orthogonaal, met als gevolg dat de volgorde van invoering van de predictoren nog belangrijker wordt.
3. Bij orthogonale codering zijn de verschillende contrasten niet meer orthogonaal.

Hoe het probleem van de ongelijke frequenties moet worden opgelost, hangt af van de gekozen coderingsvorm. Bij dummy- en effectcodering zijn er drie mogelijkheden:

- 1) Men kan aselekt een aantal proefpersonen buiten beschouwing laten zodat de frequenties toch gelijk worden.
- 2) Volgens de experimentele benadering moet men het percentage variantie dat door een factor verklaard wordt corrige-

ren voor de correlatie van die factor met alle andere factoren (op.cit. p. 190).

3) Volgens de a priori methode bepaalt men op theoretische gronden een volgorde voor de verschillende factoren zonder verder correcties toe te passen (op.cit. p. 193).

In het geval van orthogonale codering tenslotte kan men het probleem van de ongelijke frequenties opvangen door andere coderingen te kiezen. Twee vergelijkingen met ongelijke frequenties in de cellen (n_1 , n_2 en n_3) zijn dan orthogonaal als $n_1c_{11}c_{21} + n_2c_{12}c_{22} + n_3c_{13}c_{23}$ gelijk is aan nul, waarbij de verschillende c's de coderingscoëfficiënten vormen. In hoofdstuk 8 worden dergelijke orthogonale coderingen bij ongelijke frequenties toegepast.

h) Interacties van hogere orde

Cohen en Cohen (1975, p. 336) waarschuwen tegen het ongebreideld opnemen van interacties van hogere orde (bijvoorbeeld intelligentie x angst x conditie) in de regressie-analyses. Toch zijn er aanwijzingen dat dergelijke complexe interacties de werkelijkheid het best weerspiegelen (vooral in ATI-research, zie Snow, 1977). De oplossing die in deze dissertatie is gekozen, bestaat hieruit, dat op beperkte schaal interacties van hogere orde worden geëxploreerd, met name met betrekking tot die interacties waarvoor a priori een rationale kan worden bedacht (zie ook hoofdstuk 1).

i) Differentiatiegebieden

In hoofdstuk 2 werd betoogd dat ordinale interacties net zo goed belangrijk zijn als disordinale. Toch heeft het nog wel zin te bepalen of een interactie ordinaal dan wel disordinaal is. Er bestaan verschillende methoden om dat vast te stellen (zie bijvoorbeeld Bracht, 1970, Cronbach en Snow, 1977). In Nederland heeft Plomp (1977) technieken beschreven waarmee kan worden nagegaan of een interactie ordinaal of disordinaal van aard is. Deze technieken zijn gebaseerd op methoden van Johnson-Neyman en Pothof.

Volgens Plomp (1977) en anderen is een interactie alleen disordinaal wanneer de regressielijnen die op ware

scores zijn gebaseerd elkaar snijden binnen de range van gemeten waarden van een leerlingkenmerk. Om te bepalen of een interactie disordinaal is, moet men dus niet alleen weten of de regressielijnen elkaar snijden, maar ook of dit binnen de range van gemeten waarden gebeurt. Bovendien moet daarbij dan ook nog rekening worden gehouden met meetfouten. Het is dus niet voldoende dat men constateert, dat de regressielijnen die op basis van waargenomen scores (en niet de ware scores) zijn berekend elkaar snijden binnen de range van gemeten waarden op het leerlingkenmerk. Waar het om gaat is, dat de regressielijnen elkaar ook nog snijden binnen de range, na correctie voor meetfouten.

Om via deze principes te bepalen of een interactie disordinaal is of niet, werd door Johnson-Neyman een methode ontwikkeld waarbij differentiatiegebieden bepaald worden (zie Plomp, 1977). Differentiatiegebieden zijn die gemeten waarden van een leerlingkenmerk (of van meer dan één leerlingkenmerk) waarvoor geldt dat de bijbehorende voorspelde ware criteriumscores in verschillende condities significant van elkaar verschillen. Differentiatiegebieden kunnen volgens Plomp bij twee condities en één leerlingkenmerk worden bepaald via de ongelijkheid $d^2 - F_{1,df} \cdot \hat{v}_e \cdot P^2 > 0$. In deze formule is d het verschil bij een bepaalde waarde op het leerlingkenmerk tussen de voorspelde ware criteriumscores in de verschillende condities; $F_{1,df}$ de waarde die in de F-tabel voorkomt met 1 en $df = N_1 + N_2 - 4$ vrijheidsgraden (waarbij N_1 het aantal proefpersonen in de ene conditie en N_2 het aantal proefpersonen in de andere conditie); \hat{v}_e is de schatting van de fouten- of residuvariatie in de populatie en P^2 een quadratische functie in x (het leerlingkenmerk) die volgens de volgende formule kan worden bepaald:

$$P^2 = \frac{1}{N_1} + \frac{1}{N_2} + \frac{(x - \bar{x}_1)^2}{(N_1 - 1) V_{x1}} + \frac{(x - \bar{x}_2)^2}{(N_2 - 1) V_{x2}}$$

- waarin: \bar{x}_1 = De gemiddelde score op het leerlingkenmerk in conditie 1.
- \bar{x}_2 = De gemiddelde score op het leerlingkenmerk in conditie 2.
- V_{x1} = De variantie van het leerlingkenmerk in conditie 1.
- V_{x2} = De variantie van het leerlingkenmerk in conditie 2.

In deze dissertatie zullen bij significante interacties de differentiatiegebieden worden bepaald met behulp van een tweetal computerprogramma's die zijn beschreven door De Leeuw (1977).

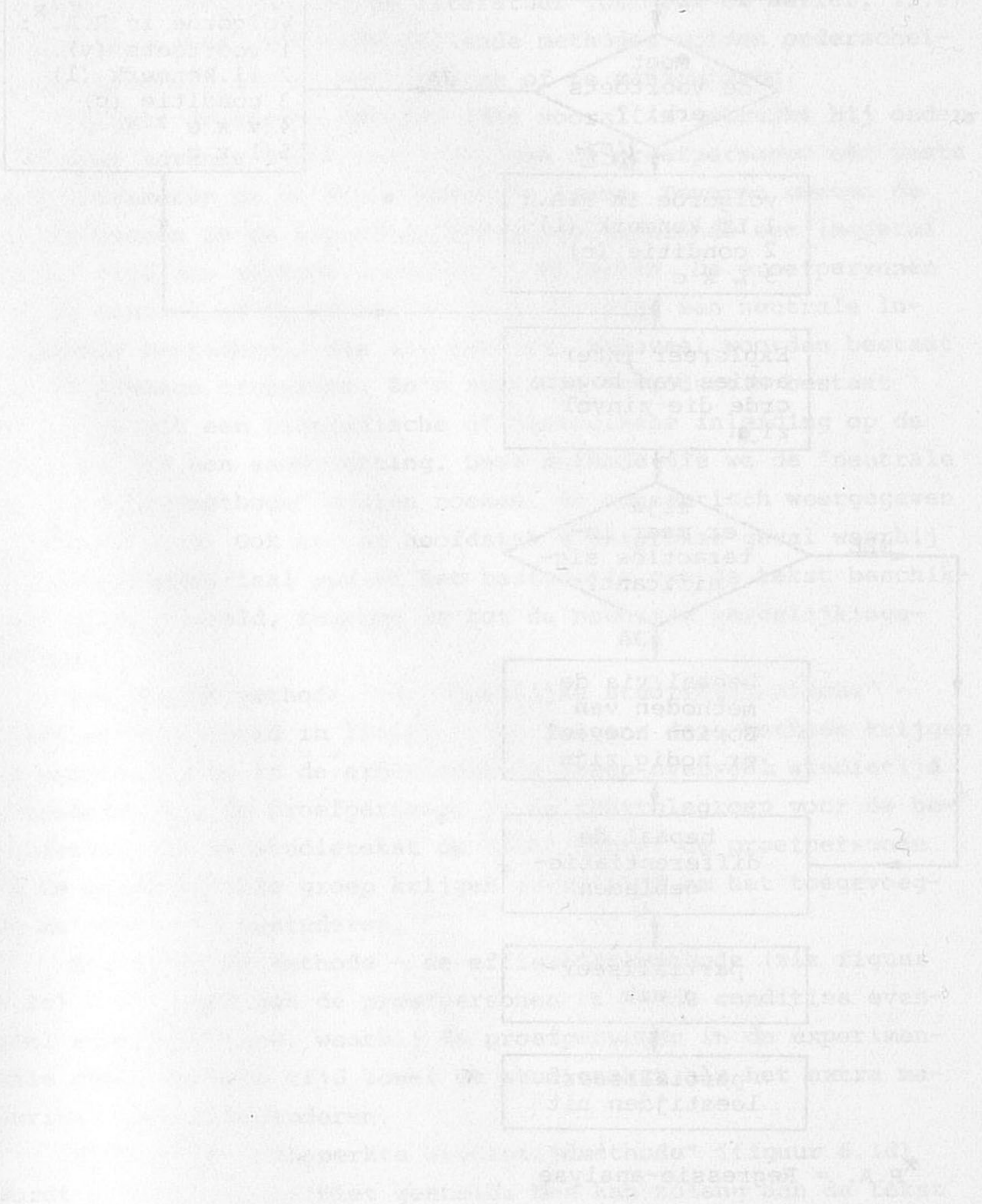
j) Uitpartialisatie van de g-factor en leestijden

Eén van de weinige ATI-hypothesen waarvoor langzamerhand enige steun bestaat werd door Snow (1977, p. 69) als volgt omschreven: "Instructional treatments differ in the informationprocessing burdens they place on, or remove from, the responsibility of the learner, and the regression slopes of cognitive outcomes on g become steeper or shallower accordingly." Soms bleken simplificaties, algoritmen en aanwijzingen zelfs nadelig voor studenten die hoog op algemene intelligentie (g) scoorden (op.cit. p. 75). Snow concludeerde dan ook: "In the meantime, the interpretation of most cognitive tests in ATI rests on g. And, for the sake of parsimony, any study investigating a more specialized aptitude must show its separation from g empirically." (p. 81).

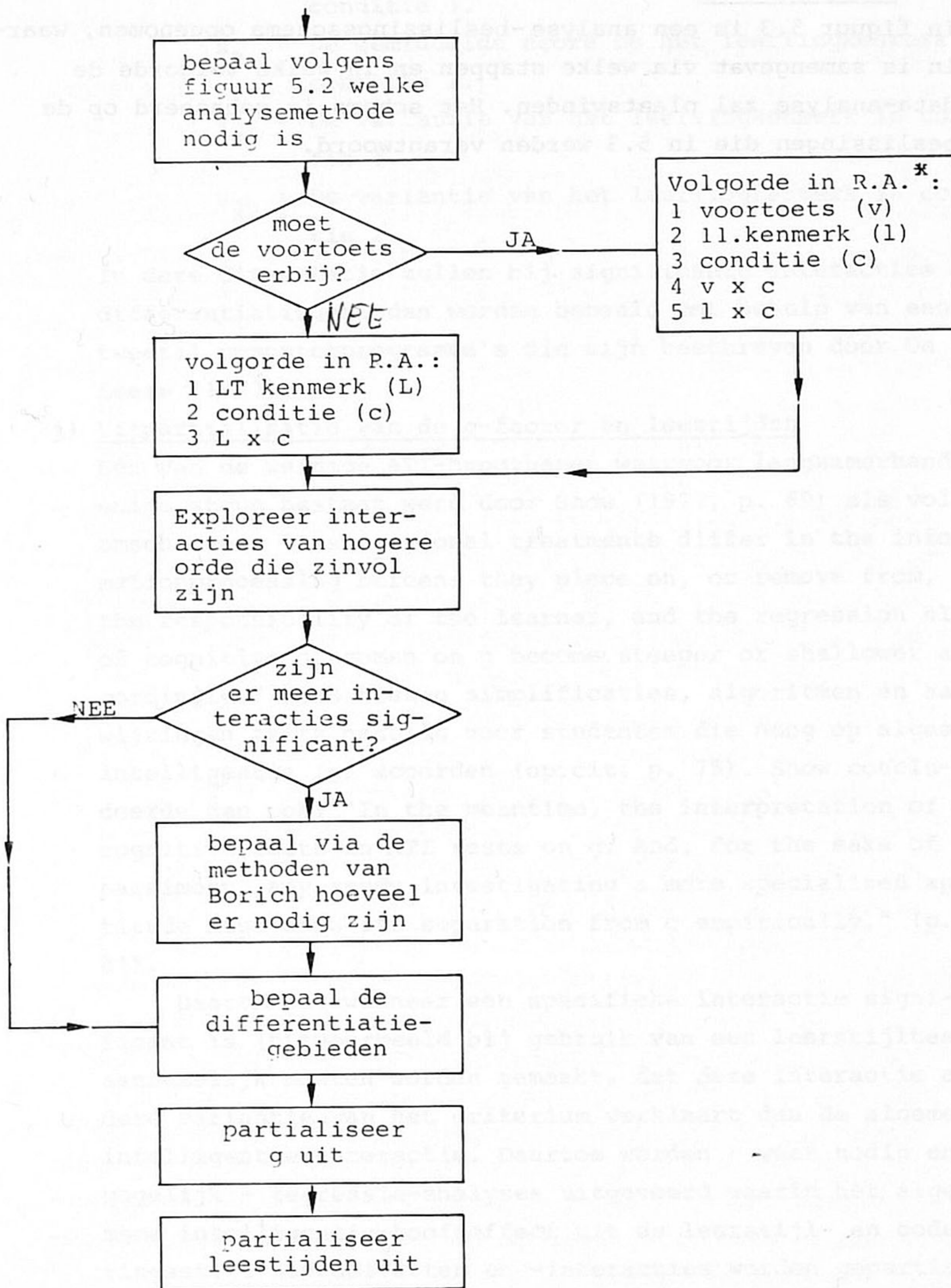
Daarom zal wanneer een specifieke interactie significant is (bijvoorbeeld bij gebruik van een leerstijltest) aannemelijk moeten worden gemaakt, dat deze interactie andere variantie van het criterium verklaart dan de algemene intelligentie-interactie. Daartoe worden - waar nodig en mogelijk - regressie-analyses uitgevoerd waarin het algemene intelligentie-hoofdeffect uit de leerstijl- en coderingsstijl-hoofdeffecten en -interacties worden gepartialiseerd. Hetzelfde zal - waar nodig en zinvol - overigens ook gebeuren met de leestijdeffecten.

5.4 Analyseschema

In figuur 5.3 is een analyse-beslissingsschema opgenomen, waarin is samengevat via welke stappen en in welke volgorde de data-analyse zal plaatsvinden. Het schema is gebaseerd op de beslissingen die in 5.3 werden verantwoord.



Figuur 5.3 Analyseschema



*R.A. = Regressie-analyse

Hoofdstuk 6 Het controleren van studietijden

6.1 Methoden om studietijden te controleren

Het gegeven dat extra studiemateriaal extra studietijd vraagt, leidt ertoe dat het extra belangrijk wordt aandacht te besteden aan studietijden. In de literatuur (cf. Faw en Waller, 1976) kunnen tenminste vier verschillende methoden worden onderscheiden om studietijden te controleren of te manipuleren.

Volgens de eerste methode (die vooral is gebruikt bij onderzoek naar advance organizers) krijgen de proefpersonen een vaste tijd toegemeten om de studietekst te leren. Tevoren moeten de proefpersonen in de experimentele groep gedurende een (meestal ruime) tijd een advance organizer bestuderen. De proefpersonen uit de controlegroep moeten in dezelfde tijd een neutrale introductie bestuderen, die uit ongeveer evenveel woorden bestaat als de advance organizer. Zo'n neutrale introductie bestaat doorgaans uit een biografische of historische inleiding op de tekst of uit een samenvatting. Deze methode die we de "neutrale vergelijkingsmethode" zullen noemen, is schematisch weergegeven in figuur 6.1a. Ook het in hoofdstuk 4 besproken geval waarbij introductiemateriaal pas na het bestuderen van de tekst beschikbaar wordt gesteld, rekenen we tot de neutrale vergelijkingsmethode.

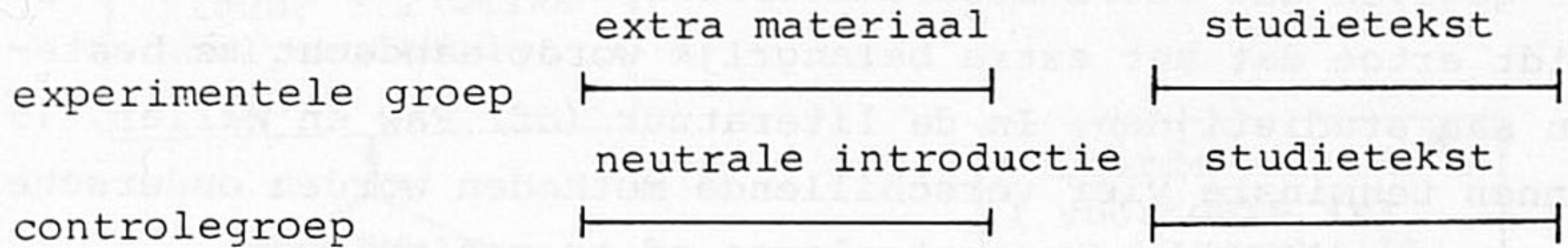
Een tweede methode - de "ongelijke studietijdmethode" - wordt geïllustreerd in figuur 6.1b. Volgens deze methode krijgen de proefpersonen in de experimentele groep evenveel studietijd toegemeten als de proefpersonen in de controlegroep voor de bestudering van de studietekst op zich. Echter de proefpersonen in de experimentele groep krijgen extra tijd om het toegevoegde materiaal te bestuderen.

Bij de derde methode - de efficiëntiemethode (zie figuur 6.1c) - kent men aan de proefpersonen in beide condities evenveel studietijd toe, waarbij de proefpersonen in de experimentele groep in deze tijd zowel de studietekst als het extra materiaal moeten bestuderen.

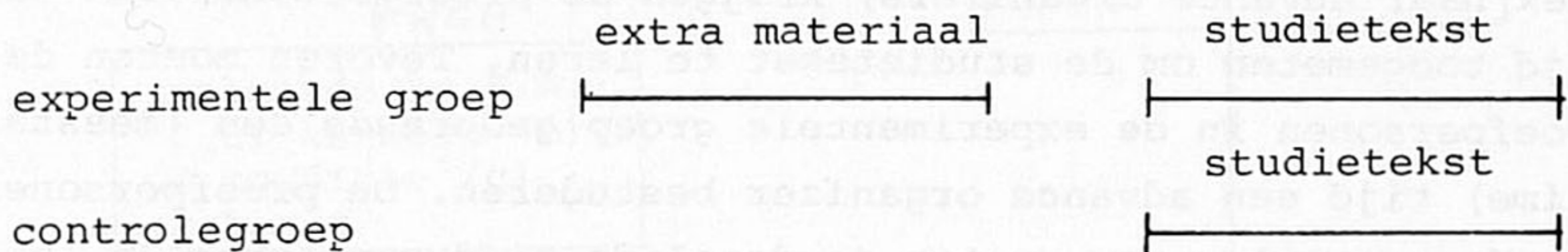
Volgens de "onbeperkte studietijdmethode" (figuur 6.1d) wordt er geen tijdslimiet gesteld: Men kan zolang aan de tekst

Figuur 6.1 Schematische weergave van de vier studietijdcontrole-
methoden*

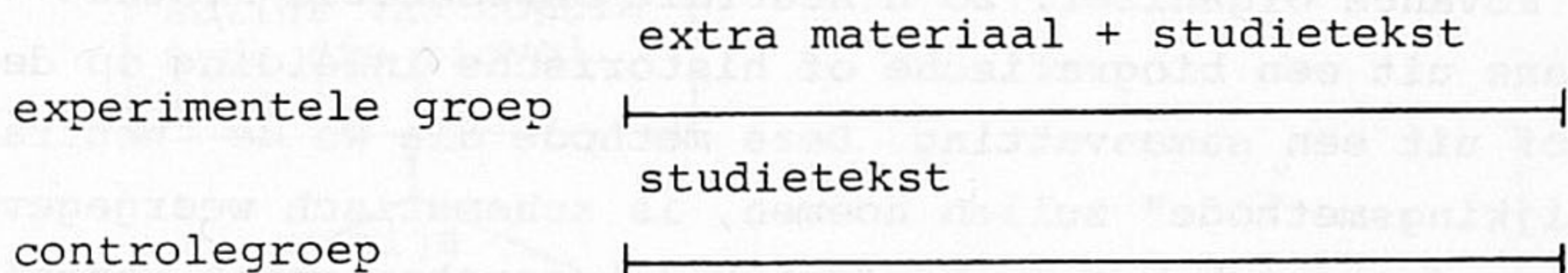
a) De neutrale vergelijkingsmethode



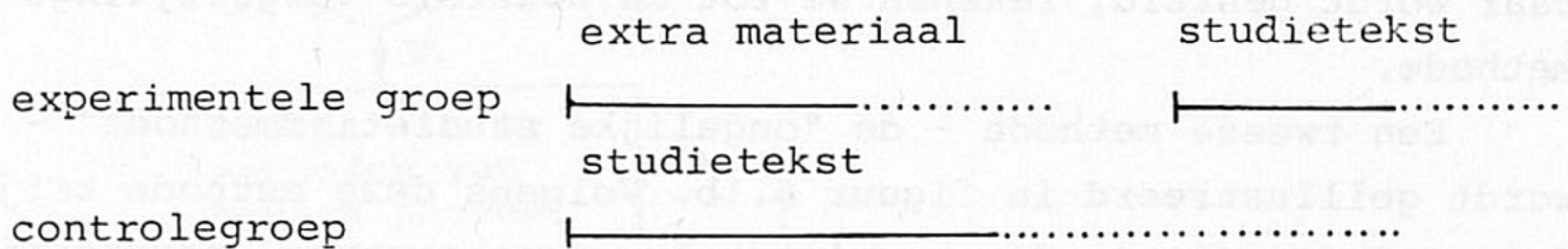
b) De ongelijke studietijdmethode



c) De efficiëntiemethode



d) De onbeperkte studietijdmethode



*De horizontale strepen stellen tijdsassen voor. De verticale streepjes geven begin en eind van een studiesessie weer en puntjes wijzen op het ontbreken van een tijdslimiet. Boven die tijdsassen is weergegeven wat in het betreffende tijdsinterval moest worden bestudeerd.

en het extra materiaal werken als men zelf wenselijk acht. Soms - maar helaas lang niet altijd - wordt gemeten hoeveel tijd de proefpersonen gebruiken en wordt een soort leerwinst per tijds-eenheidsscore berekend.

6.2 Voor- en nadelen van de verschillende studietijdcontrole-methoden

De neutrale vergelijkingsmethode biedt op een enigszins vreemde wijze controle op de studietijdverlenging in de experimentele conditie: De studietijd van de controleproefpersonen wordt kunstmatig verlengd, doordat zij bezig worden gehouden met een neutrale taak. Als er een effect van de onafhankelijke variabele wordt aangetoond, kan dan ook alleen maar worden geconcludeerd, dat het toegevoegde instructiemateriaal beter is dan (slecht) ander instructiemateriaal (Faw en Waller, 1976).

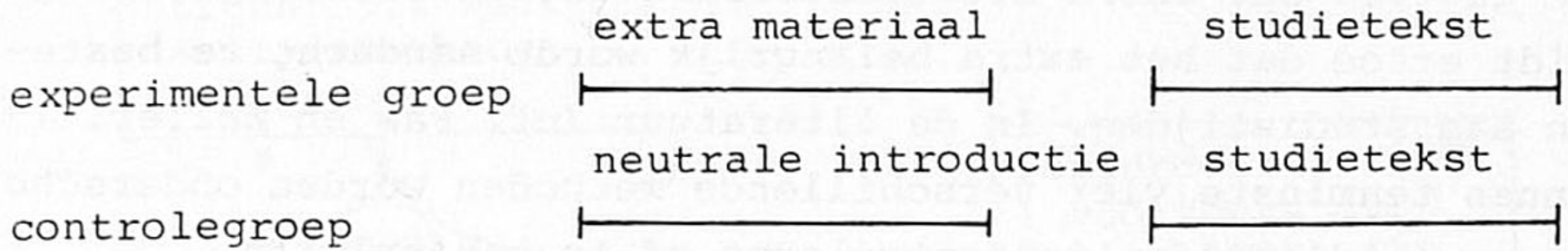
Een ernstig probleem dat aan de neutrale vergelijkingsmethode is verbonden (dat echter niet geldt voor post organizer studies) werd geconstateerd door Jaspers (1977) en heeft betrekking op de studiehoeveelheid. Over de neutrale (biografische, historische) introductie worden weliswaar geen vragen gesteld, maar dat weten de proefpersonen niet. Daarom is voor de proefpersonen in de controlegroep de studiehoeveelheid (niet het aantal woorden maar de inhoud van de stof) groter dan voor de proefpersonen in de experimentele groep, die een introductie bestuderen die nauw aansluit bij de inhoud van de studietekst.

Tegen de ongelijke studietijdmethode zijn in de literatuur nogal wat bezwaren aangetekend (Faw en Waller, 1976; Peeck, 1970, 1977). Deze bezwaren richten zich tegen het feit dat de proefpersonen in de experimentele groep langer kunnen studeren dan die in de controlegroep. Een gevolg daarvan is dat het niet mogelijk is te bepalen of een gevonden effect is veroorzaakt door de onafhankelijke variabele of door de extra tijd die nodig is om het extra materiaal te bestuderen.

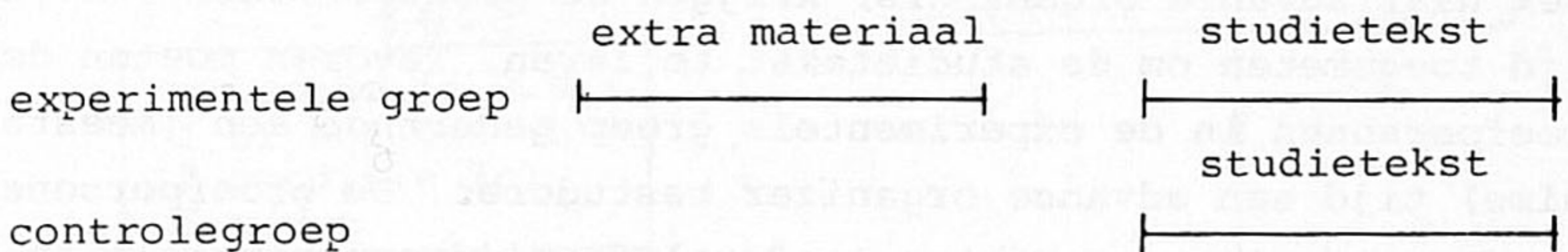
Carver (1972) voerde zelfs aan dat de totale studietijd de belangrijkste verklarende factor vormt voor de resultaten van experimenteel onderzoek naar het toevoegen van vragen.

Figuur 6.1 Schematische weergave van de vier studietijdcontrole-
methoden*

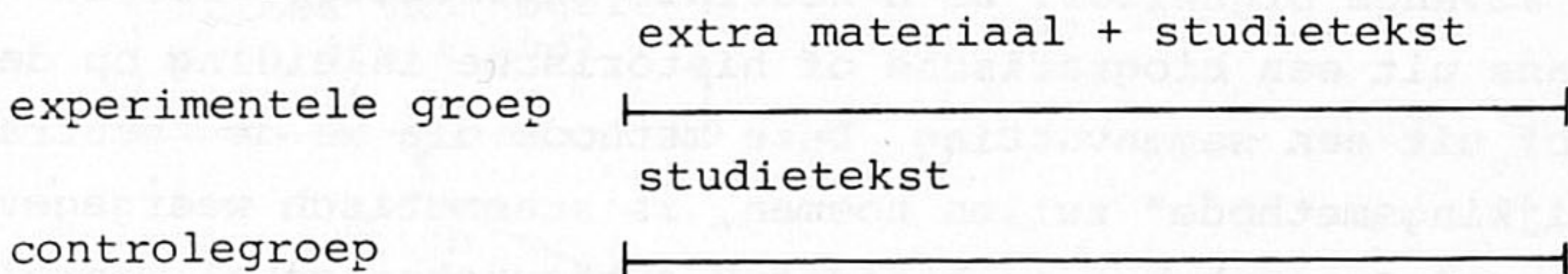
a) De neutrale vergelijkingsmethode



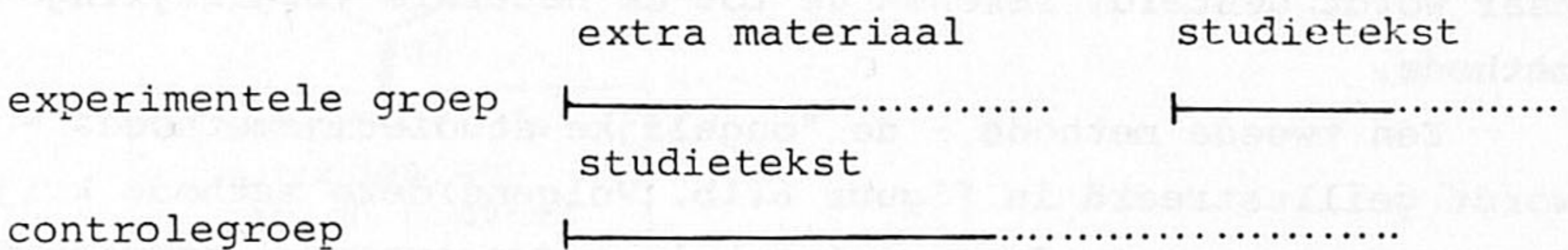
b) De ongelijke studietijdmethode



c) De efficiëntiemethode



d) De onbeperkte studietijdmethode



*De horizontale strepen stellen tijdsassen voor. De verticale streepjes geven begin en eind van een studiesessie weer en puntjes wijzen op het ontbreken van een tijdslimiet. Boven die tijdsassen is weergegeven wat in het betreffende tijdsinterval moest worden bestudeerd.

en het extra materiaal werken als men zelf wenselijk acht. Soms - maar helaas lang niet altijd - wordt gemeten hoeveel tijd de proefpersonen gebruiken en wordt een soort leerwinst per tijds-eenheidsscore berekend.

6.2 Voor- en nadelen van de verschillende studietijdcontrole-methoden

De neutrale vergelijkingsmethode biedt op een enigszins vreemde wijze controle op de studietijdverlenging in de experimentele conditie: De studietijd van de controleproefpersonen wordt kunstmatig verlengd, doordat zij bezig worden gehouden met een neutrale taak. Als er een effect van de onafhankelijke variabele wordt aangetoond, kan dan ook alleen maar worden geconcludeerd, dat het toegevoegde instructiemateriaal beter is dan (slecht) ander instructiemateriaal (Faw en Waller, 1976).

Een ernstig probleem dat aan de neutrale vergelijkingsmethode is verbonden (dat echter niet geldt voor post organizer studies) werd geconstateerd door Jaspers (1977) en heeft betrekking op de studiehoeveelheid. Over de neutrale (biografische, historische) introductie worden weliswaar geen vragen gesteld, maar dat weten de proefpersonen niet. Daarom is voor de proefpersonen in de controlegroep de studiehoeveelheid (niet het aantal woorden maar de inhoud van de stof) groter dan voor de proefpersonen in de experimentele groep, die een introductie bestuderen die nauw aansluit bij de inhoud van de studietekst.

Tegen de ongelijke studietijdmethode zijn in de literatuur nogal wat bezwaren aangetekend (Faw en Waller, 1976; Peeck, 1970, 1977). Deze bezwaren richten zich tegen het feit dat de proefpersonen in de experimentele groep langer kunnen studeren dan die in de controlegroep. Een gevolg daarvan is dat het niet mogelijk is te bepalen of een gevonden effect is veroorzaakt door de onafhankelijke variabele of door de extra tijd die nodig is om het extra materiaal te bestuderen.

Carver (1972) voerde zelfs aan dat de totale studietijd de belangrijkste verklarende factor vormt voor de resultaten van experimenteel onderzoek naar het toevoegen van vragen.

Hoewel er in laboratoriumonderzoek van gepaarde associaties wel enige steun is gevonden voor deze "total-time-hypothesis" (Bugelski, 1962), is zij beperkt en zijn de gegevens niet zonder meer generaliseerbaar naar andere soorten van leren (Rothkopf, 1974).

Rothkopf (op.cit.) heeft verder ter verdediging van de ongelijke studietijdmethode aangevoerd, dat het in het onderwijs alleen maar gaat om de "total amount learned" en dat de toename van de studietijd meestal zo gering is dat deze buiten beschouwing kan blijven. Faw en Waller (1976) reageerden daar echter weer als volgt op: "...nevertheless in the long haul, techniques which increase learning simply at the expense of additional time are more of a disservice than an aid to students, since the totality of useful (or interesting) knowledge to be acquired is so extensive and the time available to accomplish the task is so limited. On this basis, it is maintained that efficiency of learning deserves more research attention in future investigation of prose learning." (p. 712).

Enkele regels eerder hadden dezelfde auteurs echter beweerd: "It can safely be said that effects of study time on performance are large enough to demand our research attention but not so large that nothing else of importance remains to be understood."

Het lijkt verstandig niet alle aandacht te gaan richten op efficiëntiescores en alleen de efficiëntiemethode nog maar te gaan gebruiken. Zeker in de beginfasen van onderzoek naar een nieuwe vraagstelling verdient het de voorkeur, dat eerst maar eens wordt aangetoond dat het nieuwe extra materiaal een effect heeft op studiestatistiek. Pas in de latere fasen van onderzoek zou dan de efficiëntievraag aan de orde moeten komen. Een belangrijk voordeel van de ongelijke studiemethode is dat de studietijdverlenging kan worden afgewogen tegen de prestatieverhoging. Of het extra materiaal vervolgens zou moeten worden gebruikt, hangt af van vragen als: "Hoeveel tijd is er beschikbaar?", "Wat doen leerlingen in die tijd?", "Hoe is de verhouding tussen tijdsverlenging en prestatieverhoging?".

De efficiëntiemethode die is bedacht door Peeck (1970) wordt door Faw en Waller (1976) sterk aangeraden vanwege de controle over de studietijd. Immers de controlegroep kan de extra tijd, die de experimentele groep aan het extra materiaal besteedt, aan de studietekst op zich besteden. Toch kleven ook aan deze methode nadelen, zoals ook bleek uit een onderzoek van Peeck (1977). De proefpersonen in één van de controlegroepen kregen tien minuten extra studietijd voor het bestuderen van de studietekst. Toch presteerden zij niet meer (zelfs iets minder) dan de proefpersonen in een andere controlegroep die in die tien minuten een neutrale introductie bestudeerden. Kennelijk leidde de extra beschikbare studietijd niet tot verhoging van de bestede studietijden (dit bleek ook uit een vragenlijst).

De onbeperkte studietijdmethode heeft als grootste voordeel dat efficiëntiescores kunnen worden berekend (Faw en Waller, 1976): Hoeveel wordt er bij verschillende methoden per tijdseenheid geleerd? Bezwaren tegen deze studietijdcontrole-methode richten zich mijns inziens vooral op het stopcriterium dat proefpersonen hanteren. Houden ze op met studeren op grond van een tijds criterium, een herhalingscriterium of een beheersingscriterium, of houden ze misschien op omdat ze geen zin meer hebben of omdat anderen al klaar zijn?

Theoretisch zouden proefpersonen door moeten studeren tot dat ze de stof beheersen (beheersingscriterium van 100 procent) en zo gaat/ging het ook in het laboratoriumonderzoek (Underwood, 1964). Uit het feit dat er in verschillende tekstonderzoekingen nog voldoende spreiding in prestaties overbleef en het zinvol was efficiëntiescores te berekenen, kan worden afgeleid dat proefpersonen een ander dan het 100 procent beheersingscriterium hanteren.

Maar welk criterium hanteren ze dan wel, en hanteert men in verschillende condities dezelfde criteria? Het laatste zal vaak niet het geval zijn: Extra materiaal kan proefpersonen er bijvoorbeeld toe brengen meer tijd te investeren (Peeck, 1977) of minder vaak te herhalen. Opgemerkt moet worden, dat veel onderzoekers in het verleden de onbeperkte studietijdmethode gebruikten zonder de studietijden te meten (Faw en Waller, 1976).

6.3 Indirecte studietijdinvloeden

Een onderzoek naar verschillende oefenpatronen dat in 1978 in samenwerking met een doctoraalstudent onderwijspsychologie werd verricht (Biskop, 1978), bracht een tweetal indirecte studietijdinvloeden aan het licht, ^{dat} moet worden onderscheiden van de in de literatuur aangetroffen en in het bovenstaande besproken ~~methode der~~ directe studietijdverlenging. Conclusies uit het onderzoek van Biskop waren onder andere:

- a) Het vastleggen van een bepaalde studietijd door de onderzoeker garandeert niet dat al deze tijd ook door de proefpersonen wordt besteed. Een aanzienlijk deel van de door de onderzoeker toegestane studietijd werd door een groot deel van de proefpersonen niet gebruikt. Peeck (1977) introduceerde naar aanleiding van een soortgelijk onderzoeksresultaat het onderscheid tussen de nominale studietijd (die door de onderzoeker beschikbaar wordt gesteld) en de effectieve studietijden (die daadwerkelijk door de proefpersonen worden besteed). Ook Hiller (1974) trof verschillen aan tussen de nominale studietijd en de effectief bestede studietijden. In het tekstbestuderingsonderzoek is het gebruikelijk ruime nominale studietijden ter beschikking te stellen (zie de kritiek hierop van Underwood, 1964). Verschillen tussen de nominale studietijd en effectieve studietijden zullen dan ook eerder regel dan uitzondering zijn.
- b) In groepen die onder verschillende condities leren, kunnen verschillende gemiddelde effectieve studietijden optreden ondanks de gelijke nominale studietijden. In ons onderzoek (Biskop, 1978) besteedde de experimentele groep gemiddeld bijvoorbeeld bijna twee keer zo veel tijd aan de studietekst als de controlegroep, terwijl toch de nominale studietijden gelijk waren. Kennelijk kunnen onafhankelijke variabelen niet alleen invloed uitoefenen op de studieprestaties maar ook op de studietijdsbesteding.

Ook in andere onderzoekingen werd aangetoond, dat er onder invloed van de experimentele variabelen verschillen in gemiddelde effectieve studietijden kunnen optreden. Zo toonde Wolters (1975) bijvoorbeeld aan, dat proef-

personen meer tijd besteedden aan de studietekst wanneer ze vooraf vragen hadden gekregen dan wanneer ze (in de controleconditie) niet de beschikking hadden over vragen. Evenzo lieten Rickards en August (1975) zien, dat bepaalde vormen van onderstrepen langere effectieve studietijden tot gevolg hadden dan andere vormen. Ook uit het eerder vermelde onderzoek van Peeck (1977) bleek, dat proefpersonen onder sommige condities langer studeren dan onder andere.

Waarom zouden proefpersonen eerder ophouden met lezen en studeren dan is toegestaan (terwijl toch de studieprestaties zelden perfect kunnen worden genoemd)? Afgezien van factoren als vermoeidheid en geen zin meer hebben, zijn er mijns inziens drie mogelijke redenen om de studie voortijdig te staken:

- a) Omdat men vindt dat men voldoende tijd aan de stof besteed heeft: Men staakt dan de studie op grond van een tijds criterium.
- b) Omdat men vindt dat men de stof voldoende vaak herhaald heeft: Men staakt dan de studie op grond van een herhalingscriterium.
- c) Omdat men vindt dat men de stof in voldoende mate beheerst (bijvoorbeeld om de verwachte toetsvragen te beantwoorden): Men staakt dan de studie op grond van een (intern) beheersingscriterium.

Het laatste criterium werd ook door Peeck genoemd: "...subjects will spend a certain period of time on the text until some satiation point or some subjective level of familiarity is reached; this level probably corresponds to the anticipated degree of mastery required..." (1977, p. 206).

Als deze veronderstelling van Peeck juist is (en ook als de veronderstelling van het herhalingscriterium juist is), dan volgt daaruit dat het vaak erg moeilijk zal zijn effecten (in termen van studieprestaties) aan te tonen. Als proefpersonen zolang studeren tot ze een "subjective level of familiarity" of een herhalingscriterium hebben bereikt, dan kunnen allerlei didactische hulpmiddelen (als vragen, advance organizers en metaforen) ertoe leiden dat deze beheersings- of herhalingscriteria sneller worden bereikt. Met als gevolg dat er wel

kortere (effectieve) studietijden resulteren, maar dat van hogere studieprestaties geen sprake zal zijn.

Wellicht hebben we hier een verklaring gevonden voor sommige tegenvallende resultaten in onderzoek naar pre-instructievariabelen: Doordat de nominale studietijden ruim waren, heeft men enerzijds de proefpersonen in de gelegenheid gesteld in alle condities tot het persoonlijk beheersingscriterium te studeren. Anderzijds onttrokken de verschillen tussen de effectieve studietijden zich aan de waarneming van de onderzoeker die zich concentreerde op de gelijke nominale studietijden.

Conclusie: Indien men een prestatieverschil tussen experimentele en controlegroep wil aantonen, dan moet men niet toestaan dat proefpersonen kunnen doorstuderen totdat ze hun persoonlijk beheersingscriterium of persoonlijk herhalingscriterium toepassen. Hoe dat zou kunnen worden bereikt, wordt verder uiteengezet in 6.5.

Impliciet werden in het voorgaande twee indirecte studietijdinvloeden onderscheiden:

- a) Onder invloed van het extra studiemateriaal treedt verkorting op van de studietijden die door de proefpersonen aan de tekst wordt besteed, doordat de persoonlijke beheersingscriteria en/of de herhalingscriteria eerder worden toegepast.
- b) Onder invloed van extra studiemateriaal treedt verlenging op van de studietijd die door de proefpersonen aan de tekst wordt besteed, doordat ze op een andere wijze gaan studeren (bijvoorbeeld doordat ze de vooraf gegeven vragen proberen te beantwoorden; doordat het onderstrepen leidt tot een intensievere bestudering, of doordat metaforen moeten worden vergeleken met de lesstof).

In hoeverre spelen deze twee indirecte studietijdinvloeden nu een (ongecontroleerde) rol in de eerder besproken studietijdcontrolemethoden? De beantwoording van deze vraag is wat betreft de eerste drie methoden afhankelijk van de hoeveelheid (nominaal) beschikbaar gestelde studietijd. Is deze nominale tijd krap, dan zullen de effectieve studietijden niet noemenswaardig afwijken van de nominale tijd. De indirecte studietijdinvloeden spelen dan geen rol.

Overigens wordt men dan wel geconfronteerd met andere problemen: De spreiding in studieprestaties kan te gering worden (bodemeffect), de interpretatie kan bemoeilijkt worden doordat niet alle proefpersonen klaar zijn met de bestudering van de tekst (zie Van Hout-Wolters, 1980), of doordat de stof in verschillende groepen niet evenveel wordt herhaald.

Daarnaast is het ook voorstelbaar, dat extra materiaal soms alleen maar tot zijn recht kan komen als de effectieve studietijden kunnen worden verlengd. Bij het ontbreken van verschillen tussen groepen die onder verschillende condities werken, blijft, wanneer de studietijden te krap worden, de mogelijkheid open dat wel verschillen zullen optreden bij ruimere studietijden.

Bij ruime nominale studietijden kunnen de beide indirecte studietijdinvloeden een ongecontroleerde invloed uitoefenen op de resultaten bij de eerste drie studietijdcontrolemethoden: De neutrale vergelijkingsmethode, de ongelijke studietijdmethode en de efficiëntiemethode. De indirecte studietijdinvloeden kan men natuurlijk meten door na te gaan na hoeveel tijd de proefpersonen de studie staken. Zijn er geen verschillen tussen de gemiddelde effectieve studietijden dan kan men gerust zijn: De indirecte studietijdinvloeden speelden geen rol of hebben elkaars werking geneutraliseerd (tegenover de verlenging stond een verkorting). Zijn er zulke verschillen echter wel, dan zijn de resultaten vaak oninterpreteerbaar doordat de effectieve studietijden bij een deel van de proefpersonen langer zouden zijn geweest als daaraan door de nominale studietijd geen grens was gesteld.

Ook door middel van de onbeperkte studietijdmethode kunnen de beide indirecte studietijdinvloeden niet worden gecontroleerd, maar hun gezamenlijk effect kan men in ieder geval betrouwbaar meten door te registreren hoe lang de proefpersonen nodig hebben voor het extra materiaal, en hoeveel tijd ze besteden aan de tekst op zich.

6.4 Het kiezen van een studietijdcontrole methode: Vijf verschillende vragen

Bij het kiezen van een studietijdcontrole methode moet men rekening houden met de besproken voor- en nadelen van de verschillende methoden. Daarnaast is deze keuze echter afhankelijk van de soort van vragen die men in het onderzoek wil trachten te beantwoorden. De volgende onderzoeksvragen kunnen worden onderscheiden:

- a) Heeft het extra materiaal een effect op studieprestaties? (vergelijk Rothkopf, 1974).
- b) Werkt de ene introductie beter dan de andere? (vergelijk Ausubel et al., 1978; Van Hout-Wolters, 1980).
- c) Heeft het extra materiaal - gegeven een bepaalde vaste (nominale) studietijd - effect op studieprestaties? (vergelijk Peeck, 1970; Carver, 1972).
- d) Zorgt extra materiaal voor efficiënter leren?, of ook: Heeft extra materiaal naast een kwantitatief effect (via studietijdverlenging) ook een kwalitatief effect (via gunstige beïnvloeding van leerprocessen)? (vergelijk Faw en Waller, 1976; Peeck, 1970).
- e) Welke invloeden heeft het extra materiaal op de studietijdbesteding? (vergelijk Wolters, 1975; vergelijk ook het onderzoek in dit proefschrift).

In tabel 6.1 is een overzicht opgenomen van de studietijdcontrole methoden en de onderzoeksvragen. In deze tabel is per combinatie van vraag en methode, door middel van cijfers, aangegeven waarom de betreffende methode niet in aanmerking komt om de vraag te beantwoorden. Bij de overige combinaties zijn de mogelijke interpretatieproblemen door middel van letters opgesomd. De volgende betekenissen moeten worden toegekend aan de cijfercodes 1 tot en met 7:

- (1) Met deze methode kan hoogstens worden nagegaan of het extra materiaal beter is dan (slecht) ander materiaal (Faw en Waller, 1976).
- (2) Deze methoden zijn speciaal geschikt voor het vergelijken van een groep die met extra materiaal werkt met een groep die zonder extra materiaal werkt en derhalve niet geschikt

zijn voor een vergelijking van de effecten van de verschillende soorten extra materiaal.

- (3) De nominale studietijden zijn niet gelijk voor groepen die onder verschillende condities werken, terwijl de vraagstelling gelijke tijden veronderstelt.
- (4) Efficiëntiemetingen en tijdmetingen zijn onbetrouwbaar, omdat de effectieve studietijden voor een deel van de proefpersonen zijn begrensd door de nominale studietijd en voor een ander deel niet.
- (5) De studietijden worden gecontroleerd en de indirecte invloeden kunnen dus niet worden gemeten.
- (6) Het studieprestatie-effect kan worden verbloemd of opgeheven door de studietijdcontrole, terwijl men bij deze vraag niet geïnteresseerd is in studietijden.
- (7) De studieprestaties zouden kunnen worden gedeeld door de nominale studietijd, maar dat levert mijns inziens een zeer onzuivere maat op.

In de gevallen waarin in de tabel één of meer van deze cijfercodes voorkomen, is de betreffende methode niet geschikt om de onderzoeksvraag te beantwoorden. Ook in de overige gevallen kunnen echter nog interpretatieproblemen ontstaan. Deze zijn in de tabel door middel van de volgende lettercodes weergegeven:

- (A) Als er geen significante prestatieverschillen worden gevonden, blijft nog de mogelijkheid open dat deze wel zullen worden gevonden bij langere studietijden, bijvoorbeeld omdat de indirecte verlenging van de studietijd onder invloed van het extra materiaal een noodzakelijke voorwaarde vormt voor het optreden van een effect.
- (B) Als er geen significante prestatieverschillen worden gevonden, kan zulks zijn veroorzaakt door een verkorting van de effectieve studietijden.
- (C) Als er significante prestatieverschillen worden gevonden, kunnen deze zijn veroorzaakt door een (indirecte) verlenging van de effectieve studietijden.
- (D) Op welke gronden stoppen de proefpersonen de studie? Zijn deze in de verschillende groepen hetzelfde? Hoe moeten even-

tuele tijdsverschillen worden afgewogen tegen prestatiever-
schillen?

- (E) De resultaten zijn afhankelijk van de (willekeurig gekozen) (krappe) nominale studietijd. Vaak zullen effecten alleen in de resultaten tot uiting komen als de directe verlenging van de studietijd (doordat het extra materiaal moet worden bestudeerd) een kans krijgt.

Tabel 6.1 Overzicht van de verschillende studietijdcontrole-
methoden (I tot en met IV) en de onderscheiden onder-
zoeksvragen (a tot en met e)

Vraag	a	b	c	d	e
Methode van studietijdcontrole					
I Neutrale vergelijkingsmethode					
ruime studietijd	1	A, B, C	1	1, 4	1, 4
krappe studietijd	1	A	1	1, 5	1, 5
II Ongelijke studietijdmethode					
ruime studietijd	A, B*	2	3	4	4
krappe studietijd	A	2	3	7	5
III Efficiëntiemethode					
ruime studietijd	6	2	B*	4	4
krappe studietijd	6	2	*	A, E	5
IV Onbeperkte studietijdmethode	D	D	3	D	D

*Problemen A en C kunnen hier wel een rol spelen, maar dat is in het licht van de vraagstelling niet relevant.

De interpretatieproblemen A tot en met E zijn - in tegenstelling tot de eerder gesignaleerde problemen 1 tot en met 6 - in bepaalde mate oplosbaar (zie 6.5).

Het probleem van de keuze van een studietijdcontrole-
methode is nu verschoven naar de keuze van een onderzoeksvraag-
stelling. Immers daar hangt het vanaf welke studietijdcontrole-

methode het meest geschikt is. Alle genoemde vragen zijn mijns inziens interessant, al hebben sommige eerder een praktische betekenis (b en c) en andere eerder een theoretische (e). Belangrijkste vraag lijkt mij vraag d. Want zowel vanuit een theoretisch als vanuit een praktisch perspectief zijn prestatiever- schillen die uitsluitend zijn veroorzaakt door studietijdver- lenging op de lange termijn nauwelijks interessant (zie ook Faw en Waller, 1976).

Waar het om gaat is of de eventuele kwalitatieve effecten, die worden veroorzaakt doordat het extra materiaal andere leer- processen induceert, opwegen tegen de verlenging van de studie- tijd die daarmee gepaard gaat. Ook bij vraag c kan men deze af- weging maken, maar dan is men gebonden aan een bepaalde vaste studietijd die men meestal tamelijk willekeurig moet kiezen. Vraagstelling c is dan ook voornamelijk interessant wanneer er (bijvoorbeeld vanuit de praktijk) ideeën kunnen worden ge- leverd over maximale studietijden die aan het te bestuderen onderwerp worden besteed. Evenzo is vraagstelling b (welke introductie?) voornamelijk interessant wanneer een nieuw soort van introductie kan worden vergeleken met een bestaande, veel gebruikte soort van introductie.

Vraag a is vooral interessant wanneer tijd geen rol speelt of wanneer de studietijdverlenging erg gering is. Vraag a komt mijns inziens met name in beginfasen van een onderzoeksproject in aanmerking. Hetzelfde geldt voor vraag e (effect op studie- tijden).

Voordat we ons kunnen wagen aan vraag d (efficiëntie) dienen we - zeker gezien de complexiteit van de studietijd- invloeden (zie 6.3) - meer informatie te verzamelen over de verhoudingen tussen de drie studietijdinvloeden (vraag e).

De verschillende vragen zijn alle in bepaalde fasen van een project van belang en het is mijns inziens dan ook gewenst dat alle aan bod komen. Wellicht is het verstandig om te be- ginnen met de vragen a (is er wel een effect) en e (invloeden op tijdsbesteding) en pas daarna aan de andere vragen aandacht te besteden.

6.5 Oplossingen voor de gesignaleerde problemen

In de vorige paragraaf werd gesteld, dat de interpretatieproblemen A tot en met E in tabel 6.1 tenminste gedeeltelijk oplosbaar zijn. In deze paragraaf zal worden nagegaan hoe elk van deze problemen kan worden voorkomen door passende controlemaatregelen te treffen. Daarbij zullen enkele nieuwe studietijdcontrolemethoden worden voorgesteld.

Probleem A (misschien is er bij een langere studietijd wel een effect, namelijk als de indirecte studietijdverlenging op kan treden) kan men oplossen door groepen met verschillende nominale studietijden op te nemen. Is bij de groep met langere studietijd het effect groter dan bij de groep met een kortere studietijd dan is het duidelijk dat de langere studietijd effect heeft.

Om de problemen B (het ontbreken van een effect komt door verkorting van de effectieve studietijden) en C (een gevonden effect kan zijn veroorzaakt door een indirecte verlenging van de studietijd) op te lossen, moeten de effectieve studietijden worden gemeten. Als er geen verschillen in gemiddelde effectieve studietijden tussen de groepen zijn, dan vormen B en C geen probleem. Als er wel verschillen in gemiddelde effectieve studietijden zijn, weet men dat B en/of C mogelijk een rol spelen. Dan is echter verder onderzoek nodig met gebruikmaking van een andere methode.

Om probleem D (wat zijn de stopcriteria?) op te lossen, moet men (via instructies) de stopcriteria duidelijker maken en uniformeren. Uit laboratoriumonderzoek (zie Underwood, 1964) weten we, dat er drie methoden zijn om stopcriteria te manipuleren: Door een prestatie criterium te stellen, door een herhalingscriterium te stellen of door een tijds criterium te stellen. Een prestatie criterium kan inhouden, dat de studie pas mag worden afgebroken zodra bijvoorbeeld alle items één keer goed zijn onthouden. Een herhalingscriterium wordt gerealiseerd doordat de items een constant aantal keren worden gepresenteerd (één, twee of drie keer bijvoorbeeld). Een tijds criterium stelt men door een krappe studietijd beschikbaar te stellen waarbinnen niemand alle items kan leren.

Het stellen van een tijds criterium voor tekstbestudering vormt natuurlijk een oplossing voor het probleem D, maar dan is men van de onbeperkte studietijdmethode weer terechtgekomen bij de andere drie studietijdcontrolemethoden uit tabel 6.1 met de daarbij behorende problemen. Een oplossing voor probleem D kan daarom beter worden gezocht bij prestatiecriteria of herhalingscriteria. Men zou een prestatie criterium kunnen stellen door bijvoorbeeld de instructie te geven zólang door te studeren totdat 80 procent van de - vooraf bekende - vragen goed beantwoord kan worden, maar het is zeer de vraag of proefpersonen in staat zijn een dergelijke schatting van de eigen studieprestaties te maken.

Een betere oplossing lijkt mij om een herhalingscriterium te stellen: De proefpersonen krijgen dan de opdracht de studietekst één of meer keren door te nemen. We zullen deze variant op de onbeperkte studietijdmethode de vast-aantal-trials methode noemen. Andere oplossingen voor probleem D kunnen worden gevonden door het bestuderen van teksten meer te laten lijken op de natuurlijke studiesituaties van de proefpersonen. Men zou instructies kunnen geven om de tekst net zo goed te bestuderen als tentamenstof of huiswerk (ecologisch beheersingscriterium) of om net zo veel tijd aan de tekst te besteden als men thuis of op school aan een vergelijkbare tekst zou besteden (ecologisch tijds criterium) of om de tekst zo vaak te herhalen als men gewend is bij zelfstudie of voor een tentamen (ecologisch herhalingscriterium).

Welke van deze voorstellen men ook opvolgt telkens zullen zowel studieprestaties als studietijden moeten worden gemeten. Een probleem vormt dan de combinatie van deze twee afhankelijke variabelen: Moet men ze a) één voor één analyseren of b) moeten er efficiëntiescores worden berekend (leerwinst per tijdseenheid) of c) moeten de studietijden als covariabele worden opgenomen in de analyse? In deze dissertatie is gekozen voor een combinatie van a) en c).

Probleem E (de directe verlenging van de studietijd vormt misschien een noodzakelijke voorwaarde voor het optreden van een effect) kan alleen een rol spelen bij gebruik van de efficiëntiemethode.

Binnen deze methode is het probleem onoplosbaar (het speelt overigens alleen wanneer er geen significante verschillen worden gevonden bij vraagstelling d). Een oplossing wordt echter wel bereikt via een gecombineerde efficiëntie- en ongelijke studietijdmethode, waarbij twee controlegroepen worden opgenomen (zie Peeck, 1970, 1977). Eén controlegroep krijgt voor de tekst op zich evenveel tijd als de experimentele groep, en de andere controlegroep krijgt evenveel tijd als de experimentele groep krijgt voor de bestudering van het extra materiaal en de studietekst samen.

6.6 De studietijdcontrolemethoden in deze dissertatie

In het voorgaande passeerden de verschillende studietijdcontrolemethoden de revue, werden hun voor- en nadelen opgesomd met speciale aandacht voor de indirecte studietijdinvoeden en werd aangegeven welke methode het beste past bij welke vraagstelling. Tevens werden in de laatste paragraaf oplossingen voorgesteld voor enkele resterende interpretatieproblemen. Rest ons nog te beargumenteren welke methoden en vraagstellingen inclusief bijbehorende controlemaatregelen in deze dissertatie zijn gekozen.

In 6.4 werd reeds voorgesteld om in de beginfasen van een onderzoeksproject de vragen a (is er los van de studietijden een effect?) en e (welke invloed heeft het extra materiaal op de studietijdsbesteding?) te onderzoeken, en pas daarna de vragen d (is het leren met het extra materiaal efficiënt?), c (werkt het extra materiaal ook gegeven een vaste nominale studietijd?) en b (welke soort introductie werkt het best?). Deze volgorde is in onze experimenten aangehouden: In deel III worden eerst twee experimenten gerapporteerd waarin de vragen a en e aan bod komen en vervolgens wordt een experiment gepresenteerd waarbij de vragen c en d zijn betrokken. In de laatste drie experimenten (hoofdstuk 8) staat vraag b centraal, maar komen ook de vier andere vragen nog aan de orde.

Uit tabel 6.1 kan worden afgelezen dat voor vraagstelling a de ongelijke studietijdmethode of de onbeperkte studietijdmethode in aanmerking komt, terwijl voor vraagstelling e alleen de onbeperkte studietijdmethode is geïndiceerd. In de eerste

twee experimenten is gekozen voor de onbeperkte studietijdmethode omdat dan zowel studieprestaties (vraag a) als studietijden (vraag e) kunnen worden geanalyseerd, terwijl er ook met betrekking tot vraag d reeds perspectieven aanwezig zijn (zie tabel 6.1). Probleem D (wat zijn de stopcriteria) is opgelost door een herhalingscriterium te stellen (de vast-aantal-trials-methode). Binnen deze methode zijn verschillende versies mogelijk (zie figuur 6.2): Men kan de tekst één, twee of drie keer laten bestuderen (of lezen); men kan tussen de verschillende lezingen door tijdsregistraties doen of alleen aan het eind; men kan tussendoor toetsen of alleen aan het eind; men kan de tijd voor het extra materiaal apart meten of inbegrijpen in de eerste lezing; men kan de verschillende lezingen in verschillende sessies laten doen of binnen één sessie. Daarbij moet natuurlijk worden bedacht, dat het "tussendoor toetsen" van invloed kan zijn op de volgende bestudering van de tekst.

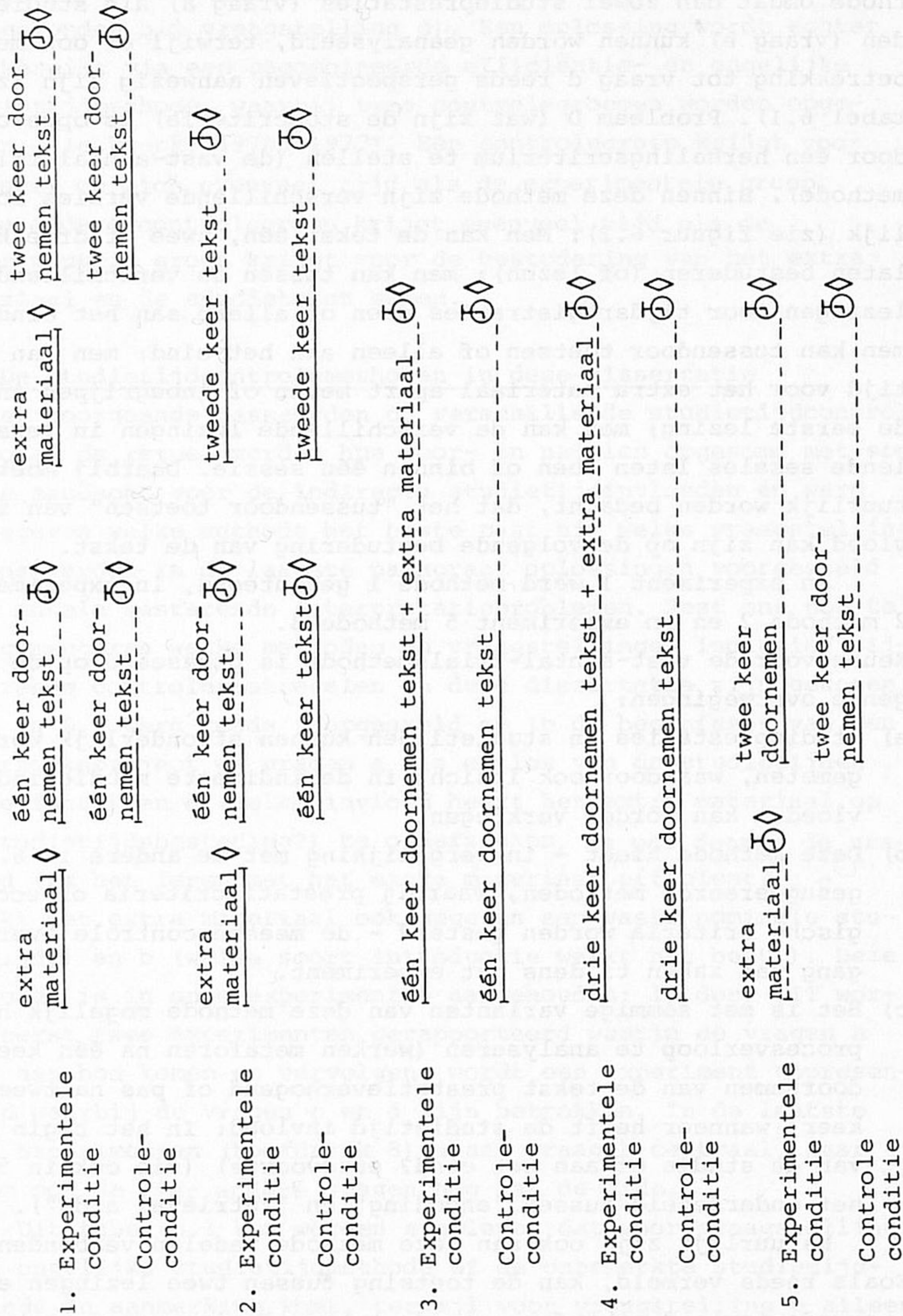
In experiment 1 werd methode 1 gehanteerd, in experiment 2 methode 2 en in experiment 5 methode 3.

Keuze voor de vast-aantal-trials-methode is gebaseerd op de volgende overwegingen:

- a) Studieprestaties en studietijden kunnen afzonderlijk worden gemeten, waardoor ook inzicht in de indirecte studietijdinvloeden kan worden verkregen.
- b) Deze methode biedt - in vergelijking met de andere in 6.5 gesuggereerde methoden, waarbij prestatiecriteria of ecologische criteria worden gesteld - de meeste controle over de gang van zaken tijdens het experiment.
- c) Het is met sommige varianten van deze methode mogelijk het procesverloop te analyseren (werken metaforen na één keer doornemen van de tekst prestatieverhogend of pas na twee keer; wanneer heeft de studietijd invloed: In het begin van de studie of aan het eind? enzovoorts) (zie ook in 5.2.1 het onderscheid tussen "encoding" en "retrieval aids").

Natuurlijk zijn ook aan deze methode nadelen verbonden. Zoals reeds vermeld, kan de toetsing tussen twee lezingen een storende invloed uitoefenen op de tweede lezing van de tekst. Dit bezwaar geldt echter alleen voor de tweede en verdere le-

Figuur 6.2 Enkele mogelijke versies van de vast-aantal-trialsmethode



Voor de betekenis van de strepen en puntjes zie figuur 6.1: ◊ = tijdsmeting, ◊ = toetsafname

zingen van de tekst. Een belangrijk bezwaar tegen de vast-aantal-trials methode vormt de ecologische validiteit. Zelden zal men leerlingen opdragen een tekst één, twee of drie keer door te nemen, veeleer is men geïnteresseerd in de studieprestaties. Wat dat betreft zou een methode met een ecologisch criterium (zie 6.5) wellicht de voorkeur verdienen.

In de latere experimenten (3 en 4) wordt de efficiëntiemethode gebruikt, omdat daar vraagstellingen c en d (efficiëntievraagstellingen) centraal staan. Er wordt daarbij gekozen voor een krappe nominale studietijd, zodat de problemen B en C geen rol kunnen spelen (zie tabel 6.1). De keuze van een studietijd is echter problematisch: Of een studietijd krap of ruim is, blijkt uiteindelijk pas tijdens het experiment. De problemen A (bij vraagstelling c) en E (bij vraagstelling d) kunnen de interpretatie van de resultaten bemoeilijken als er geen significante verschillen worden aangetroffen en de nominale studietijd krap bleek te zijn. Wanneer de nominale studietijd onverhoopt toch (te) ruim blijkt, kan ten aanzien van vraag c (effect bij vaste studietijd?) probleem B (geen effect door verkorting van de effectieve studietijden) een rol spelen, terwijl vraag d (de efficiëntievraag) onbeantwoordbaar is geworden door probleem 4 (de effectieve tijdsmetingen zijn onbetrouwbaar geworden doordat de nominale studietijd een plafond aangeeft voor de effectieve studietijden).

In experiment 6, tenslotte, wordt de ongelijke studietijdmethode gebruikt. Hier vindt namelijk geen tekstbestudering door de leerlingen plaats, maar wordt een mondelinge presentatie van de lesstof gegeven. De andere studietijdmethoden zijn in geval van mondelinge presentatie moeilijk uitvoerbaar.

III Onderzoeksdeel

Hoofdstuk 7 Drie experimenten: Invloeden van metaforen op studieprestaties en studietijden bij verschillende leerstijlen

In dit en het volgende hoofdstuk wordt een zestal studies gerapporteerd rondom de in hoofdstuk 1 geformuleerde algemene vraagstelling: "Voor welke typen van leerlingen vergemakkelijken welke soorten van metaforen het begrijpen en onthouden (op korte en lange termijn) van lesstof". De drie experimenten in dit hoofdstuk richtten zich op invloeden van metaforen op zich, terwijl de drie experimenten in het volgende hoofdstuk daarnaast ook waren gericht op invloeden van verschillende soorten van metaforen.

Experiment 1 richtte zich - met behulp van de vast-aantal-trials methode (zie 6.6) - op invloeden van metaforen op studieprestaties (vraagstelling a in 6.5) en op studietijden (vraagstelling e in 6.5) bij proefpersonen met verschillende door Pask (1976c) onderscheiden leerstijlen (operation learning versus comprehension learning). In experiment 2 werd het eerste onderzoek gerepliceerd bij een heterogener samengestelde steekproef (MAVO- en VWO-leerlingen). In experiment 3 werd, met hetzelfde stimulusmateriaal en vergelijkbare steekproef, de efficiëntie van het leren via metaforen (vraagstelling c in 6.5), bij proefpersonen met verschillende door Boekaerts (1978) onderscheiden leerstijlen (visualizers versus verbalizers), onderzocht met behulp van de efficiëntiemethode.

7.1 Experiment 1: Operation learning versus comprehension learning¹⁾

7.1.1 Vraagstellingen en hypothesen

Op basis van de (theoretische) functies die metaforen in een onderwijsleerproces kunnen vervullen (hoofdstuk 3) en op basis van het voorgaand onderzoek (hoofdstuk 4) werd verwacht, dat

¹⁾ Dit onderzoek werd uitgevoerd door P. Domen en H. Kreutzer in het kader van hun doctoraalstudie onderwijspsychologie (zie Domen en Kreutzer, 1979).

metaforen een faciliterend effect zouden hebben op het begrijpen en onthouden van relatief onbekende leerstof met een moeilijk te doorgronden structuur. Dit effect zou zich zeker moeten uiten in de studieprestaties wanneer niet wordt gecontroleerd voor de extra tijd die men aan de metaforen kan besteden (hetgeen het geval is bij gebruik van de vast-aantal-trials methode). Omdat metaforen volgens Rummelhart en Ortony (1977) een belangrijke rol kunnen spelen bij het construeren van nieuwe schemata (zie hoofdstuk 3) zou vooral de eerste lezing of verwerking van een studietekst beter moeten verlopen. Evenzo werd verwacht dat de eerste lezing (verwerking) van een studietekst sneller zou geschieden wanneer via een metafoor de constructie van een nieuw schema werd vergemakkelijkt en de gelegenheid werd geboden de nieuwe informatie te koppelen aan een bestaand schema. Ook werd verwacht, dat de metaforen een lange termijn effect zouden hebben (drie weken na het drie keer doornemen van de lesstof).

Geen duidelijke verwachtingen bestonden er met betrekking tot de invloeden van de metaforen op studieprestaties en studietijden wanneer de lesstof drie keer werd doorgenomen: Zou de voorspelde voorsprong blijven bestaan, of zouden de proefpersonen in de controlegroep de opgelopen achterstand inhalen door langer over de tweede en derde lezing te doen dan de experimentele proefpersonen?

Het onderzoek werd opgezet volgens het Aptitude-Treatment-Interaction-paradigma. Via de correspondentie-analysestrategie (zie hoofdstuk 2) werd gezocht naar leerlingkenmerken die in het onderzoek zouden kunnen worden opgenomen. Centraal daarbij stond dat deze leerlingkenmerken moesten kunnen worden vertaald in onderliggende processen (vergelijk leerling-analyse) die in verband konden worden gebracht met moeilijkheden van de leertaak en functies van de onderwijsmaatregel (situatie-analyse).

Via een intuïtieve leertaakanalyse werd als één van de moeilijkheden van de in de experimenten 1 tot en met 3 te bestuderen leerstof aangemerkt: Het opbouwen van een adequate representatie van het netwerk van begrippen, en het blijven

vasthouden van de grote lijn in het systeem van relaties (vergelijk Simons en Lodewijks, 1979). Deze moeilijkheid correspondeerde met een van de in hoofdstuk 3 beschreven functies van metaforen: Het bieden van hulp bij de constructie van een nieuw schema doordat een formeel vergelijkbaar schema als grote lijn of als kapstok kan dienen.

Bij deze correspondentie tussen een leertaakmoeilijkheid en een situatiefunctie werd als corresponderend leerlingkenmerk, de leerstijl "operation versus comprehension learning" van Pask (1976c) gevonden. Een hoge score op comprehension learning wijst volgens Pask op een stijl van leren, waarbij vooral de grote lijn in de leerstof wordt opgespoord, terwijl de details pas later worden ingevuld. Een hoge score op operation learning daarentegen wijst erop, dat regels en details van buiten worden geleerd zonder dat er een grote lijn wordt ontdekt (zie verder 7.1.2). Metaforen zouden aansluiten bij, c.q. capitaliseren op, de leerstijl van comprehension learners en niet van operation learners. Zij zouden vooral het leren van eerstgenoemde vergemakkelijken, terwijl ze het leren van operation learners eerder zouden verstoren, omdat ze niet passen bij hun leerstijl.

In experiment 1 werden aldus de volgende hypothesen getoetst:

- 1) Als metaforen aan een relatief onbekende studietekst met een moeilijk te ontdekken grote lijn worden toegevoegd, wordt na één keer lezen meer van die tekst begrepen dan zonder metaforen.
- 2) Als aan zo'n tekst metaforen worden toegevoegd, zal de eerste lezing van die tekst sneller gaan dan zonder metaforen.
- 3) Als aan zo'n tekst metaforen worden toegevoegd, zal de lange termijnretentie hoger zijn dan wanneer dat niet het geval is.
- 4) De effecten van metaforen zijn afhankelijk van de leerstijl van leerlingen: Comprehension learners profiteren er meer van dan operation learners.
- 5) Welke invloeden hebben metaforen op het begrip van de leerstof en op de studietijden na het drie keer doorlezen van de studietekst?

7.1.2 Operation versus comprehension learning: Theorie en onderzoek

In het kader van theorievorming en onderzoek met betrekking tot conversaties tussen een student en een leerkracht of een computer, maakte Pask (1976a, 1976c) oorspronkelijk onderscheid tussen holisten en serialisten: "The holist has many goals and working topics under his aim topic; the serialist has one goal and working topic, which may be the aim topic"... "Evidence suggests that the holist is assimilating information from many topics in order to learn the "aim" topic, while the serialist moves on to another topic only when he is completely certain about the one he is currently studying"... "Holist students ask questions about broad relations and form hypotheses about generalisations. Serialists ask questions about much narrower relations and their hypotheses are specific." (Pask, 1976c, p. 130). Holisten zijn voortdurend op zoek naar analogieën met het gevolg dat ze wel eens analogieën zien die er helemaal niet zijn (globetrotting). Serialisten gaan voorbij aan valide analogieën en maken het zich daardoor moeilijker dan nodig (improvidence) (zie Entwistle, 1978).

Het onderscheid tussen holisten en serialisten werd zeer strikt gedefinieerd binnen de grenzen van computer gestuurde onderwijssystemen (conversaties). Deze tweedeling, die door via de apparatuur opgelegde restricties kunstmatig werd bereikt, beschouwde Pask (1976c) als een extreme manifestatie van meer fundamentele processen. Daarom noemde hij het onderscheid tussen holisten en serialisten een onderscheid naar twee leerstrategieën dat onder minder restrictieve omstandigheden samenhangt met het onderscheid tussen twee leerstijlen (operation versus comprehension learning). Operation learners zijn geneigd tot serialisme en comprehension learners tot holisme, maar of dat inderdaad zo is, hangt af van allerlei restricties.

De twee leerstijlen, die kunnen worden gemeten met de Spy Ring History Test (zie 7.1.3), worden door Pask (1976c) als volgt beschreven: "Comprehension learners readily pick up an overall picture of the subject matter, for example, redundancies in a taxonomic scheme or relations between distinguished

classes and recognise clearly where information can be obtained. These individuals are able to build descriptions of topics and to describe the relation between topics. Their cognitive repertoire includes effective, though individually distinctive, description building operations, although such learners may not be able to apply these operations to specific subject matter information (for example, to classify specimens) until the procedures underlying the concepts in question are specifically taught.

Left to their own devices, operation learners pick up rules, methods and details, but are often unaware of how or why they fit together. They have, at most a sparse mental picture of the material and their recall of the way they originally learned is guided by arbitrary number schemes or accidental features of the presentation. On the other hand if an operation learner is provided with a specific description (by external means) he assimilates procedures and builds concepts for isolated topics. His cognitive repertoire includes accessible or effective procedure building operations." (p. 133).

Naast de operation en de comprehension learners onderscheidde Pask ook nog de "versatiles", die zowel goed zijn in operation learning als in comprehension learning: "A student who is versatile is not prone to vacuous globetrotting: He does indeed build up descriptions of what may be known, by a rich use of analogical reasoning, but subjects the hypotheses to test and operationally verifies the validity of an analogy and the limits of its applicability. By the same token, the versatile student is less likely to show the improvidence of failing to see that one area of knowledge (which has been operationally mastered) is analogous to a second area (which is to be learned)." (Pask et al., 1977, p. 68).

In de eerste experimenten werd door Pask en zijn medewerkers (zie bijvoorbeeld Pask en Scott, 1972) aangetoond, dat in het kader van computer gestuurd onderwijs het afstemmen van leerstrategie en onderwijsprogramma op elkaar betere resultaten opleverde dan "mismatching". Dus holisten leerden bijna niets met een serialistisch programma en serialisten bijna niets met een holistisch programma, terwijl beide groepen goede presta-

ties leverden wanneer het onderwijsprogramma aansloot bij hun leerstrategie (zie voor een overzicht Pask, 1976c).

In een later onderzoek werden ook de leerstijlen (comprehension en operation learning) in minder restrictieve onderwijs-situaties gemeten en vergeleken met de in meer restrictieve situaties gemeten leerstrategieën. Er bleek een goede overeenstemming te bestaan tussen de leerstrategieën en de leerstijlen en tevens werd aangetoond, dat ook "mismatching" tussen leerstijl en instructiewijze het leren totaal verstoort (op.cit.).

In een laatste serie experimenten (op.cit.) werd aangetoond, dat via de Spy Ring History Test gemeten leerstijlen correspondeerden met typische fouten tijdens het leerproces (globetrotting en improvidence). Bovendien werd aangetoond, dat deze fouten bij de verschillende typen leerlingen op een theoretisch voorspelde wijze konden worden versterkt of verzwakt doordat veranderingen in het instructieprogramma werden aangebracht.

Verder onderzoek van Pask en zijn collega's richtte zich op de mogelijkheden om de leerstijlen te veranderen in de richting van "versatility" (Pask, 1976a) en op de constructvaliditeit en betrouwbaarheid van de Spy Ring History Test. In een recent voortgangsrapport (Pask et al., 1978) werden enigszins tegenvallende correlaties gerapporteerd tussen de Spy Ring History Test en een paralleltest, de Smugglers Test. Een ernstig probleem met de leerstijlen van Pask werd zeer recentelijk gesuggereerd door Laurillard (1979). Zij meent te hebben aangetoond, dat de leerstijlen van de context afhankelijk zijn. Verder onderzoek zal nog moeten uitwijzen of zij daarin gelijk heeft.

7.1.3 Methode

7.1.3.1 Proefpersonen

Aan het onderzoek namen in totaal 81 proefpersonen deel, afkomstig uit zes tweede klassen van vier MAVO-scholen te Tilburg, 38 jongens en 43 meisjes. De leeftijden van de proefpersonen varieerden van dertien tot vijftien jaar. De toewijzing aan de experimentele en de controlegroep geschiedde aselekt met dien

verstande dat - via een gestratificeerde steekproeftrekking - werd gezorgd voor gelijke aantallen comprehension en operation learners in de twee groepen. In de experimentele groep zaten oorspronkelijk 42 leerlingen en in de controlegroep 39. Door ziekte en dergelijke deden er uiteindelijk slechts 32 experimentele en 29 controleproefpersonen aan alle vier de sessies mee. Na afloop van het experiment werden de proefpersonen beloond met een cadeaubon.

7.1.3.2 Testmateriaal

a) De Spy Ring History Test

In deze test worden drie verschillende spionage-netwerken gepresenteerd (met betrekking tot de jaren 1985, 1986 en 1987). Deze netwerken worden weergegeven door middel van berichtenlijsten en door middel van pijlverbindingen (gerichte grafen) tussen de spionnen. Verder wordt er uitvoerige informatie verstrekt over de spionnen zelf, over de landen waarin ze wonen en over de verschillende rollen die de spionnen kunnen spelen. Na de bestudering van de achtergrondinformatie moeten de netwerken en de bijbehorende berichtenlijsten van de drie achtereenvolgende jaren worden bestudeerd totdat ze perfect kunnen worden gereproduceerd. Tot slot volgt dan een testboekje waarin allerlei vragen zijn opgenomen over de ligging van de landen, de ontwikkelingen in de netwerken (hoe zien netwerk en lijst eruit in 1988?), de rollen van de verschillende spionnen, en de lijsten en netwerken van de drie jaren. Belangrijk in verband met het onderscheid tussen comprehension en operation learning zijn de door de proefpersonen te geven voorspelling voor 1988 en de vraag of men liever eerst netwerken dan wel lijsten wil reproduceren.

Er werd gebruik gemaakt van een door G. van de Veer van de Vrije Universiteit van Amsterdam ontwikkelde vertaling van de tweede versie van de test van Pask (de eerste versie omvatte vijf achtereenvolgende jaren). Scoring vond plaats aan de hand van een door Pask et al. ontwikkelde scoringshandleiding. Conform de richtlijnen van Pask werden proefpersonen die te lage totaal-resultaten op de test be-

haalden (som van operation en comprehension learning kleiner dan 20 procent) buiten beschouwing gelaten. De afnameduur varieert tussen de anderhalf en twee en een half uur. De test is geschikt voor kinderen vanaf + veertien jaar (Pask et al., 1978).

b) De Hidden Figures Test

Bij deze test die ontwikkeld werd om veld(on)afhankelijkheid te meten, moet men ontdekken welke van vijf eenvoudige figuren in een ingewikkeld patroon verborgen is. De test bestaat uit twee delen met elk zestien patronen. Voor elk deel is tien minuten beschikbaar (speedtest). De test werd opgenomen als opvultraak in de controlegroep en ter validering van de Spy Ring History Test.

c) De DAT-analogieën-subtest

De subtest verbale analogieën uit de Differentiele Aanleg Test bestaat uit 50 items van het type ... staat tot water, als eten staat tot Voor elk van de open plaatsen moet de proefpersoon uit vier opgegeven begrippen het juiste kiezen. De testtijd bedraagt 30 minuten. Omdat deze test in eerste instantie als opvultraak en in tweede instantie ter validering van de Spy Ring Test werd opgenomen, werd hij in dit onderzoek in twee delen gesplitst (25 items in vijftien minuten).

7.1.3.3 Lesmateriaal

a) De natuurkundecursus

Er werd gebruik gemaakt van een korte schriftelijke cursus electriciteitsleer. Deze cursus die eerder in verschillende onderzoeken van de vakgroep onderwijspsychologie was gebruikt, bestond uit een aanpassing aan tweede klas MAVO-niveau van de "logische volgorde versie", die in het kader van het onderzoek van Lodewijks (1978) was ontwikkeld.

De cursus werd opgesplitst in drie delen. In het eerste boekje werden de begrippen stof, molecuul, atoom en electron behandeld. In het tweede boekje kwamen de begrip-

pen potentiaalverschil, coulomb, volt, elektrische stroom en ampère aan de orde. Het laatste boekje behandelde de begrippen weerstand en soortelijke weerstand.

In de drie boekjes kwamen, naast de begrippen en hun definities, ook eigenschappen van de begrippen, voorbeelden, relaties tussen verschillende begrippen, regels (wet van Ohm bijvoorbeeld) en eenvoudige reken-voorbeelden aan de orde. In totaal besloeg de cursus veertien bladzijden. De oorspronkelijk in de cursus voorkomende analogieën werden eruit verwijderd.

b) Metaforen

Er werden verschillende wegen bewandeld om geschikte metaforen te verkrijgen. Eerst werd, zonder noemenswaardig resultaat, gezocht in diverse natuurkundemethoden en boeken over vakdidactiek. Vervolgens werden enkele natuurkundedocenten geraadpleegd: Zij gebruikten echter bij het onderwijs in de electriciteitsleer geen metaforen. Tenslotte werden door ons zelf ongeveer twintig metaforen bedacht. Deze werden uitvoerig bediscussieerd en getoetst aan de volgende criteria:

- 1) Zijn ze gemakkelijk te begrijpen?
- 2) Is de inhoud vertrouwd?
- 3) Zijn er veel formeel analoge begrippen en relaties?
- 4) Roepen ze gemakkelijk een mentale voorstelling op?
- 5) Hebben ze betrekking op een ander vakgebied dan de lesstof zelf?

De beste metaforen werden voorgelegd aan MAVO-docenten en op grond van hun commentaar herschreven. Uiteindelijk resulteerden vijf metaforen: Twee voor het eerste leerboekje, één voor het tweede leerboekje en twee voor het laatste leerboekje.

De eerste metafoor ging over de relaties tussen speculaas, een kruimel speculaas, en bestanddelen van speculaas (boter, suiker en dergelijke) (bij atomen, moleculen en stof) (175 woorden). De tweede metafoor bestond uit een beschrijving van een weegschaal (bij het evenwicht tussen kernlading en aantal electronen in een atoom) (200 woorden).

De derde metafoor beschreef de trek van de stad naar het platteland als er in de stad voedseltekort is (bij potentiaal verschil en elektrische stroom) (380 woorden). In de vierde metafoor werden verschillende verschijnselen beschreven die zich voor kunnen doen wanneer de waterdruk, de lengte en de dikte van een slang en het aantal verdraaiingen in de slang variëren (bij weerstand en de wet van Ohm) (400 woorden). De laatste metafoor beschreef de doorlaatbaarheid van verschillende soorten regenjassen (bij soortelijke weerstand) (200 woorden).

Er werd ook een uitvoerige instructie geschreven, waarin aan de hand van een tweetal voorbeelden werd uitgelegd hoe men metaforen kan gebruiken bij het studeren.

7.1.3.4 Toetsmateriaal

a) Voortoets

Uit de in het kader van voorgaande onderzoeken geconstrueerde itempool werden 30 vierkeuze-items geselecteerd die zo representatief mogelijk waren voor de totale lesstof.

b) Begripstoets

Uitgaande van de richtlijnen van Anderson (1972) werden 53 parafrase- en getransformeerde parafrasevragen geconstrueerd (vierkeuze-items). Deze vragen werden afgenomen bij 39 vierde klas VWO-leerlingen. Op basis van een itemanalyse werden twintig begripsitems geselecteerd.

c) Retentietoets

De retentietoets bestond uit de 30 voortoets items, aangevuld met tien nieuwe items alsmede de twintig begripsitems.

d) Metaforentoetsen

Er werden verschillende toetsen geconstrueerd, waarmee kon worden nagegaan, of de leerlingen de metaforen begrepen (drie metafoortoetsen van vijf items elk), of ze na afloop van de studie nog wisten waarover de metaforen gingen (één metafoortoets van zes items), en of ze verband konden leggen tus-

sen de metaforen en de lesstof (een vergelijkingstoets van zes items, waarin telkens een begrip uit de les moest worden gerelateerd aan een begrip uit de metafoor en een analogie-invultoets van zes items, waarin steeds analogieën moesten worden aangevuld). Een voorbeeld van een item van de analogie-invultoets is: "Geleider staat tot isolator is als spijkerjasje staat tot ...". Via simpele vijf puntenschaaltjes werd de proefpersonen gevraagd te beoordelen of metaforen in het algemeen een hulpmiddel bij het studeren vormen en of ze studeren met behulp van metaforen leuk vinden. Tevens werden de verschillende metaforen afzonderlijk beoordeeld.

7.1.3.5 Procedure

In tabel 7.1 is de procedure, voor de experimentele en controlegroep afzonderlijk, samengevat. Het onderzoek besloeg vier sessies van gemiddeld twee uur en vond plaats in localiteiten van de Katholieke Hogeschool Tilburg. Tussen de sessies verliep precies één week, alleen tussen sessie drie en vier verliepen twee weken. In de eerste sessie werd de Spy Ring History Test afgenomen in groepen van tien tot twintig proefpersonen met steeds twee of drie proefleiders. Op basis van de resultaten van de Spy Ring Test vond de indeling in experimentele en controlegroep plaats; Er werden vier groepen van leerlingen gevormd: Met hoge scores op operation learning en op comprehension learning (HH), met lage scores op beide (LL) en met hoge scores op de ene en lage op de andere (HL en LH). Aselect werd steeds uit elk van deze vier groepen één leerling aan de controlegroep en één aan de experimentele groep toegewezen. De volgende sessies vonden afzonderlijk - maar tegelijkertijd - plaats voor de experimentele en de controlecondities in groepen die in aantal variëerden van tien tot vijftwintig. Deze variaties in aantal waren noodzakelijk omdat het onderzoek plaats vond op de vrije middagen van de leerlingen en meer leerlingen op woensdagmiddag vrij hadden dan de donderdag- of de vrijdagmiddag.

In sessie twee werd begonnen met de voortoets. Vervolgens werden de drie leerboekjes één keer doorgewerkt (vast-aantal-

Tabel 7.1 Samenvatting van de procedure in experiment 1

Sessie	Conditie		
	Experimentele	Experimentele en Controle	Controle
Sessie 1		- Spy Ring History Test	
Sessie 2	<ul style="list-style-type: none"> - Metafoor instructie - Metaforen 1+2 metafoortoets - Metafoor 3 metafoortoets - Metaforen 4+5 metafoortoets metafoortoets vergelijkings- vragen analogietoets 	<ul style="list-style-type: none"> - Voortoets - Lesbrief 1 (één keer lezen) - Tijdregistratie - Lesbrief 2 (één keer lezen) - Tijdregistratie - Lesbrief 3 (één keer lezen) - Tijdregistratie - Begripstoets 	<ul style="list-style-type: none"> - Instructie - Analogietest dl. I - Analogietest dl. II
Sessie 3	<ul style="list-style-type: none"> - Metaforen 1+2 +vragen - Metafoor 3 +vragen - Metaforen 4+5 +vragen - HFT dl. I metafoortoets vergelijkings- vragen analogietoets beoordelings- schaal 	<ul style="list-style-type: none"> - Lesbrief 1 (2x) - Tijdregistratie - Lesbrief 2 (2x) - Tijdregistratie - Lesbrief 3 (2x) - Tijdregistratie 	<ul style="list-style-type: none"> - HFT dl. I - HFT dl. II
Sessie 4		- Retentietoets Cadeaubonnen	

trialsmethode) en werd de begripstoets afgenomen. In de experimentele groep werden bovendien de metaforen bestudeerd en de verschillende metaforentoetsjes afgenomen. De leerlingen in de controlegroep maakten de DAT-analogieëntest. Na elk leerboekje registreerden de proefpersonen zelf - onder strenge controle van de proefleiders - de studietijden.

Sessie drie was geheel hetzelfde als sessie twee, met dien verstande, dat de drie leerboekjes nu twee keer werden doorgevoerd en de voortoets natuurlijk niet opnieuw werd afgenomen. De controlegroep vulde nu de HFT in in plaats van de analogieëntest. Voor de metaforen werd steeds vijf minuten studietijd beschikbaar gesteld in sessie twee en zoveel minder als nodig was in sessie drie. In sessie vier vond de afname van de retentietoets plaats en het uitdelen van de cadeaubonnen.

7.1.3.6 Verwerking van de gegevens

Conform het gestelde in 5.3 werden op de data regressie-analyses uitgevoerd met gestandaardiseerde predictoren via een hiërarchische methode. De regressie-analyses werden uitgevoerd met behulp van het SPSS-programma Regression (Nie et al., 1975) op basis van een effectcodering. Als afhankelijke variabelen fungeerden de drie begripstoetsen (sessie twee, drie en vier), de retentietoets (kennisitems) en de leestijden. De volgorde van invoering van de variabelen geschiedde conform het analyseschema (zie 5.3). De hoofdeffecten werden getoetst doordat alleen het conditie-effect als predictor werd ingevoerd. De interactie-effecten werden getoetst door eerst het leerlingkenmerk, vervolgens het conditie-effect en tenslotte de interactieterm in te voeren. De hoofdeffecten werden dus op twee wijzen getoetst.

7.1.3.7 Design

Het design was een pretest-posttest-control-group-design met aselechte toewijzing aan de groepen.

7.1.4 Resultaten van experiment 1

De gemiddelde scores op de voortoets waren maar iets boven de verwachte waarde en de correlaties van de voortoets met de vier prestatietoetsen waren lager dan het gestelde criterium (respectievelijk .16, .20, .17 en .25). Daarom werden de voortoets-scores niet in de analyses betrokken. De descriptieve gegevens die volgens Cronbach en Snow (1977) in geen enkele ATI-rapportage mogen ontbreken, zijn weergegeven in bijlage 1.

Hoofdeffecten: hypothesen 1, 2 en 3

Volgens de hypothesen 1, 2 en 3 zouden op de begripstoets 1, de leestijd (eerste keer), de begripstoets 3 (na drie weken) en de kennistoets conditiehooftdeffecten op moeten treden. Dit bleek inderdaad het geval te zijn bij begripstoets 1 ($F(1,59) = 6.79, p < .05$), de kennistoets ($F(1,59) = 4.72, p < .05$) en de leestijd ($F(1,59) = 33.06, p < .01$), echter niet bij de derde begripstoets ($F(1,59) = 2.65, p < .10$). In tabel 7.2 zijn naast de genoemde F-waarden ook de percentages verklaarde variantie en de gemiddelde scores en standaarddeviaties vermeld.

De resultaten met betrekking tot hypothese 2 (leestijdeffecten) waren tegengesteld aan de voorspelling: De experimentele proefpersonen besteedden in totaal gemiddeld vijf minuten langer aan de lesstof dan de controle-proefpersonen. De overige resultaten waren in overeenstemming met de hypothesen.

Interactie-effecten: Hypothese 4

In de tabellen 7.3, 7.4 en 7.5 zijn de resultaten vermeld van de afzonderlijke regressie-analyses, waarin per afhankelijke variabele (drie begripstoetsen, kennistoets, leestijd eerste keer, leestijd tweede en derde keer, en totale leestijd) achtereenvolgens werden ingevoerd: Leerlingkenmerk (operation learning in tabel 7.3 en comprehension learning in tabel 7.4), conditie en interactie.

Volgens hypothese 4 zouden comprehension learners (hoog op comprehension learning en laag op operation learning scorend) meer van de metaforen moeten profiteren dan operation learners (die hoog op operation learning en laag op comprehension lear-

Tabel 7.2 Overzicht van de percentages door conditie verklaarde variantie, de F-waarden en de gemiddelden en standaarddeviaties op de verschillende afhankelijke variabelen (df=59)²⁾

	% var	F	\bar{X} exp	\bar{X} cont	SD exp	SD cont
Begripstoets 1 (na één keer lezen)	10.03	6.79**	7.94	6.17	3.09	2.04
Begripstoets 2 (na drie keer lezen)	3.7	2.28	8.41	7.34	2.49	3.00
Begripstoets 3 (na drie weken)	4.3	2.65	6.75	5.69	2.71	2.33
Kennistoets (na drie weken)	7.4	4.72**	15.19	12.62	5.36 ³⁾	3.60 ³⁾
Leestijd (eerste keer)	35.9	33.06***	25.28	19.90	3.75	3.54
Leestijd (tweede en derde keer)	1.8	1.10	34.28	35.97	6.30	6.25
Leestijd totaal	4.9	3.06*	59.56	55.86	8.05	8.47

* .05 < p < .10

** .01 < p < .05

*** p < .01

ning scores). De resultaten zijn echter niet in overeenstemming met deze hypothese:

De interacties van comprehension learning met de condities verklaarden nauwelijks enige variantie en geen van deze interacties bleek significant (zie tabel 7.4). De interacties van operation learning met de condities waren alle tegengesteld (qua richting) aan hypothese 4: Degenen die hoog scoorden op

²⁾ De betekenissen van de in dit en het volgende hoofdstuk gebruikte afkortingen en symbolen zijn weergegeven in bijlage VII.

³⁾ In dit geval zijn de varianties in de verschillende condities niet homogeen. Ook een t-toetsing die gebaseerd werd op een "separate variance estimate" leverde echter significante resultaten op.

Tabel 7.3 Overzicht van de percentages verklaarde variantie en de F-waarden bij operation learning als leerlingkenmerk (df=57)

Predictor	df	Begrips- toets 1 na 1x lezen % var	F	Begrips- toets 2 na 3x lezen % var	F	Begrips- toets 3 na 3 weken % var	F	Kennis- toets na 3 weken % var	F	Leestijd 1e keer % var	F	Leestijd 2e+3e keer % var	F	Leestijd totaal % var	F
Totale model	3	15.4	3.46**	9.1	1.90	22.4	5.48**	31.2	8.61**	44.1	15.01**	6.4	1.31	14.8	3.30**
Operation learning	1	3.8	2.56	3.9	2.45	3.2	2.35	17.1	14.17**	0.7	0.71	1.6	0.97	2.0	1.34
Conditie	1	11.0	7.41**	4.1	2.57	4.7	3.45*	8.6	7.13**	36.5	37.20**	1.7	1.04	5.3	3.55*
Interactie OXC	1	0.7	0.45	1.1	0.68	14.5	10.62**	5.5	4.56**	7.0	7.12**	3.1	1.89	7.6	5.08**
Residu	57	84.6		90.9		77.6		68.8		55.9		93.6		85.2	

* .05 < p < .10

** .01 < p < .05

*** p < .01

Tabel 7.4 Overzicht van de percentages verklaarde variantie en de F-waarden bij comprehension learning als leerlingkenmerk (df=57)

Predictor	df	Begrips- toets 1 na 1x lezen % var	F	Begrips- toets 2 na 3x lezen % var	F	Begrips- toets 3 na 3 weken % var	F	Kennis- toets na 3 weken % var	F	Leestijd 1e keer % var	F	Leestijd 2e+3e keer % var	F	Leestijd totaal % var	F
Totale model	3	12.0	2.59	13.9	3.08	10.1	2.13	25.2	6.40	40.4	12.87	4.5	0.90	10.4	2.21
Comprehen- sion 1	1	0.4	.03	7.5	4.97	2.6	1.65	12.2	9.29	0.5	0.48	3.4	2.03	3.2	2.04
Conditie	1	11.6	7.51	6.2	4.10	5.9	3.74	12.0	9.14	39.0	37.30	1.0	0.60	6.8	4.33
Interactie- cx cond	1	0.0	0.00	0.3	0.20	1.6	1.01	1.0	0.76	0.8	0.77	0.1	0.06	0.4	0.03
Residu	57	88.0		86.1		89.9		74.8		59.6		95.5		89.6	

* .05 < p < .10

** .01 < p < .05

*** p < .01

Tabel 7.5 Overzicht van de ongestandaardiseerde regressie-coëfficiënten per conditie voor het leerlingkenmerk operation learning

	constante in experimentele conditie	constante in controle-conditie	B voor operation-learning in experimentele conditie	B voor operation learning in controle-conditie
Begripstoets na één keer lezen	8.01	6.19	.82	.36
Begripstoets na drie keer lezen	8.46	7.34	.87	.29
Begripstoets na drie weken	6.80	5.69	1.49	-.47
Kennistoets na drie weken	15.35	12.59	3.18	.94
Leestijd eerste keer	25.44	20.00	1.74	-.66
Leestijd tweede en derde keer	34.28	35.89	1.91	-.32
Leestijd totaal	59.72	55.89	3.64	-.98

operation learning profiteerden meer van de metaforen dan degenen die laag scoorden op operation learning (zie tabel 7.5 voor de regressie-coëfficiënten). Alleen op de rententietoetsen (zie tabel 7.3) waren deze operation learning x conditie-interacties significant (kennistoets $F(1,57)=4.56, p<.05$) en begripstoets 3 ($F(1,57)=10.62, p<.01$). De interactie met betrekking tot de begripstoets is getekend in figuur 7.1. De differentiatiegebieden staan in tabel 7.6. De interactie met betrekking tot de begripstoets bleek disordinaal (twee differentiatiegebieden), de kennistoetsinteractie was ordinaal (één differentiatiegebied). Ook ten aanzien van de leestijden (zie tabel 7.3) traden deze operation learning x conditie-interacties op: Degenen die hoog scoorden op operation learning besteedden in de experimentele conditie meer tijd aan de lesstof dan in de controleconditie (zie tabel 7.5 voor de regressiecoëffi-

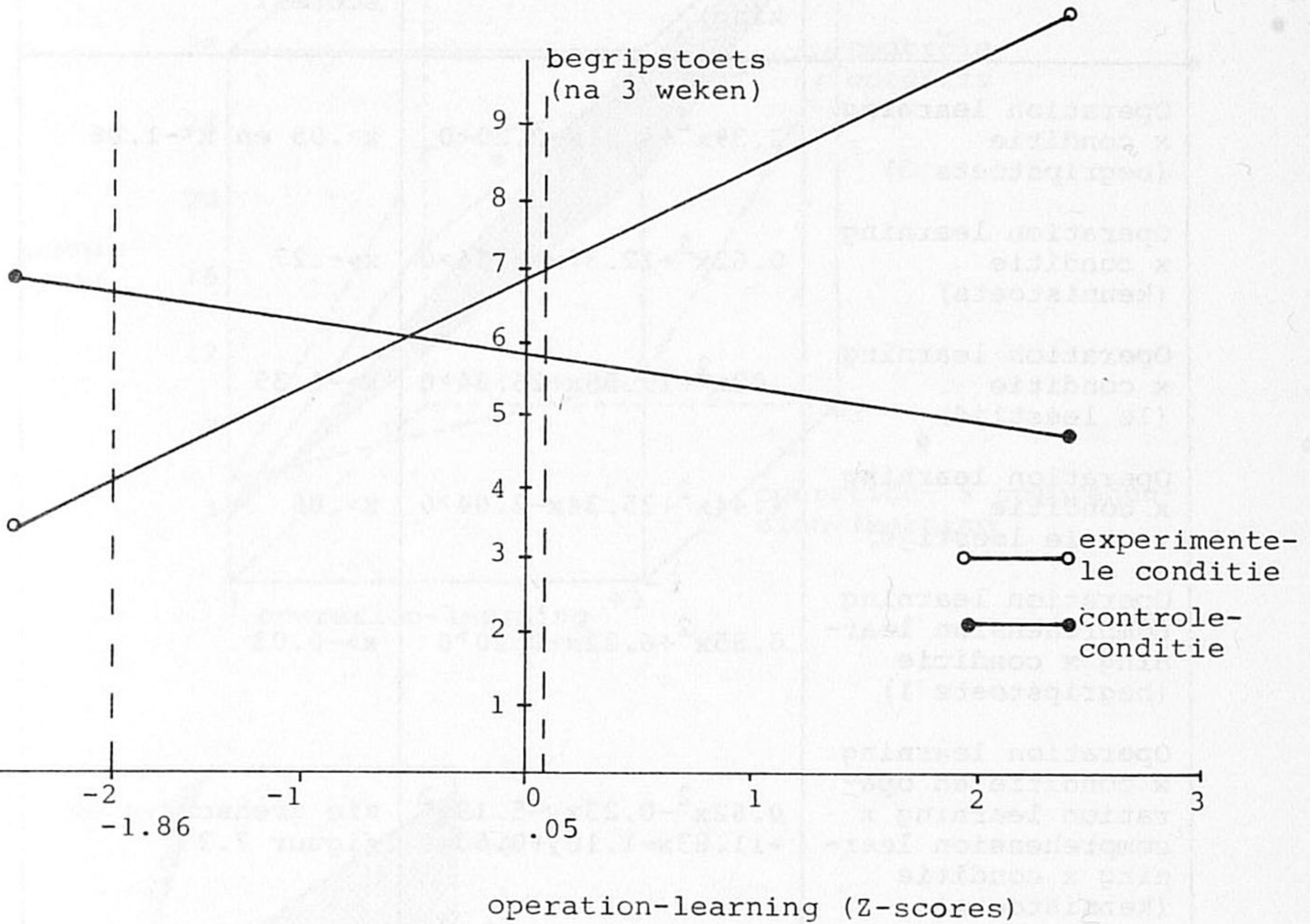
ciënten). Deze interacties waren significant (zie tabel 7.3) met betrekking tot de eerste lezing ($F(1,57)=7.12, p<.05$) en met betrekking tot de totale leestijd ($F(1,57)=5.08, p<.05$). De differentiatiegebieden zijn weergegeven in tabel 7.6. Beide interacties bleken ordinaal van aard. Om na te gaan of de interacties op de retentietoetsen konden zijn veroorzaakt door de interacties op de leestijden werden regressie-analyses uitgevoerd, waarbij de leestijd als eerste predictor werd ingevoerd (zie onder overige resultaten).

Omdat geen van de conditie x comprehension learning-interacties variantie verklaarde, was het weinig zinvol modellen met beide predictoren te toetsen. Wel werden interacties van hogere orde (operation learning x comprehension learning x conditie) getoetst op significantie (zie tabel 7.7). Op de eerste begripstoets ($F(1,54)=4.05, p<.05$) en op de kennistoets ($F(1,54)=2.90, p<.10$) werden deze operation x comprehension x conditie-interacties significant bevonden, op de overige afhankelijke variabelen niet. Zie figuur 7.2 waar de interactie voor de kennistoets is getekend in een kubus, omdat immers op deze afhankelijke variabele ook de operation learning x conditie-interactie significant was. De differentiatiegebieden zijn weergegeven in tabel 7.6.

Vraagstelling 5: Effecten van metaforen na drie keer lezen

Zoals ook uit tabel 7.2 kan worden opgemaakt, presteerden de experimentele-proefpersonen ook op begripstoets 2 (na drie keer lezen) meer dan de controle-proefpersonen; het verschil was echter niet significant ($F(1,59)=2.28, p<.10$). Na uitparialisatie van het comprehension learning hoofdeffect (zie tabel 7.4) werd het verschil echter wel significant ($F(1,57)=4.10, p<.05$). De tweede en de derde lezing van de tekst ging bij de experimentele-proefpersonen iets sneller dan bij de controle-proefpersonen (zie tabel 7.2); dit verschil in leestijd was echter niet significant ($F(1,59)=1.10, p<.10$). De totale leestijd over de drie lezingen te zamen bleef voor de experimentele-proefpersonen langer dan voor de controle-proefpersonen; het verschil was nu echter niet meer significant ($F(1,59)=$

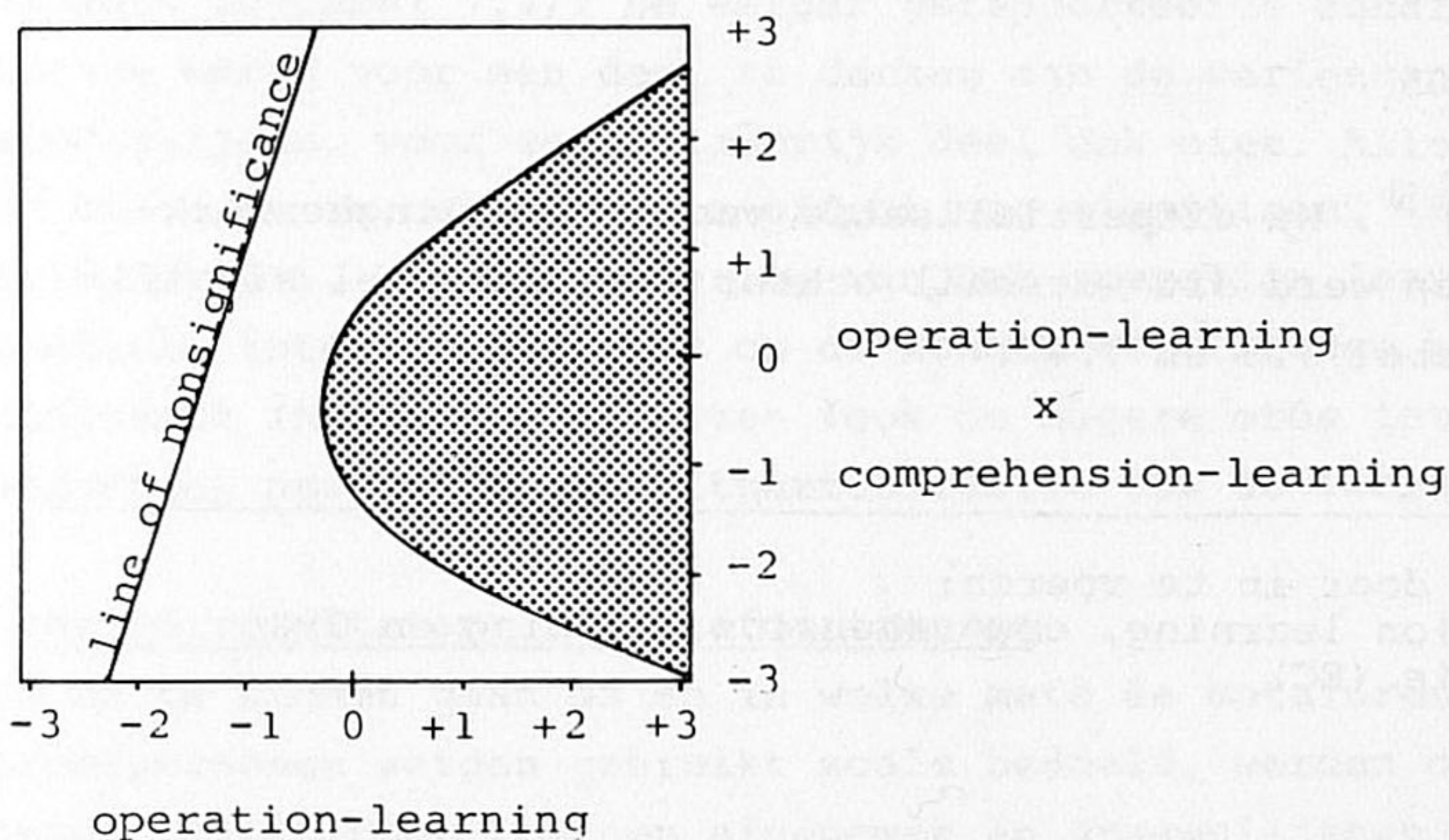
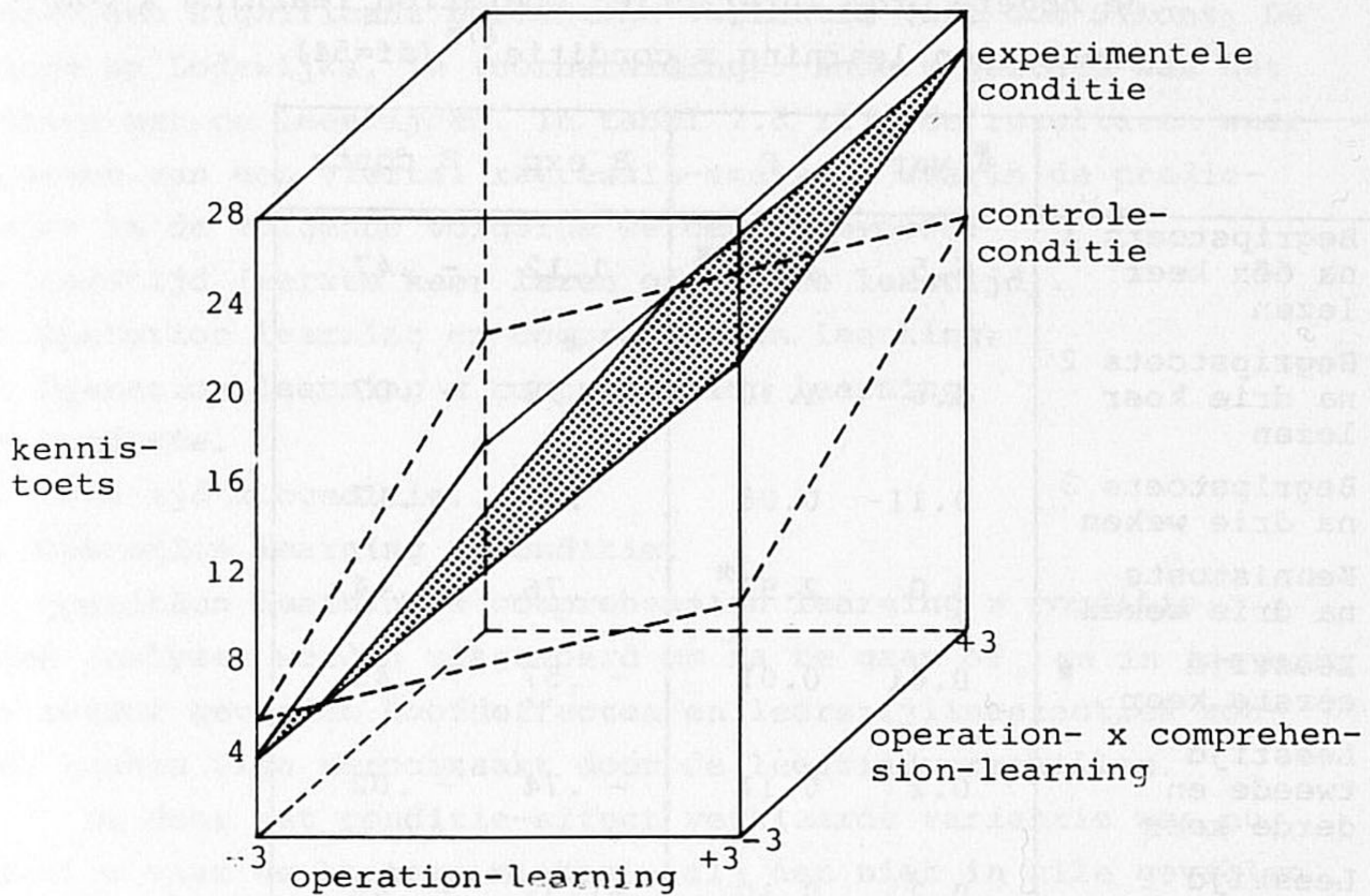
Figuur 7.1 De interactie van de condities met operation learning met betrekking tot het begripstoetsgedeelte van de retentietoets, alsmede de differentiatiegebieden



Tabel 7.6 Overzicht van de volgens de techniek van Plomp bepaalde differentiatiegebieden

Interactie	Ongelijkheid die de differentiatiegebieden bepaalt (of de grenscurvevergelijking)	Differentiatiegebieden (binnen de range van predictor-scores)
Operation learning x conditie (begripstoets 3)	$2.39x^2 + 4.34x - 0.20 > 0$	$x > .05$ en $x < -1.86$
Operation learning x conditie (kennistoets)	$0.62x^2 + 12.37x + 3.34 > 0$	$x > -.27$
Operation learning x conditie (1e leestijd)	$.02x^2 + 19.58x + 26.34 > 0$	$x > -1.35$
Operation learning x conditie (totale leestijd)	$4.44x^2 + 35.34x - 2.04 > 0$	$x > .06$
Operation learning comprehension learning x conditie (begripstoets 1)	$0.85x^2 + 6.22x + 0.20 > 0$	$x > -0.03$
Operation learning x conditie en operation learning x comprehension learning x conditie (kennistoets)	$0.62x^2 - 0.23xy - 5.19y^2 + 11.83x - 1.10y + 0.50 = 0$	zie grenscurve en figuur 7.2

Figuur 7.2 De interacties van de condities met operation learning en het product van operation learning met comprehension learning met betrekking tot de kennis-toets, alsmede de differentiatiegebieden



De vergelijking van de projectielijn op het X-Y-vlak luidt:
 $Y = 3.32x + 4.29$. De grenscurve van de differentiatiegebieden (hyperbool) luidt:
 $0.62x^2 - 0.23xy - 5.19y^2 + 11.33x - 1.10y + 0.50 = 0$

Tabel 7.7 Overzicht van de percentages verklaarde variantie en de bijbehorende F-waarden en ongestandaardiseerde B-gewichten voor de twee condities met betrekking tot de hogere orde interactie: Operation learning x comprehension learning x conditie⁴⁾ (df=54)

	% var	F	B exp	B contr
Begripstoets 1 na één keer lezen	6.5	4.65**	1.12	- .47
Begripstoets 2 na drie keer lezen	1.6	1.10	.92	- .07
Begripstoets 3 na drie weken	0.11	0.08	.40	.17
Kennistoets na drie weken	3.0	2.90*	.76	-1.44
Leestijd eerste keer	0.01	0.01	- .57	- .43
Leestijd tweede en derde keer	0.2	0.11	- .74	- .02
Leestijd totaal	0.2	0.10	-1.31	- .46

* .05 < p < .10

** .01 < p < .05

3.06, p < .10)⁵⁾. Na uitpartialisatie van de leerlingkenmerkhoofdeffecten werd dit verschil echter toch weer wel significant (zie tabel 7.3 en 7.4).

4) Verkregen door in te voeren:

- 1) operation learning, comprehension learning en OxC
- 2) conditie (EC)
- 3) OxEC
- 4) OxCxEC

5) Hoofdeffecten werden getoetst bij een α van .05, interactie-effecten bij een α van .10 (zie 5.3).

Overige resultaten: Uitpartialisatie van de HFT en de leestijden

Noch het HFT-hoofdeffect noch de HFT- x conditie-interacties verklaarden met betrekking tot één van de afhankelijke variabelen een significant percentage variantie (zie ook Simons, De Klerk en Lodewijks, in voorbereiding). Anders gesteld was het echter met de leestijden. In tabel 7.8 zijn de resultaten weergegeven van een viertal regressie-analyses waarin de predictoren in de volgende volgorde werden ingevoerd:

- 1) Leestijd (eerste keer lezen of totale leestijd).
- 2) Operation learning en comprehension learning.
- 3) Operation learning x comprehension learning.
- 4) Conditie.
- 5) Leestijd x conditie.
- 6) Operation learning x conditie.
- 7) Operation learning x comprehension learning x conditie.

Deze analyses werden uitgevoerd om na te gaan of, en in hoeverre de eerder gevonden hoofdeffecten en leerstijlinteracties zouden kunnen zijn veroorzaakt door de leestijdverschillen.

De door het conditie-effect verklaarde variantie was nu op alle vier de toetsen minder - zij het niet in alle gevallen significant minder - dan in de vorige analyses (nu respectievelijk 3.7, 4.3, 2.7 en 8.4 procent tegen 11.6, 6.2, 5.9 en 12.0 procent in tabel 7.4). De eerder gerapporteerde conditiehoofdeffecten waren voor een deel te danken aan de verlenging van de studietijden, voor een belangrijk deel ook niet. Alleen op de kennistoets is het conditie-effect nog significant ($F(1,52) = 8.16, p < .01$). Met uitzondering van het operation learning x conditie-interactie-effect op de kennistoets bleken alle significante interactie-effecten (ook de hogere orde interactie-effecten) bestand tegen uitpartialisatie van de leestijden.

Overige resultaten: De metafoortoetsen

Om na te kunnen gaan of en in welke mate de metaforen door de proefpersonen werden gebruikt zoals bedoeld, werden op verschillende tijdstippen toetsen afgenomen en vragenlijsten voorgelegd.

Tabel 7.8 Regressie-analyses met betrekking tot de voor leestijd gecorrigeerde prestaties (df=52)

Predictor	df	Begripstoets 1 na één keer lezen ⁶⁾	Begripstoets 2 na drie keer lezen	Begripstoets 3 na drie weken	Kennistoets na drie weken
		% var F	% var F	% var F	% var F
Totale model	8	31.0 2.92**	21.6 1.79*	26.8 2.38**	46.5 5.64***
Leestijd (L)	1	9.3 7.00**	3.7 2.45	9.1 6.46**	12.8 12.44***
Operation learning (O)	1	2.9 2.19	3.0 1.99	1.9 1.35	13.4 13.06***
Comprehen- sion lear- ning (C)	1	0.0 0.0	4.8 3.18*	0.8 0.57	5.7 5.54**
OxC	1	1.9 1.43	3.3 2.19	2.3 1.63	0.2 0.19
conditie (EC)	1	3.7 2.79	4.3 2.85	2.7 1.92	8.4 8.16***
LxEC	1	5.0 3.77*	0.6 0.40	0.7 0.50	0.0 0.0
OxEC	1	0.0 0.0	0.4 0.27	9.0 6.39**	2.8 2.72
OxCxEC	1	8.2 6.18**	1.5 0.99	0.2 0.14	3.0 2.92*
Residu	52	69.0	78.4	73.2	53.5

* .05 < p < .10

** .01 < p < .05

** p < .01

6) Hier vond correctie plaats voor de tijd die de eerste lezing kostte, in de andere drie gevallen betrof het de totale leestijd.

Er werd getoetst of de proefpersonen (in de experimentele conditie) de metaforen begrepen (onmiddellijk na het lezen van de metaforen), en of zij ze zich na het bestuderen van de lesstof nog herinnerden. Tevens werd getoetst of men in staat was verband te leggen tussen de begrippen in de lesstof en de begrippen in de metaforen. In tabel 7.9 zijn de resultaten met betrekking tot deze toetsen weergegeven.

Tabel 7.9 Resultaten met betrekking tot de verschillende metafoortoetsen (in sessie 2 en sessie 3) (N=38)

Toets	Sessie 2 mediaan	Sessie 3 mediaan	Aantal items
Inhoud metaforen 1 en 2	3.9	3.8	5
Inhoud metafoor 3	4.3	3.8	5
Inhoud metaforen 4 en 5	3.5	3.5	5
Inhoud van de me- taforen - na af- loop van de studie	4.6	4.8	6
Vergelijking meta- foor-les	2.9	3.1	6
Analogie-aanvul- vragen	3.1	3.2	6

De resultaten kunnen als volgt worden samengevat. De meeste proefpersonen begrepen de metaforen en onthielden ze ook tot na afloop van de bestudering van de lesstof. Dit blijkt uit de mediane scores op de metafoortoetsen: De meeste leerlingen beantwoordden drie à vier van de vijf vragen goed, zowel onmiddellijk na het bestuderen van de metaforen als aan het einde van de sessies.

De scores op de vergelijkings- en analogie-items waren echter aan de lage kant. Tenminste een deel van de proefpersonen slaagde er onvoldoende in verband te leggen tussen de lesstof en de metaforen: De helft van de proefpersonen beantwoordde minder dan de helft van de items juist.

Ook uit de beoordelingsvragen, waarvan de resultaten zijn weergegeven in tabel 7.10, bleek, dat veel proefpersonen het moeilijk vonden om verband te leggen tussen de metaforen en de lesstof. Hoewel de manier van studeren afweek van hun normale manier van studeren, vonden ze de metaforen echter wel een hulpmiddel dat nog leuk was ook.

Tabel 7.10 Beoordeling door de proefpersonen van het studeren via metaforen (N=38) (1=heel erg, 5=niet)⁷⁾

	Sessie 1 mediaan	Sessie 2 mediaan
Is de metafoor een hulpmiddel?	1.3	2.0
Lijkt deze manier van studeren op je eigen manier?	3.3	3.5
Vind je het een leuke manier van studeren?	2.4	2.6
Is het moeilijk om de lesstof te vergelijken met de verhalen vooraf?	2.7	2.7

7.1.5 Discussie

Voornaamste conclusies die op basis van de resultaten van het eerste experiment kunnen worden getrokken, zijn:

- a) Metaforen leiden tot een beter begrip - na één keer doorlezen - van (onbekende, abstracte) lesstof.
- b) Dit betere begrip komt gedeeltelijk (maar niet uitsluitend) tot stand, doordat proefpersonen meer studietijd uittrekken voor die eerste lezing. In tegenstelling tot hypothese 2 leiden de metaforen namelijk juist niet tot een snellere maar tot een langzamere eerste lezing van de lesstof.
- c) Metaforen hebben ook lange termijn effecten.
- d) Geen steun werd gevonden voor hypothese 4: Wel traden enkele onverwachte interacties tussen leerlingkenmerken en conditie op.
- e) Het tijdsverlies dat onder invloed van de metaforen bij de eerste lezing wordt geleden, wordt gedeeltelijk gecompens-

⁷⁾ De vragen werden niet letterlijk zo gesteld als ze hier zijn geformuleerd.

seerd door een snellere tweede en derde lezing van de lesstof.

f) MAVO-leerlingen vinden het studeren via metaforen moeilijk en anders, maar wel leuk en nuttig.

Het onverwachte leestijdeffect, dat tegengesteld was aan hypothese 2, vloeit waarschijnlijk voort uit de aard van de gebruikte metaforen. Veronderstelling achter hypothese 2 was dat sneller zou worden gelezen, doordat de lesstof ingepast werd in een bestaand schema.

Omdat geen overbekende schemata in aanmerking kwamen (zie 7.1.3), refereerden de metaforen echter (noodgedwongen) aan minder bekende verschijnselen. Dat betekende dat de proefpersonen eigenlijk twee nieuwe schemata moesten construeren: Eén relatief eenvoudig en minder abstract schema (de metafoor), en één daaraan gekoppeld moeilijker en abstracter electriciteitsleerschema. Wellicht was daarvoor extra studietijd nodig.

Een andere verklaring van de extra studietijd kan ook gelegen zijn in het feit, dat de leerlingen in de experimentele conditie op een andere wijze met de lesstof bezig waren. Zij moesten immers niet alleen de lesstof leren en begrijpen, ze moesten ook de relaties tussen de lesstof en de metafoor leggen. Waarschijnlijk kostte dat tijd. Overigens verlenen de geconstateerde studietijdverschillen steun aan de theoretische analyses in hoofdstuk 6.

De vraag naar de efficiëntie van het leren (vergelijk Faw en Waller, 1976 en hoofdstuk 6) krijgt nog extra gewicht in het licht van de geconstateerde studietijdverlenging (studietijdinvloed b in hoofdstuk 6). Weliswaar waren de prestatieverschillen slechts gedeeltelijk te wijten aan de langere studietijden (zie vorige paragraaf), maar hoe zou het zijn wanneer we ook nog de aan de metaforen op zich bestede studietijd (directe studietijdinvloed a in hoofdstuk 6) zouden meerekenen?

Een grove schatting daarvan kan als volgt worden gemaakt. Bij de eerste lezing gebruikten de experimentele-proefpersonen gemiddeld 5.38 minuten meer leestijd dan de controle-proefpersonen. Voeg daarbij de tijd die (nominaal) beschikbaar was voor de metaforen (3x5 minuten=15 minuten) dan komen we uit op

20 minuten extra. Dat is $20.38/19.90 \times 100 \% = + 100 \%$ extra studietijd. De leerwinst die daar tegenover staat, is $(7.94-6.17)/6.17 \times 100 \% = 29 \%$.

Dus tegenover 100 procent tijdverlies staat 29 procent leerwinst. Wanneer we de totale studietijden op eenzelfde manier vergelijken met de verschillen op de kennistoets, dan resulteert een tijdsverlies van $+ 40$ procent tegenover een leerwinst van 20 procent. Zo bezien lijkt het leren via metaforen een inefficiënte methode van leren.

Er zijn echter wel enkele complicaties:

- a) Leerwinst is waarschijnlijk geen lineaire functie van studietijd. Zou 40 procent extra studietijd die wordt besteed aan de lesstof zelf tot 40 procent leerwinst leiden?
- b) Wellicht vergroten metaforen de bereidheid van de proefpersonen om tijd te besteden aan de studie. Vergelijk ook de controleconditie in de studie van Peeck (1977) die in hoofdstuk 6 werd besproken: De controle-proefpersonen staakten de studie voortijdig.
- c) Uit de absolute hoogte van de studiestatistieken blijkt, dat de lesstof door de meeste proefpersonen nog lang niet voldoende werd beheerst. Er zal derhalve nog verdere studie plaats dienen te vinden. Halen degenen die metaforen tot hun beschikking hadden het tijdsverlies later nog meer in?
- d) De tijdsverschillen kunnen voor een deel ook te wijten zijn aan een onervarenheid met de wijze van studeren, die bij voortgaande oefening verdwijnt (zie ook hoofdstuk 4).

Voorlopig moeten we echter concluderen, dat het leren via metaforen geen efficiënte vorm van leren lijkt te zijn. Deze conclusie maakt de vraag naar interacties van de onderwijsvariabelen en leerlingkenmerken nog belangrijker dan zij al was. Op basis van die interacties zou immers kunnen worden bepaald voor welke leerlingen de extra tijdsinvestering in metaforen de moeite waard is en voor welke niet. De in deze studie gevonden interacties waren enigszins onverwacht. Daarom dient eerst replicatie van deze resultaten plaats te vinden, voordat verdere conclusies (bijvoorbeeld ten aanzien van tijdsinvesteringen) verantwoord zijn.

Volgens hypothese 4 hadden zowel de operation learning x conditie-interactie als de comprehension learning x conditie-interactie significant moeten zijn. De resultaten wezen echter uit dat drie operation learning x conditie-interacties weliswaar significant waren maar met een onverwachte richtingscoëfficiënt en dat de comprehension learning x conditie-interacties in het geheel niet significant waren. Het lijkt erop dat de aan hypothese 4 ten grondslag liggende activatieredenering (zie hoofdstuk 2), volgens welke metaforen zouden aansluiten bij de geprefereerde stijl van informatieverwerking van comprehension learners, onjuist was.

De geconstateerde operation learning x comprehension learning-interacties (op de eerste begripstoets en de kennistoets) wijzen erop, dat die leerlingen die zowel hoog scoorden op de operation learning dimensie als op de comprehension learning dimensie het meest profiteerden van de metaforen. Dit resultaat zou men kunnen interpreteren als steun voor een activatieredenering waarbij metaforen juist aansluiten bij de stijl van leren van "versatiles" (zie 7.1.2), die volgens Pask immers zowel goed zijn in operation learning als in comprehension learning.

Betrekken we bij deze beschouwing echter de aard van de steekproef van leerlingen (MAVO) en de gemiddelden op de operation en de comprehension dimensies dan ontstaat een ander beeld. De gemiddelden op de beide dimensies (zie bijlage 1) waren aan de lage kant (respectievelijk 32.5 en 42.0 voor operation en comprehension learning, met spreidingen van respectievelijk 13.4 en 15.1 bij een range van 0-100). Vooral de operation learning scores zijn laag te noemen. In dat licht bezien, zouden de gevonden interacties van hogere orde wellicht toch nog in overeenstemming kunnen zijn met hypothese 4. Onderzoek bij een heterogener samengestelde steekproef waarin wellicht meer spreiding in de operation en comprehension learning scores kan worden aangetroffen, zal moeten uitmaken welke interpretatie juist is.

De met betrekking tot de leestijden geconstateerde operation learning x conditie-interacties riepen de vraag op in hoe-

verre deze interacties met betrekking tot de begrips- en kennis-toetsen veroorzaakt waren door deze leestijdinteracties. De resultaten van de in tabel 7.8 gerapporteerde analyses wezen uit, dat de operation x conditie-interactie op de kennistoets voor een belangrijk deel werd verklaard doordat de leerlingen die hoog scoorden op de operation learning dimensie in de experimentele conditie langer studeerden (het percentage verklaarde variantie daalde van 5.5 naar 2.6 (procent)). De operation x conditie-interactie op de begripstoets (na drie weken) bleek slechts voor een deel te zijn verklaard door de studietijdverschillen (het percentage variantie daalde van 14.5 procent naar nog altijd 9 procent). De interacties van hogere orde operation learning x comprehension learning bleken in het geheel niet te zijn verklaard door leestijdverschillen (de percentages door de interactietermen verklaarde variantie namen zelfs iets toe respectievelijk van 7.2 naar 8.2 (procent) en van 2.9 naar 3.0 (procent)).

De ervaringen die in dit experiment werden opgedaan met de vast-aantal-trials methode waren zeer positief. De methode bleek goed hanteerbaar. De leerlingen hadden geen problemen met de betekenis van de opdracht: Eén, twee of drie maal doorwerken van de lesstof, de controle daarop verliep goed en ook de tijdsregistratie vormde geen probleem. Bovendien toonden de resultaten van het experiment aan, dat de methode een welkome aanvulling vormt op het arsenaal van studietijdcontrolemethoden. Er werden met betrekking tot studietijdsbesteding resultaten gevonden die met geen der andere bestaande methoden hadden kunnen worden opgespoord, doch wel de prestaties kunnen beïnvloeden.

In een volgend onderzoek zouden de volgende aspecten voor verbetering in aanmerking komen:

- a) De steekproef kan beter wat heterogener worden samengesteld (zie boven).
- b) Er dient een algemene intelligentietest te worden afgenomen, zodat kan worden nagegaan of interacties veroorzaakt kunnen zijn door intelligentie.
- c) Er dient voor te worden gezorgd, dat de leerlingen nog be-

- ter de relaties tussen de lesstof en de metaforen leggen (de resultaten op de vergelijkings- en analogie-invultoets waren nu te laag).
- d) De Spy Ring History Test dient vereenvoudigd en verkort te worden (de afname duurt nu te lang en de test leek aan de moeilijke kant).
- e) Sommige metaforen kunnen misschien nog wat worden verbeterd.

7.2 Experiment 2: Een replicatie bij een heterogener samengestelde steekproef³⁾

7.2.1 Doelstelling, hypothesen en vraagstellingen

Experiment 2 werd opgezet om na te gaan of voor de in experiment 1 bevestigde hypothesen ook steun kon worden gevonden, wanneer de in 7.1.5 gesuggereerde verbeteringen waren ingevoerd, en om na te gaan of de onverwachte resultaten in experiment 1 ten aanzien van de studietijden en aptitude-treatment-interacties ook zouden worden gevonden bij een heterogener samengestelde steekproef.

De hypothesen waren derhalve voor een deel (1 en 3) dezelfde als in experiment 1 (zie 7.1.1) en voor een deel nieuw (2 en 4). Met betrekking tot de leestijden (hypothese 2) werd nu - conform de resultaten van experiment 1 - verwacht dat de experimentele proefpersonen bij de eerste lezing meer tijd zouden gebruiken (hypothese 2a) en bij de volgende lezingen evenveel of minder (hypothese 2b).

Hypothese 4 werd op basis van de resultaten van experiment 1 veranderd in:

Hypothese 4: Het effect van metaforen is afhankelijk van de leerstijl van de leerlingen: Operation learners profiteren er meer van dan comprehension learners. In plaats van de activatieredenering die in experiment 1 de grondslag vormde voor de hypothese wordt hierbij geredeneerd volgens een supplantatiemodel (zie 2.4).

8) Dit experiment werd uitgevoerd door R. Feijen, R. Kleinmoedig en J. Tooten in het kader van hun doctoraalstudie onderwijspsychologie (zie Feijen, Kleinmoedig en Tooten, 1980).

Er werd op basis van de resultaten van experiment 1 ook nog een hypothese 5 toegevoegd:

Hypothese 5: Ook op een natoets (na het twee keer bestuderen van de tekst) leiden metaforen tot een prestatieverhoging.

7.2.2 Methode

7.2.2.1 Proefpersonen

Aan het onderzoek namen in totaal 114 proefpersonen deel, afkomstig uit zeven tweede klassen van twee MAVO-scholen en uit vier eerste klassen van een lyceum. De leeftijden van de proefpersonen variëerden van twaalf tot zestien jaar met een gemiddelde van 14.1. De toewijzing aan de condities geschiedde net als in experiment 1 aselekt op basis van de operation en comprehension learning scores.

Er deden 39 lyceum-leerlingen en 75 MAVO-leerlingen aan het onderzoek mee. Na afloop van het experiment werden de proefpersonen beloond met een cadeaubon. Van te voren werd meegedeeld, dat de hoogte van de beloning in zekere mate afhankelijk was van de studiestatistiek en de geleverde inspanning. Er waren 65 jongens en 49 meisjes.

Door ziekte en dergelijke deden slechts 84 proefpersonen aan alle vier sessies mee: 41 Experimentele en 43 controle-proefpersonen. Om gelijke aantallen leerlingen in de verschillende groepen te krijgen (zie 5.3) werden twee controle-proefpersonen aselekt verwijderd.

7.2.2.2 Testmateriaal

- a) De Spy Ring History Test.
- b) De DAT Analogieën Test.
- c) De Hidden Figures Test.

De Spy Ring History Test (zie 7.1.3.2) werd op basis van de ervaringen in experiment 1 verbeterd en vereenvoudigd: Alle moeilijke woorden werden verwijderd en de moeilijke zinnen werden vereenvoudigd. De instructie werd verduidelijkt en bepaalde gedeelten werden ingekort.

7.2.2.3 Lesmateriaal

Omdat de studieprestaties in experiment 1 gemiddeld laag konden worden genoemd, werd de natuurkundecursus herschreven: De verhaaltrant werd onderhoudender gemaakt, de zinnen werden vereenvoudigd en er werden herhalingen en rekenvoorbeelden ingebouwd. Bovendien werd een nieuwe paragraaf geschreven over "vrije electronen". Op basis van een taakanalyse werd ook de volgorde waarin de begrippen aan bod kwamen enigszins veranderd. De uiteindelijke cursus bestond uit twee delen. Deel I ging over de begrippen molecuul, atoom, electron, kernlading, ion en coulomb, en deel II ging over de begrippen potentiaal, potentiaal verschil, spanningsbron, stroomsterkte, weerstand, geleiding en de wet van Ohm. Deel I besloeg zeven pagina's en deel II negen.

Ook de metaforen werden aangepast. Op basis van onder andere de beoordelingen van de metaforen door de leerlingen werd slechts één metafoor gehandhaafd ("speculaas"). De andere vier metaforen werden vervangen door twee nieuwe - aanzienlijk kortere: Het "eethuisje" en het "schoolfeest".

De speculaasmetafoor (één pagina) had weer betrekking op de begrippen molecuul, atoom en stof. Het verhaal over het eethuisje (anderhalve pagina) ging over allerlei verschijnselen die zich voordoen wanneer een hele groep kinderen tegelijkertijd een eethuisje binnenkomt en elk kind dan een stoel moet zien te bemachtigen (bij electronen, ionen, kern en lading).

Het verhaal over het schoolfeest (twee pagina's) ging over de snelheid van doorstroming door de gangen van de school (smal en breed) wanneer alle jongens een drankje gaan halen bij de bar en de afhankelijkheid van die doorstroming van de snelheid waarmee het barpersoneel werkt (bij de wet van Ohm).

Ook de schriftelijke instructie over het studeren met behulp van metaforen (twee pagina's) werd verbeterd.

Om de leerlingen nog meer te dwingen gebruik te maken van de metaforen werden in de les op enkele plaatsen in totaal achttien meerkeuzevragen opgenomen, waarin verband moest worden gelegd tussen de lesstof en de metaforen. De proefpersonen in de controlegroep kregen op dezelfde plaats in de tekst neutrale vragen te beantwoorden (bijvoorbeeld: Wat hoort niet in het rijtje thuis: Gas, benzine, petroleum en dieselolie).

7.2.2.4 Toetsmateriaal

- a) Voortoets : 32 Vierkeuze items uit de kennistoets.
- b) Natoetsen : Twee toetsen van zestien items elk: Acht uit de voortoets en acht nieuwe.
- c) Begripstoetsen : Twee begripstoetsen van tien items elk (zie experiment 1).
- d) Retentietoetsen : 60 Vierkeuze items, namelijk de 32 voortoetsitems, de zestien nieuwe natoetsitems en twaalf nieuwe items.
- e) Transfertoets : Twintig vierkeuze items die waren gericht op toepassing van het geleerde in nieuwe probleemtypen.
- f) Relatietoets : Van de 77 paren begrippen die in de lessen voorkwamen, moesten de proefpersonen door middel van een vijf-puntsschaal aangeven of er een logische relatie bestond tussen die twee begrippen en (zo ja) wat de aard van die relatie was (zie Lodewijks, 1978).
- g) Metaforentoetsen: Evenals in experiment 1 werden verschillende metafoortoetsen afgenomen: Om na te gaan of de proefpersonen de metaforen begrepen en onthielden (elf items), en om na te gaan of ze de lesstof en de verhalen met elkaar in verband konden brengen (achttien items). Tevens werden ook weer beoordelingsschalen voorgelegd (veertien vragen).

7.2.2.5 Procedure

De procedure is samengevat in tabel 7.11. Verschillen met experiment 1 zijn:

- a) De voortoets werd in sessie 1 afgenomen in plaats van in sessie 2.
- b) In sessie 2 werd lesstof deel I bestudeerd en in sessie 3 deel II, terwijl in het vorige experiment alle lesstof in beide sessies werd bestudeerd.
- c) De lesstof werd slechts twee maal doorgenomen en niet drie maal.

Tabel 7.11 Samenvatting van de procedure in experiment 2

Sessie	Conditie		
	Experimentele	Experimentele en Controle	Controle
Sessie 1		<ul style="list-style-type: none"> - Voortoets - Spy Ring History Test 	
Sessie 2 (1 week na sessie 1)	<ul style="list-style-type: none"> - Metafoorinstructie - Metaforen 1+2 (10 min.) - Toets over metaforen - Vergelijkingsvragen - Metafoortoetsen 	<ul style="list-style-type: none"> - Lesbrieft 1 (1x lezen) - Tijdregistratie - Begripstoets 1 - Lesbrieft 1 (2e keer lezen) - Tijdregistratie - Natoets 1 	<ul style="list-style-type: none"> - Instructie HFT - Neutrale vragen
Sessie 3 (2 weken na sessie 2)	<ul style="list-style-type: none"> - HFT - Metafoor 3 (10 min.) - Metafoortoets - Vergelijkingsvragen - Metafoortoetsen 	<ul style="list-style-type: none"> - Lesbrieft 2 (1x lezen) - Tijdregistratie - Begripstoets 2 - Pauze - Lesbrieft 2 (2e keer lezen) - Tijdregistratie - Natoets 2 	<ul style="list-style-type: none"> - DAT dl. 1 - Neutrale vragen
Sessie 4 (1 week na sessie 3)	<ul style="list-style-type: none"> - DAT 1+2 	<ul style="list-style-type: none"> - Retentietoets - Transfertoets - Relatietoets - Interessevragen - Beloning 	<ul style="list-style-type: none"> - DAT dl. 2

- d) De begripstoets werd na het een maal doorlezen van de lesbrief afgenomen en de natoets na de tweede keer.
- e) De begripstoets werd niet ten tijde van de retentiemeting afgenomen.
- f) Er werden enkele nieuwe toetsen afgenomen: De transfertoets en de relatietoets.
- g) De DAT-analogieëntest werd ook bij de experimentele proefpersonen afgenomen.
- h) Verspreid door de les waren vergelijkingsvragen opgenomen in de experimentele en neutrale vragen in de controleversie van de lesbrief.
- i) De metaforen bleven tijdens de tekstbestudering voor de proefpersonen beschikbaar.
- j) Tussen sessie 2 en 3 verstreken twee weken (in verband met vakantie) en tussen sessie 3 en 4 verstreek één week.

7.2.2.6 Verwerking van de gegevens

Soorgelijke regressie-analyses als in experiment 1 werden volgens het analyseschema (figuur 5.3) uitgevoerd op de volgende afhankelijke variabelen: De begripstoets (=de som van de twee begripstoetsen), de natoets (=de som van de twee natoetsen), de retentietoets en de transfertoets, alsmede op de eerste leestijd (=de som van de eerste lezing van lesbrief 1 en de eerste lezing van lesbrief 2), de tweede leestijd (=de som van de tweede lezing van lesbrief 1 en de tweede lezing van lesbrief 2) en de totale leestijd (=de som van de voorgaande twee). De hoofdeffecten werden weer op twee wijzen getoetst.

7.2.2.7 Design

Het design was weer een pretest-posttest-control-group-design met aselechte toewijzing aan de condities.

7.2.3 Resultaten van experiment 2

De gemiddelde scores op de voortoets waren iets boven de verwachte waarde ($M=12.20$, $SD=2.86$ bij een range van 0-32) en de correlaties van de voortoets met de begripstoets, de natoets, de retentietoets en de transfertoets waren lager dan het ge-

stelde criterium van .60 (respectievelijk .48, .52, .48 en .39). De homogeniteitsindex van de voortoets (N=85) was .21, terwijl die van de natoets, begripstoets, retentietoets en transfertoets respectievelijk .83, .63, .90 en .71 bedroeg. Op basis van het beslissingsschema in figuur 5.2 werd de voortoets niet in de analyse betrokken. De descriptieve gegevens zijn weergegeven in bijlage II.

Hoofdeffecten: Hypothesen 1, 2, 3 en 5

Volgens de hypothesen 1, 2, 3 en 5 zou op de begripstoets, de natoets, de retentietoets, de transfertoets alsmede op de eerste leestijd en wellicht op de tweede leestijd een significant conditie-hoofdeffect moeten optreden. De resultaten in termen van percentages verklaarde variantie, F-waarden, gemiddelden en standaarddeviaties per conditie zijn weergegeven in tabel 7.12

Tabel 7.12 Overzicht van de percentages door het conditie-effect verklaarde variantie, de F-waarden, en de gemiddelden en standaarddeviaties per conditie op de verschillende afhankelijke variabelen (df=80)

	% var	F	\bar{x} exp	\bar{x} cont	SD exp	SD cont
Begripstoets	3.2	2.63	12.12	11.10	2.61	3.09
Natoets	4.9	4.13**	20.46	18.00	5.59	5.38
Retentietoets	1.7	1.40	29.10	26.41	10.25	10.28
Transferstoets	3.6	2.99*	9.32	7.93	3.70	3.58
Eerste leestijd	3.1	2.53	40.05	37.32	8.43	7.05
Tweede leestijd	14.4	13.51***	18.05	22.27	5.50	4.88
Totale leestijd	0.5	0.42	58.10	59.59	11.11	9.63

* .05 < p < .10

** .01 < p < .05

*** p < .01

Alle resultaten waren conform de hypothesen. Ten aanzien van de hypothesen 1 (begripstoets), 2a (eerste leestijd) en 3 (retentietoets en transfertoets) waren de verschillen echter niet significant. Wel significant waren de verschillen met betrekking tot hypothese 2b (tweede leestijd; $F(1,80)=13.51, p<.01$) en met betrekking tot hypothese 5 (natoets; $F(1,80)=4.13, p<.05$). Het verschil op de transfertoets (hypothese 3) was marginaal significant ($p<.10$).

In tabel 7.13 zijn de resultaten weergegeven van een viertal regressie-analyses (op de begripstoets, de natoets, de transfertoets en de retentietoets), waarbij - teneinde een significantietoetsing met meer onderscheidingsvermogen te verkrijgen - werd gecorrigeerd voor leestijdverschillen en voor de verschillende leerlingkenmerken. Na deze correcties bleek ook het conditie-effect met betrekking tot de transfertoets significant ($F(1,75)=6.60, p<.05$) en werd het conditie-effect met betrekking tot de begripstoets marginaal significant ($F(1,75)=3.61, p<.10$).

Interactie-effecten: Hypothese 4

In tabel 7.14 zijn de resultaten samengevat van een serie regressie-analyses waarin de interacties van operation learning, comprehension learning en operation x comprehension learning met conditie werden getoetst. Geen der in het eerste experiment aangetroffen interacties werd teruggevonden. De enige significante interacties (zie tabel 7.15) waren de operation learning x comprehension learning x conditie-interacties met betrekking tot de leestijden (zowel van de eerste lezing, als van de tweede als de totale). Deze interacties zijn zodanig, dat de leerlingen die hoog scoren op OxC (het product operation learning x comprehension learning) langer over de eerste lezing in de experimentele conditie deden en ook een langere totale leestijd hadden. De leerlingen die laag scoorden op OxC deden langer over de tweede lezing in de controleconditie en ook hun totale leestijden waren in de controleconditie langer. De interacties met betrekking tot de eerste en de tweede leestijd waren, blijkens de differentiatiegebieden, ordinaal en de

Tabel 7.13 Regressie-analyses met betrekking tot voor leestijd, intelligentie en cognitieve stijl gecorrigeerde studieprestaties (df=75)

Predictor	df	Begripstoets		Natoets		Retentietoets		Transfertoets	
		% var	F	% var	F	% var	F	% var	F
Totale model	6	39.8	8.25***	48.2	11.62***	42.2	9.13***	48.9	11.94***
Leestijd	1	3.1	3.86*	8.2	11.87***	2.1	2.72	2.6	3.82*
Vier leerling-kenmerken	4	33.8	10.52***	34.9	12.63***	38.2	12.39***	41.8	15.34***
Conditie	1	2.9	3.61*	5.1	7.38***	1.9	2.47	4.5	6.60**
Residu	75	60.2		51.8		57.8		51.1	

* .05 < p < .10

** .01 < p < .05

*** p < .01

Tabel 7.14 Overzicht van de percentages verklaarde variantie betreffende de interactie-effecten: Operation learning x conditie (OxEC), comprehension learning x conditie (CxEC) en operation x comprehension x conditie (OxCxEC) voor de gehele steekproef en voor de MAVO-leerlingen afzonderlijk

Afhan- kelijke varia- bele	Totale steekproef (N=82)			MAVO-leerlingen apart (N=50)		
	OxEC ⁹⁾	CxEC ¹⁰⁾	OxCxEC ¹¹⁾	OxEC ⁹⁾	CxEC ¹⁰⁾	OxCxEC ¹¹⁾
Begrips- toets	1.7	0.1	0.0	9.9**	0.0	0.1
Natoets	0.1	0.0	0.0	0.3	2.4	0.3
Kennis- retentie- toets	0.1	0.4	0.2	6.0*	0.9	0.7
Transfer- toets	0.3	0.3	1.1	0.2	0.0	1.0
Leestijd eerste keer	0.6	0.2	5.0**	0.1	0.0	2.4
Leestijd tweede keer	0.6	0.1	5.1**	0.2	0.5	7.6**
Leestijd totaal	1.1	0.3	8.4***	0.0	0.0	7.1*

* .05 < p < .10

** .01 < p < .05

*** p < .01

9) Verkregen via een regressie-analyse met als predictoren respectievelijk: O, EC, OxEC.

10) Verkregen via een regressie-analyse met als predictoren respectievelijk: C, EC, CxEC.

11) Verkregen via een regressie-analyse met als predictoren respectievelijk: 1) O, C
2) OxC
3) EC
4) OxEC, CxEC
5) OxCxEC.

Tabel 7.15 Overzicht van de significante interactie-effecten: F-waarden en regressiecoëfficiënten per conditie voor de totale steekproef en voor de MAVO-leerlingen apart

Inter-actie-effect	totale steekproef (N=82)				MAVO apart (N=50)			
	df	F	B in exp cond	B in contr cond	df	F	B in exp cond	B in contr cond
OxCxEC								
- eerste leestijd	74	4.26**	2.88	-.88	42	n.s.	-	-
- tweede leestijd	74	5.25**	1.09	-1.61	42	4.10**	3.09	-1.59
- totale leestijd	74	7.68***	3.97	-2.49	42	3.40*	5.95	-2.44
OxC								
- begripstoets	78	n.s.	-	-	46	6.10**	-0.09	1.55
- kennisretentie-toets	78	n.s.	-	-	46	2.98*	-1.94	2.04

* .05 < p < .10

** .01 < p < .05

*** p < .01

interactie met betrekking tot de totale leestijd bleek disordinaal (zie tabel 7.16).

Opgemerkt moet worden dat een hoge OxC score kan worden bereikt door hoge operation en comprehension learning scores maar ook door lage operation en comprehension learning scores (min maal min maakt plus!). Lage OxC scores ontstaan als gevolg van een lage operation of een lage comprehension score. Om na te gaan of het uitblijven van de interacties werd veroorzaakt door de grotere heterogeniteit van de steekproef, werden de verschillende regressie-analyses ook voor de groep MAVO-leerlingen afzonderlijk uitgevoerd. Uit de tabellen 7.14 en 7.15 kan worden afgelezen, dat er nu twee significante operation learning x conditie interacties werden geconstateerd: Met betrekking tot de begripstoets ($F(1,46)=6.10, p<.05$) en met betrekking tot de kennisretentietoets ($F(1,46)=2.98, p<.10$). De richtingen van deze interacties zijn precies omgekeerd aan de interacties in experiment 1: Degenen die laag scoorden op operation learning profiteerden meer van de metaforen dan degenen die hoog op deze dimensie scoorden.

Resultaten met betrekking tot de relatietoets

In tabel 7.17 zijn de resultaten vermeld met betrekking tot de relatietoets. In deze tabel zijn de resultaten samengevat van regressie-analyses met betrekking tot de conditie-effecten en met betrekking tot de interactie-effecten (OxEC en CxEC) voor twee uit de relatietoets afgeleide variabelen Rep en Prod. Deze variabelen werden volgens de procedure van Lodewijks (1978) als volgt berekend: Er werden drie soorten relaties tussen begrippen onderscheiden:

- a) door experts als niet bestaand beoordeelde relaties (niet bestaand (NB), in totaal 32 relaties);
- b) door experts als bestaand beoordeelde relaties die expliciet in de tekst werden behandeld (expliciet bestaand (EB), in totaal 26 relaties), en
- c) door experts als bestaand beoordeelde relaties die niet expliciet in de studietekst werden behandeld (impliciet bestaand (IB), in totaal 22 relaties).

Tabel 7.16 Overzicht van de differentiatiegebieden bepaald via de techniek van Plomp

Interactieterm	Ongelijkheid die de differentiatiegebieden bepaalt	Differentiatiegebied(en) (binnen de range van predictorscores)
Operation learning x comprehension learning x conditie (1e leestijd)	$2.98x^2 + 17.10x - 13.08 > 0$	$x > 0.68$
Operation learning x comprehension learning x conditie (2e leestijd)	$2.61x^2 - 25.92x + 26.65 > 0$	$x < 1.17$
Operation learning x comprehension learning x conditie (totale leestijd)	$23.44x^2 - 44.92x + 1.94 > 0$	$x < 0.04$ en $x > 1.87$
Operation learning x conditie (begripstoets, Mavo)	$0.89x^2 - 2.54x - 1.07 > 0$	$x < -0.37$

De Rep score geeft aan hoeveel procent van de expliciet bestaande relaties door de leerling als bestaand werd beoordeeld en de Prod score geeft aan hoeveel procent van de impliciet bestaande relaties door de leerling als bestaand werd beoordeeld. Beide scores (Rep en Prod) worden gecorrigeerd voor het percentage niet bestaande relaties dat door de leerling ten onrechte als bestaand werd gekwalificeerd. De formules luiden:

$$\text{Rep} = \left(\frac{\text{EB}}{26} \times 100\right) - \left(\frac{\text{NB}}{32} \times 100\right)$$

$$\text{Prod} = \left(\frac{\text{IB}}{22} \times 100\right) - \left(\frac{\text{NB}}{32} \times 100\right)$$

De Rep score geeft aan hoe goed iemand de expliciet in de tekst vermelde relaties kan reproduceren en de Prod score geeft aan hoe goed iemand de niet expliciet in de tekst behandelde relaties kan produceren.

Tabel 7.17 Resultaten met betrekking tot de Prod en Rep scores van de relatietoets

	Rep-score	Prod-score
Door het conditie-effect verklaarde variantie	3.2	8.1
F-waarde (df=1,80)	2.63	7.05 ^{***}
\bar{X} in experimentele conditie	18.55	11.78
\bar{X} in controleconditie	12.81	2.43
SD in experimentele conditie	16.69	17.03
SD in controleconditie	15.34	14.79
Door operation learning x conditie verklaarde variantie	0.4	1.0
F-waarde bij Oxec (Oxec)	0.34	0.83
Door comprehension learning x conditie verklaarde variantie	1.7	0.0
F-waarde bij Cxec (Cxec)	1.40	0.00

*** $p < .01$

De metaforen leidden tot zowel hogere Rep als hogere Prod scores. Het verschil met betrekking tot de Rep score is echter niet significant, het verschil met betrekking tot de Prod score is significant op het één procentniveau ($F(1,80)=7.05$). Geen der interacties bleek significant.

De metafoortoetsen en de beoordelingsvragen

Ook in dit experiment begrepen de leerlingen de metaforen in voldoende mate en onthielden ze de inhoud van de metaforen ook tot na de studie (zie tabel 7.16). Maar evenals in het vorige experiment waren de scores op de analogievragen (waar een verband moest worden gelegd tussen de lesstof en de metaforen) te laag. Gemiddeld werd slechts de helft van deze vragen goed beantwoord. Dit is te meer opmerkelijk, daar deze vragen in dit experiment tijdens het bestuderen van de les, met de metaforen daarbij nog beschikbaar, moesten worden beantwoord.

Tabel 7.18 Resultaten met betrekking tot de verschillende metafoortoetsen (N=42)

Toets	Aantal items	Mediaan
Metaforen 1 en 2		
voor de studie	6	5.2
na de studie	6	4.9
Metafoor 3		
voor de studie	5	4.2
na de studie	5	4.3
Analogievragen		
metaforen 1 en 2	9	5.4
metafoor 3	9	4.2

In tabel 7.19 tenslotte zijn de resultaten vermeld van de beoordelingsvragen. De leerlingen beoordeelden de metaforen en het leren via metaforen over het algemeen vrij gunstig; 90 procent van hen vond bijvoorbeeld dat de aan de metaforen be-

stede tijd nuttig besteed was (zie vraag 8) en 50 procent van hen vond het leren via metaforen leuk (tegenover slechts 10 procent niet leuk). Opvallend was verder dat slechts 17 procent van de leerlingen één of meer metaforen als verwarringstichtend beoordeelden en dat slechts 17 procent van de leerlingen aangaf moeite te hebben gehad met het vergelijken van de lesstof en de metaforen. Dit laatste gegeven lijkt tegenstrijdig met de relatief slechte prestaties op de analogie-items.

Tabel 7.19 Resultaten met betrekking tot beoordelingsvragen:
Frequenties van de verschillende antwoorden (N=42)¹²⁾

Vraag	nee/ weinig	?	ja/ veel
1. Hebben de metaforen ¹²⁾ geholpen bij het leren?	7	13	22
2. Waren de metaforen duidelijk?	4	8	30
3. Vond je de metaforen leuk?	7	17	17
4. Kon je de les en de metaforen vergelijken?	7	15	20
5. Vond je de metaforen ingewikkeld?	28	11	3
6. Hadden de metaforen uitgebreider moeten zijn?	13	20	7
7. Hoe bekend kwamen de metaforen je voor?	19	11	13
8. Vond je de tijd aan de metaforen besteed nuttig?	4	-	36
9. Begreep je de les de tweede keer beter dan de eerste?	6	12	24
10. Vond je de metaforen verwarringstichtend?	34	-	7
11. Vond je de metaforen een hulpmiddel voor het begrijpen?	12	-	30
12. Vond je de metaforen een hulpmiddel voor het onthouden?	17	-	25
13. Leek deze manier van leren op je normale manier?	23	15	4
14. Vond je deze manier van leren leuk?	4	17	21
15. Kwam de lesstof je bekend voor?	25	12	5

¹²⁾ De vragen zijn anders gesteld dan ze hier uit presentatie-overwegingen geformuleerd zijn. De term metafoor werd in de vragenlijst bijvoorbeeld niet gebruikt. In plaats daarvan werd over "verhalen" gesproken.

Overige resultaten: De DAT en de HFT

Er werden ook exploraties verricht naar de interacties tussen de DAT-analogieëntest respectievelijk de HFT met de condities. Op geen der afhankelijke variabelen traden echter significante AT-interacties op.

7.2.4 Discussie

De resultaten kunnen als volgt worden samengevat:

- a) Na één keer lezen van de lesstof begrepen de leerlingen iets meer van de lesstof als ze vooraf metaforen ter beschikking hadden dan wanneer ze die niet hadden; het verschil was echter niet significant.
- b) Voor de eerste lezing hadden de leerlingen in de experimentele groep iets meer tijd nodig (niet significant) dan die in de controlegroep; voor de tweede gebruikten ze echter significant minder tijd, zodat ze ook in totaal iets minder tijd gebruikten.
- c) De lange-termijn-effecten van de metaforen waren alleen met betrekking tot de transfertoets significant, niet met betrekking tot de kennistoets.
- d) Op de natoets (na twee keer lezen) waren er duidelijke verschillen tussen de condities ten gunste van de experimentele groep.
- e) De eerder geconstateerde interacties werden niet teruggevonden; wel werden enkele andere interacties aangetroffen die betrekking hadden op verschillende studietijdbestedingen onder de twee condities.
- f) De metaforen leidden ertoe, dat de leerlingen in de experimentele groep significant meer relaties tussen begrippen die niet expliciet in de tekst werden behandeld konden produceren.
- g) De metaforen en het leren door middel van metaforen werden gunstig beoordeeld: Men vond de tijd nuttig besteed.

De resultaten van experiment 2 zijn wat betreft de hoofdeffecten in grote lijn vergelijkbaar met die van experiment 1. De verschillen tussen de groepen zijn echter wat de studieprestaties betreft wat minder pregnant dan in het vorige experiment.

Een mogelijke oorzaak hiervoor zou kunnen zijn gelegen in de herschrijving van het lesmateriaal. De metaforen zouden minder invloed kunnen hebben gehad doordat het nieuwe lesmateriaal beter was georganiseerd dan het vorige. Mayer (1979b) beschreef vier studies waarin materiaal x conditie-interacties optraden: Metaforen en andere organizers werkten alleen bij relatief slecht georganiseerd lesmateriaal. Een andere - mijns inziens meer waarschijnlijke - oorzaak van deze verschillen tussen de twee experimenten is gelegen in verschillen in studietijdbesteding. Waren de prestatieverschillen tussen de groepen in experiment 1 voor een deel te danken aan de langere totale studietijden, in experiment 2 traden de prestatieverschillen op ondanks kortere totale studietijden in de experimentele groep. In dat licht bezien, zijn de resultaten wat betreft de studieprestaties in de beide experimenten vergelijkbaar.

Er kunnen vele mogelijke verklaringen worden aangevoerd ten aanzien van het gegeven, dat in dit experiment de experimentele proefpersonen minder extra tijd gebruikten voor de eerste lezing en ten aanzien van het feit dat de verschillen in studietijd tussen de groepen bij de tweede lezing groter waren dan in het vorige experiment (onder andere kunnen zij worden gezocht in de aard van het lesmateriaal, het beschikbaar hebben van de metaforen, het moeten invullen van de neutrale vragen in de controleconditie, de verdeling van de lesstof over de sessies, en de aard van de metaforen enzovoorts). De beschikbare gegevens leveren echter geen aanwijzingen op voor een keuze tussen deze verklaringen.

Kan het leren via metaforen in dit experiment een efficiënte wijze van leren worden genoemd? Na de eerste lezing van de lesstof staat tegenover een tijdsverlies van 47.5 procent $((40.05+15-37.32)/37.32) \times 100\%$ een winst op de begripstoets van 9.2 procent $((12.12-11.10)/11.10) \times 100\%$. Na twee keer lezen is het tijdsverlies echter nog maar 13,5 procent en daar staat dan leerwinst tegenover van respectievelijk 13.7 procent, 10.2 procent en 17.5 procent op respectievelijk de natoets, de retentietoets en de transfertoets. Met grote voorzichtigheid (zie de interpretatieproblemen in 7.1.5) zou men kunnen conclu-

deren, dat het leren via metaforen wellicht toch, met name wat betreft transfereffecten, een efficiënte wijze van studeren is.

Dat de metaforen op de lange termijn vooral een effect hadden op de transferresultaten, en slechts in mindere mate op de kennisresultaten is in overeenstemming met de assimilatietheorie en de resultaten van het onderzoek van Mayer (1979a, 1979b). In experiment 1 kon dit verschil niet naar voren komen, omdat daarbij geen transfertoets werd afgenomen. Wellicht is in dat experiment het retentietoetseffect ook veroorzaakt door transferachtige items in de retentietoets. De leerlingen in de experimentele groep produceerden op de relatietoets meer relaties tussen begrippen dan de leerlingen in de controlegroep. Dit gegeven wijst erop, dat leerlingen in de metaforengroep een betere representatie van het netwerk van begrippen hebben gevormd (zie hoofdstuk 3) dan leerlingen in de controlegroep. Een hoge Rep score kan men nog bereiken door, zonder overzicht over het netwerk van begrippen, geïsoleerde kennis te reproduceren; een hoge Prod score wijst echter op een beter overzicht over de stof. Het is dan ook van groot belang dat de metaforen juist met betrekking tot deze Prod scores een significant effect hadden. Overigens betekent ook dit resultaat een ondersteuning van de assimilatietheorie van Mayer (1979a, 1979b).

De in het vorige experiment geconstateerde interactie-effecten werden in dit experiment niet teruggevonden. Een situatie die zich helaas veelvuldig voordoet in aptitude-treatment-interactie-onderzoek (zie Cronbach en Snow, 1977).

Een voor de hand liggende verklaring (zie ook 7.1.5) voor het uitblijven van interacties zou kunnen worden gezocht in de grotere heterogeniteit van de steekproef in het tweede experiment. Deze verklaring is echter onjuist: Wanneer men de resultaten voor de MAVO-leerlingen apart beziet om een betere vergelijkbaarheid tussen de twee experimenten te verkrijgen (tabellen 7.14 en 7.15), dan blijken sommige operation learning x conditie-interacties wel op te treden maar in een richting die tegengesteld is aan de interacties uit experiment 1. Overigens zijn de bij de MAVO-leerlingen optredende interacties nu wel in overeenstemming met de oorspronkelijke hypothese 4.

Een andere verklaring voor het niet terugvinden van de interactie-effecten, zou kunnen worden gezocht in de studietijdsbesteding. In het vorige experiment bleken de operation learning x conditie-interacties immers voor een deel te zijn veroorzaakt door langere studietijden, terwijl in het tweede experiment ook de studietijdinteracties uitbleven. Het is echter onduidelijk waarom deze studietijdinteracties uitbleven, zodat het probleem alleen maar is verschoven.

Een laatste reden voor het niet terugvinden van de interactie-effecten kan betrekking hebben op de Spy Ring History Test. Wellicht is deze test te moeilijk voor de in ons experiment betrokken leerlingen met als gevolg dat de metingen onvoldoende betrouwbaar en valide waren. Hoewel Pask et al. (1978) aangaven, dat de test geschikt is voor kinderen vanaf + veertien jaar, bleken veel leerlingen in onze experimenten moeilijk de concentratie op te kunnen brengen om gedurende twee uur of meer gemotiveerd te blijven werken aan de testopgaven. De door ons ingevoerde vereenvoudigingen (zie 7.2.2) betekenden een verbetering (vergeleken met het vorige experiment), maar losten de problemen toch nog onvoldoende op. Er is dan ook behoefte aan onderzoek naar betrouwbaarheid en validiteit van de Spy Ring History Test.

Duidelijk is in ieder geval, dat de twee experimenten samen een weinig consistent beeld geven van de AT-interacties en dat vooralsnog geen steun is gevonden voor hypothese 4, namelijk dat comprehension learners meer van metaforen zouden profiteren dan operation learners.

Evenals in het vorige experiment waren ook nu de resultaten op de vergelijkingstoetsen, waarbij de lesstof en de metaforen met elkaar moesten worden vergeleken, laag (+ 50 procent goede antwoorden gemiddeld). Dit is een des te opmerkelijker resultaat wanneer we bedenken, dat de vergelijkingen nu tijdens het bestuderen van de stof (met de metaforen erbij) moesten worden gemaakt.

Hieruit zou men kunnen afleiden, dat een deel van de leerlingen onvoldoende in staat was de juiste analogierelaties te leggen. Misschien is niet het voornaamste probleem dat leer-

lingen moeten worden aangespoord om de relaties tussen de lesstof en de metaforen te leggen (zoals we op basis van experiment 1 dachten), maar dat een deel van de leerlingen onvoldoende in staat is tot analogieredeneren.

Dan zou men echter verwachten, dat er een interactie zou optreden van de vaardigheid om in analogieën te denken (DAT-analogieëntest) met de condities, hetgeen echter niet het geval was. Evenzo zou men een hoge correlatie verwachten van de scores op vergelijkingstoetsen en de DAT-scores, hetgeen evenmin het geval was (correlaties rondom de .40).

Op drie manieren zou iets aan het probleem van de lage vergelijkingstoetsscores kunnen worden gedaan:

- a) De verbanden tussen de lesstof en de metaforen zouden explicieter gelegd kunnen worden (zie experimenten 3 en 6).
- b) Leerlingen zouden kunnen worden getraind in het leggen van relaties tussen lesstof en metaforen.
- c) De invloed van metaforen op het leren van oudere proefpersonen zouden kunnen worden vastgesteld (zie experimenten 4 en 5).

7.3 Experiment 3: Visualizers versus verbalizers¹³⁾

7.3.1 Doelstelling, vraagstellingen en hypothesen

Doel van het experiment was de efficiëntie van het leren via metaforen te onderzoeken met behulp van de efficiëntiemethode (zie 6.1). Een ander doel van het experiment was de resultaten die werden verkregen met de vast-aantal-trials methode te vergelijken met resultaten van een experiment waarin de efficiëntiemethode werd gebruikt. Experiment 3 is dan ook in veel opzichten (qua materiaal, proefpersonen en procedure) vergelijkbaar met experiment 1.

In paragraaf 6.4 werden twee efficiëntievraagstellingen onderscheiden. Vraag c: "Heeft het extra materiaal - gegeven een vaste (nominale) studietijd - effect op studieprestaties?" en vraag d: "Zorgt extra materiaal voor efficiënter leren?" of ook "Heeft het extra materiaal naast een kwantitatief effect een kwalitatief effect?"

¹³⁾ Dit onderzoek werd uitgevoerd door A. Fijten en M. Koene in het kader van hun doctoraalstudie onderwijspsychologie (zie Fijten en Koene, 1979).

Uit tabel 6.1 kan worden afgelezen, dat de efficiëntiemethode bij vraag c geen interpretatieproblemen oplevert bij een krappe studietijd, maar wel een tweetal potentiële interpretatieproblemen (A en E) bij vraag d. Bij een ruime studietijd is vraag d met de efficiëntiemethode onbeantwoordbaar geworden en kan bij vraag c interpretatieprobleem B optreden (zie 6.4). Er is in dit experiment dan ook gekozen voor krappe nominale studietijden.

Evenals in experiment 1 werd verwacht, dat metaforen een effect zouden hebben op de korte- en de lange-termijn-retentie van de relatief onbekende en abstracte natuurkundestof (evenwel in minder sterke mate dan in experiment 1, door de strengere studietijdcontrole). Deze verwachtingen waren weer gebaseerd op de theoretische analyses met betrekking tot de functies die metaforen kunnen vervullen (hoofdstuk 3) en op het voorgaande onderzoek naar metaforen (hoofdstuk 4).

Ook nu werd weer via de principes van de correspondentieanalyse gezocht naar leerlingkenmerken die met de condities zouden kunnen interacteren (zie hoofdstuk 2). Dat betekent dat werd gezocht naar leerlingkenmerken die zich functioneel laten beschrijven in termen van onderliggende processen, die in verband konden worden gebracht met moeilijkheden van de leertaak en met functies van de metaforen.

Een intuïtieve leertaakanalyse (vergelijk Simons en Lodewijks, 1979) leverde als één van de moeilijkheden van de leertaak op: De vorming van adequate mentale voorstellingen. De natuurkundebegrippen zijn abstract en formeel van aard en het kan moeilijk zijn zich daar iets bij voor te stellen. Deze moeilijkheid correspondeerde met één van de in hoofdstuk 3 beschreven functies van metaforen: Het vergemakkelijken van de vorming van mentale voorstellingen en het "voorstelbaar" maken van abstracte informatie. Als corresponderend leerlingkenmerk bij deze correspondentie tussen een leertaakmoeilijkheid en een functie van metaforen werd de dimensie visualizers-verbalizers gekozen. Het op basis van deze dimensie te maken onderscheid tussen leerlingen, dat onlangs zowel theoretisch als empirisch

nieuw leven werd ingeblazen in de dissertatie van Boekaerts (1978), wordt nader beschreven in paragraaf 7.3.2.

De hypothesen in experiment 3 waren:

1. Als metaforen worden toegevoegd aan relatief abstracte lesstof, zijn - bij een krappe nominale studietijd - de studieprestaties op korte en lange termijn hoger dan zonder metaforen.
2. De effecten van metaforen zijn afhankelijk van de coderingsstijl (visualizers- versus verbalizersstijl) van de leerlingen.
3. De effecten van metaforen zijn minder groot bij gebruik van de efficiëntiemethode vergeleken bij gebruik van de vast-aantal-trials methode.

7.3.2 Visualizers versus verbalizers: Theorie en onderzoek

Mentale voorstellingen hebben onder andere bekendheid gekregen door het werk van Paivio (zie bijvoorbeeld Paivio, 1969). Diens "dual-code"-hypothese hield in dat er twee afzonderlijke - maar wel met elkaar verband houdende - representatiesystemen zouden bestaan: Een verbaal opslagsysteem en een "imagery" opslagsysteem. Anderson en Bower (1973) en Pylyshyn (1973) hebben echter aannemelijk gemaakt, dat er niet twee verschillende representatiesystemen bestaan (verbaal en imagery), maar dat alle opslag (storage) van informatie propositioneel van aard is.

Boekaerts (1978) heeft onlangs een nieuwe interpretatie van de "dual-code"-hypothese voorgesteld, die niet in strijd is met de aanname van propositionele informatie-opslag. Daartoe moet men - aldus Boekaerts - een onderscheid maken tussen informatie-opslag, en de codering en "retrieval" van informatie. Ook al is de opslag van informatie propositioneel van aard, dan kunnen er toch nog verschillende soorten van codering en retrieval bestaan: Verbale codering (en retrieval) en "imagery" codering (en retrieval). In haar "multi-stage-coding-and-retrieval-model" (a personal summary of the current state of research, op.cit. p. 117) worden deze twee verwerkingskanalen dan ook van elkaar onderscheiden.

Naast een "modality-free" lange-termijn-geheugenopslag postuleert Boekaerts dus "modality-specific" korte-termijn-geheugencodering en -retrieval. Argumenten voor deze "modality-specific" STM (short-term-memory) processen zijn enerzijds gebaseerd op introspectieve gegevens: Subjecten rapporteren subjectief zowel mentale voorstellingen als verbale codes; anderzijds zijn er ook aanwijzingen uit experimenteel onderzoek. Boekaerts (op.cit. p. 115) vermeldt bijvoorbeeld een onderzoek van Brooks, waaruit bleek dat reactietijden bij een visuele taak sneller waren wanneer een verbale response moest worden gegeven dan wanneer een antwoord moest worden aangewezen, terwijl het omgekeerde het geval bleek bij verbale informatie.

Kennelijk interfereerde de verbale reactie met verbale processen en niet met imageryprocessen, terwijl de ruimtelijk-visuele reactie interfereerde met imageryprocessen en niet met verbale processen. Andere onderzoekers toonden aan dat in het korte-termijn-geheugen (STM) twee afzonderlijke vormen van representatie bestaan die in bepaalde mate onafhankelijk van elkaar zijn (Kosslyn et al., 1976).

Boekaerts (op.cit. p. 150) veronderstelde, dat er ook individuele verschillen bestaan in de verwerkingskanalen die mensen prefereren. Sommigen zullen stimulusmateriaal of abstracte proposities uit het LTM (long-term-memory) bij voorkeur transformeren in verbale codes, terwijl anderen de voorkeur geven aan mentale voorstellingen. Zo zal de één wanneer hem/haar de weg wordt uitgelegd zich bij voorkeur een mentale voorstelling van het stratenplan vormen, terwijl de ander de informatie bij voorkeur codeert via zinnen als eerste rechts, tweede links, bij de stoplichten rechts (etcetera).

Hoewel de lange-termijnopslag volgens Boekaerts propositioneel van aard is zullen de verschillende vormen van codering - dan wel niet naar aard maar wel naar inhoud - verschillende informatie-opslag tot gevolg hebben. Alleen wanneer men de verbale wijze van codering heeft gebruikt, zal men bijvoorbeeld later in staat zijn zich letterlijke (in tegenstelling tot semantische) informatie te herinneren. En door een imagerycodering kunnen andere details in het LTM terecht komen dan door een verbale codering.

Om de individuele verschillen in voorkeur voor verbale dan wel imagerycodering op het spoor te komen, bedacht Boekaerts een experimentele taak, waarin twee dimensies konden worden gemeten: Het onderscheid tussen verbalizers en weak-verbalizers en het onderscheid tussen visualizers en weak-visualizers. Visualizers zijn degenen die in staat zijn visueel-ruimtelijke informatie in plaatjes of foto's alsmede de semantische informatie daarin te onthouden en weak-visualizers zijn diegenen die alleen de algemene inhoud van plaatjes of foto's onthouden. Verbalizers zijn diegenen die in staat zijn de letterlijke vorm van zinnen te onthouden alsmede de algemene inhoud van de zinnen, terwijl de weak-verbalizers alleen de algemene inhoud van zinnen onthouden en niet de letterlijke.

De gekozen experimentele taak is in wezen een interferentietetaak waarmee wordt voortgebouwd op het eerder besproken onderzoek van Brooks. Er wordt onderzocht in hoeverre iemand bij het coderen van verbale informatie wordt gehinderd doordat hij tegelijkertijd visuele informatie (zindimensie) moet coderen. En er wordt onderzocht in hoeverre iemand bij het coderen van visuele informatie er hinder van ondervindt, dat hij tegelijkertijd verbale informatie (diadimensie) moet coderen.

Boekaerts veronderstelde dat iemand die bij het verwerken van visueel stimulusmateriaal de beeldcode gebruikt geen last zal ondervinden van een verbale paralleltaak. Evenzo veronderstelde zij, dat iemand die bij het verwerken van verbaal stimulusmateriaal de verbale code gebruikt geen last zal ondervinden van een visuele paralleltaak. Om de mate van interferentie te bepalen, worden de prestaties op dubbeltaken vergeleken met prestaties op ongestoorde enkelvoudige taken.

In het eerste onderzoek werden eerst acht dia's gepresenteerd en vervolgens acht zinnen (single memory tasks), waarbij de subjecten (+ zestien jaar) zich moesten opstellen alsof zij spionnen waren die zoveel mogelijk details moesten onthouden. Vervolgens werden na een korte pauze de twee dubbele taken (dual memory tasks) uitgevoerd. Tegelijk met acht andere dia's moest nu een auditief gepresenteerde tekst over verschillende sporten worden beluisterd en tegelijk met acht andere zinnen

moest een film worden bekeken waarin allerlei dieren voorkwamen. Ter controle op de werking van de interferentietaken werd de proefpersonen gevraagd alle dieren uit de film en alle sporten uit de tekst op te schrijven. Na een retentieperiode van twee uur volgde de toets, waarin retentie van de details van de dia's en zinnen via meerkeuzevragen werd getoetst.

De resultaten van het eerste onderzoek van Boekaerts waren als volgt: Zeven proefpersonen werden geclassificeerd als indefinites (relatief veel verlies op beide dubbele taken), acht proefpersonen werden geclassificeerd als bicognitives (weinig of geen verlies op beide dubbele taken) en elf proefpersonen werden geclassificeerd als verbalizers (weinig verlies op de zinnen en relatief veel verlies op de dia's). Deze indeling in drie groepen correspondeerde met de op basis van introspectie verkregen gegevens.

In het tweede onderzoek werd de stabiliteit van de test onderzocht met behulp van een parallelversie. Op één na kwamen alle proefpersonen in dezelfde categorie terecht als in het vorige onderzoek en de test-hertest correlaties tussen de verschillende scores op de vier dimensies waren redelijk hoog.

In het derde onderzoek werd een aantal verbeteringen in de procedure ingevoerd: a) In plaats van de film en de tekst als interferentietak diende nu een "shadowing"-taak (het onmiddellijk meespreken met een auditief gepresenteerde zin) en een "tracing"-taak (het overtrekken van een tekening). De zinnen en tekeningen waren inhoudelijk gerelateerd aan de stimuli waarmee ze moesten interfereren. b) De toets werd veranderd: De dia's werden nu ook visueel getoetst door middel van tekeningen. Bij elke dia werden vier tekeningen gemaakt, waarvan er één inhoudelijk fout was, één helemaal goed en twee wel inhoudelijk goed maar qua "oppervlaktestructuur" fout waren.

Uit het derde onderzoek bleek dat er 14 bicognitives waren, 25 verbalizers, 70 indefinites en 3 visualizers (weinig verlies op de dia's en veel verlies op de zinnen). Nu werden derhalve - zij het in beperkte mate - ook visualizers aangetroffen. Itemanalyses leverden redelijke homogeniteitsindices per subtest op.

Een controleconditie, waarin de "shadowing"- en "tracing"-taken werden vervangen door respectievelijk een toon-detectie-taak en een taak waarbij getallen moesten worden geschreven, leverde geen significante verschillen op tussen de vier subtests. Dit resultaat betekent volgens Boekaerts dat de verschillen niet te danken zijn aan de dubbele taak op zich maar aan het blokkeren van het alternatieve kanaal.

Het onderzoek van Boekaerts en de daarbij aansluitende theorievorming vormen een goed uitgangspunt voor verder onderzoek naar individuele verschillen in coderingsstijl met name voor aptitude-treatment-interaction onderzoek. Er doen zich echter wel enkele problemen voor:

- a) In de theorie met betrekking tot coderingsstijlen is mijns inziens een tegenstrijdigheid geslopen: Enerzijds worden door Boekaerts in theorie verschillen in voorkeur voor het ene dan wel het andere verwerkingskanaal verwacht, anderzijds is de geconstrueerde test een vermogenstest (pp. 200, 208, 220). De test wordt in het empirisch deel ook geïnterpreteerd als een vermogenstest zoals ook blijkt uit het volgende citaat: "The results of this experiment support the conclusion that different pupils are differentially succesful at retrieving visuo-spatial and verbatim information." (p. 220).

De relaties tussen voorkeur en vermogen zijn meestal complex: Een vermogen kan tot een voorkeur leiden en een voorkeur tot een vermogen, maar nooit mag men uit een gebrek aan vermogen tot het ontbreken van een voorkeur besluiten of uit een voorkeur voor de ene coderingswijze tot een gebrek aan vermogen met betrekking tot de andere.

Voor een ATI-onderzoeker is het echter wel van groot belang om te weten of een leerlingkenmerk betrekking heeft op een vermogen of op een preferentie. Immers, gebrek aan vermogen kan worden gecompenseerd door bepaalde typen van instructie, terwijl men niet kan compenseren voor een tekort in voorkeur (zie hoofdstuk 2).

- b) Een ander probleem betreft het indelen van de proefpersonen in de vier categorieën (verbalizers, visualizers, bicognitives en indefinites) op basis van hun scores op de twee dimensies (de zindimensie en de diadimensie). Voor het omzetten van de twee continue dimensies in vier categorieën zijn "cut-off" criteria nodig: Op welke verschillen tussen prestaties op dubbele en enkelvoudige taken moet men de indeling baseren? En wanneer noemt men prestaties op de enkelvoudige taken hoog en wanneer laag?

De keuze van dergelijke criteria is willekeurig. De door Boekaerts gekozen criteria (verschillend voor de zinnen- en diadimensie op basis van de moeilijkheidsgraad van de items) zijn afhankelijk van de steekproef en bepalen volledig de verdeling over de verschillende categorieën. Bovendien is de aanname dat iemand die laag scoort op de diadimensie nooit een visualizer kan zijn, en iemand die laag scoort op de zindimensie nooit een verbalizer (op.cit. p. 218), op zijn minst aanvechtbaar. Daarom lijkt het mij veiliger de indeling in categorieën voorlopig achterwege te laten en alleen de twee dimensies als zodanig als maat te hanteren. Daarbij kan men dan overwegen proefpersonen die te laag scoren op de enkelvoudige taken (bijvoorbeeld volgens de verwachte waarde of lager) als onscorbaar buiten beschouwing te laten.

- c) Een derde probleem vormt de hoogte van de scores op de enkelvoudige taken: In de eerste twee experimenten waren deze wat betreft de dia's aan de hoge kant (er werden nauwelijks fouten gemaakt). Het plafondeffect maakt de verschilcores (enkelvoudige taak min dubbele taak prestatie) extra onbetrouwbaar.

In het derde experiment deed zich daarnaast ook nog het probleem voor dat de scores op de enkelvoudige zintaak aan de lage kant waren. Ook dit bodemeffect maakt de verschilcores nog onbetrouwbaarder dan ze normaal al zijn. Het eerste probleem zou men kunnen oplossen door meer dia's (en zinnen) als stimulusmateriaal te nemen en het tweede probleem is wellicht oplosbaar via vergemakkelijking van de zinnen en

verbetering van de instructie.

- d) Een volgend probleem betreft de "shadowing"-taak in het derde experiment. Hoewel deze taak mijns inziens een verbetering vormt ten opzichte van de stoortask in de eerste twee experimenten, omdat de blokkering van het verbale kanaal daardoor beter controleerbaar is, vormt de moeilijkheid ervan een nieuw probleem (zie ook Boekaerts op.cit. p. 215). De "shadowing"-taak vraagt voor onervaren proefpersonen een uitvoerige oefenperiode (Boekaerts, mondelinge mededeling).

Het is mijns inziens dan ook beter deze "shadowing"-taak te vervangen door een naspreektaak, waarbij korte zinnen pas nagesproken worden als ze "uitgeluisterd" zijn. Ook dan bereikt men het beoogde doel, namelijk blokkering van het verbale kanaal, terwijl deze taak gemakkelijker uitvoerbaar is.

- e) Een laatste probleem vormt de semantische interferentie tussen de hoofdtask en de "stoortask". De "stoortask" had betrekking op objecten die ook in de hoofdtask voorkwamen. Het is echter de vraag of dat zo gunstig is. Wellicht is het verstandiger deze semantische interferentie te vermijden.

7.3.3 Methode

7.3.3.1 Proefpersonen

Aan experiment 3 namen 81 tweede klas MAVO-leerlingen deel, afkomstig uit zeven klassen van drie verschillende MAVO's uit Tilburg. De leeftijd van de proefpersonen varieerde van dertien tot zeventien jaar, met een gemiddelde van 13.6. Er waren 43 meisjes en 38 jongens. De toewijzing aan de verschillende groepen geschiedde aselekt. 74 Proefpersonen deden aan alle vier de sessies mee: 34 in de experimentele groep en 40 in de controlegroep. (Gelijke aantallen proefpersonen in de twee condities werden verkregen (zie 5.3) door aselekt zes proefpersonen uit de controlegroep te verwijderen). Dit had een lichte stijging van de gemiddelde prestaties in de controlegroep tot gevolg.

7.3.3.2 Testmateriaal

In dit onderzoek werd de visualizers-verbalizers-test van Boekaerts afgenomen (zie 7.3.2), met enkele voor Nederland nodige aanpassingen (het woord "tas" betekent in Nederland bijvoorbeeld iets anders dan in België!), bestaande uit twee maal twaalf dia's en twee maal twaalf zinnen (in plaats van acht bij Boekaerts). De dia's en zinnen werden collectief gepresenteerd aan groepjes van maximaal twaalf personen. Eerst werden twaalf dia's gepresenteerd via een door een diastuurapparaat gestuurde diaprojector. De proefpersonen werd gevraagd zowel op ruimtelijk-visuele details (links-rechts, voor-achter, enzovoorts) als op betekenisaspecten te letten. De presentatietijd bedroeg per dia vier en een halve seconde en de wisseltijd anderhalve seconde. Vervolgens werden twaalf op een cassetteband opgenomen zinnen gepresenteerd via een hoofdtelefoon (zes seconden per zin). Ook nu werden de proefpersonen geïnstrueerd zowel op de betekenis als op de details (letterlijke vorm van de zin) te letten. Na een korte pauze werd geoefend met de verbale stoortaak. In de microfoon van een cassette-recorder moesten de proefpersonen via de hoofdtelefoon gepresenteerde zinnen nazeggen. Bij de verbaal gestoorde visuele taak moest informatie van twaalf dia's worden onthouden, terwijl tijdens de presentatie van elke dia een zin moest worden nagesproken. De zinnen hadden betrekking op andere onderwerpen dan de dia's. Tenslotte volgde de visueel gestoorde verbale taak. Tijdens de presentatie van zinnen (die nu moesten worden onthouden), moesten driedimensionale kubussen met afbeeldingen op de vlakken (per dia één kubus) worden overgetrokken.

Na een retentieperiode van twee uur volgde de toets. Deze toets bestond uit 48 meerkeuze-items, over elke zin en elke dia één item. Ten einde vragen over de dia's te kunnen stellen, werden telkens vier tekeningen waarop de dia's waren nagetekend, gepresenteerd. Naast het juiste alternatief waren er daarbij telkens drie afleiders: a) Een tekening waar de links-rechts relaties waren omgedraaid; b) Een tekening waarin het perspectief (focus of voor-achter relatie) was veranderd; c) een

tekening waarin zowel de links-rechts relaties als het perspectief waren veranderd.

Ook de vragen over de zinnen bestonden uit vier alternatieven¹⁴⁾: a) de letterlijke zin (de patiënt gaf de verpleegster een klap); b) een alternatief waarin de actor en het object omgedraaid waren (de verpleegster gaf de patiënt een klap); c) een alternatief waarin het werkwoord van actief passief was gemaakt (de verpleegster kreeg van de patiënt een klap); d) een alternatief waarin de veranderingen b) en c) beide waren uitgevoerd (de patiënt kreeg van de verpleegster een klap).

Op basis van deze toets konden zes scores per proefpersoon worden berekend: a) Het aantal juist onthouden ongestoorde dia's (range 0-12); b) het aantal juist onthouden ongestoorde zinnen (range 0-12); c) het aantal juist onthouden verbaal gestoorde dia's (range 0-12); d) het aantal juist onthouden visueel gestoorde zinnen (range 0-12); e) het verschil tussen a) en c) (range -12 tot +12); f) het verschil tussen b) en d) (range -12 tot +12).

7.3.3.3 Lesmateriaal

Het lesmateriaal en de metaforen waren in hoofdzaak gelijk aan het materiaal van experiment 1. Het enige verschil bestond uit een aantal verwijzingen in de tekst van de natuurkundecursus naar de metaforen. Zo werd bij een gedeelte over moleculen en atomen bijvoorbeeld tussen haakjes de volgende opmerking geplaatst: "(Vergelijk de moleculen met de kruimels en de atomen met de bestanddelen van speculaas)."

7.3.3.4 Toetsmateriaal

- a) Voortoets : 30 Meerkeuzevragen
- b) Drie natoetsen : Van elk tien meerkeuzevragen (waarvan vijf uit de voortoets)

¹⁴⁾ Sommige alternatieven komen in het dagelijks spraakgebruik minder voor dan andere. Dit vormt echter geen probleem, omdat de "onnatuurlijke" versies ook in het stimulusmateriaal voorkwamen.

- c) Retentietoets : 57 Items (alle voortoets en natoets-items + twaalf nieuwe items)
- d) Metafoortoetsen: Evenals in experiment 1 werden weer metafoortoetsen (twaalf items), vergelijkings-toetsen (vijftien items) en analogietoetsen (twaalf items) afgenomen. Verder werd een beoordelingsschaal gebruikt.

7.3.3.5 Procedure

De procedure is samengevat in tabel 7.20. De eerste sessie begon - na de kennismaking - met de afname van de voortoets (25 minuten), waarna werd begonnen met de bestudering van de lesstof. De leerlingen in de experimentele groep kregen eerst nog een instructie over het leren met behulp van metaforen. Hierin was een tweetal voorbeelden van metaforen in relatie tot lesstof uitgewerkt (bloedsomloop en centrale verwarming, en de bouw van de darmen in relatie tot de bouw van een radiator). Voor het eerste lesdeel (plus voor de experimentele groep de metaforen "speculaas" en "weegschaal") werd 50 minuten studietijd beschikbaar gesteld. Wie eerder klaar was noteerde zelf, onder strenge controle van de proefleiders, de eindtijd. Tenslotte werden natoets I en de metafoortoetsen afgenomen.

De sessies II en III waren vergelijkbaar met sessie I. Nu werd echter (op basis van de ervaringen in sessie I) slechts 30 minuten studietijd beschikbaar gesteld. De gereviseerde Boekaerts-test werd in groepen van maximaal twaalf leerlingen afgenomen voor de aanvang van de studiesessies en de toets vond na de sessie plaats. Voor elke sessie moest een groepje van twaalf leerlingen een half uur eerder komen en dezelfde leerlingen moesten na afloop een half uur langer blijven om de toets te maken.

In sessie I werd lesstof deel I bestudeerd, in sessie II lesstof deel II en in sessie III lesstof deel III. Tussen deze sessies verliep precies een week. De vierde sessie (retentiemeting) vond drie weken na afloop van sessie III plaats.

Tabel 7.20 Samenvatting van de procedure in experiment 3

Sessie	Conditie		
	Experimentele	Experimentele en Controle	Controle
Sessie 1	<ul style="list-style-type: none"> - Instructie m.b.t. het leren via metaforen - Plus de metaforen 1+2 - Metafoortoetsen 	<ul style="list-style-type: none"> - Voortoets (25 min.) - Lesbrieft 1 (50 min.) - Natoets 1 	<ul style="list-style-type: none"> - Boekaertstest (een deel) - Geen metaforen - Boekaertstoets (een deel)
Sessie 2 (1 week later)	<ul style="list-style-type: none"> - Plus metafoor 3 - Metafoortoetsen 	<ul style="list-style-type: none"> - Boekaertstest (een deel) - Lesbrieft 2 (30 min.) - Natoets 2 - HFT - Boekaertstoets (een deel) 	<ul style="list-style-type: none"> - Geen metaforen
Sessie 3 (1 week later)	<ul style="list-style-type: none"> - Boekaertstest (een deel) - Plus metaforen 4+5 - Metafoortoetsen - Beoordelingschalen - Boekaertstoets (een deel) 	<ul style="list-style-type: none"> - Lesbrieft 3 (30 min.) - Natoets 3 	<ul style="list-style-type: none"> - Geen metaforen
Sessie 4 (3 weken later)		<ul style="list-style-type: none"> - Retentietoets - Uitdelen cadeau-bonnen 	

Verschillen met experiment I betroffen:

- a) De methode van studietijdcontrolering: De efficiëntiemethode in plaats van de vast-aantal-trialsmethode (zie 6.5).
- b) Per sessie werd één deel van de lesstof bestudeerd, terwijl in experiment I alle lesstof in beide sessies werd doorgenomen.
- c) Er werd in de lesstof gerefereerd aan de metaforen.
- d) Er werden geen begripstoetsen afgenomen.
- e) Het leerlingkenmerk: De gereviseerde Boekaerts-test werd in plaats van de Spy Ring History Test afgenomen.
- f) De metaforen waren tijdens de studie van de lesstof beschikbaar.

7.3.3.6 Verwerking van gegevens

Er werden regressie-analyses uitgevoerd op de volgende afhankelijke variabelen: De totale natoets (som van de drie natoetsen) en de retentietoets. De hoofdeffecten werden op twee manieren getoetst.

7.3.3.7 Design

Evenals in de vorige twee experimenten was het design een pre-test-posttest-control-group-design met aselechte toewijzing aan de twee condities.

7.3.4 Resultaten van experiment 3

De gemiddelde score op de voortoets was iets boven de verwachte waarde ($M=8.82$, $SD=2.71$, bij een range van 0-30). De correlaties van de voortoets met de natoets en de retentietoets bedroegen respectievelijk .04 en -.07. Op basis van het beslissingsschema in figuur 5.2 werd de voortoets dan ook niet in de analyse betrokken. De descriptieve gegevens zijn weergegeven in bijlage III.

Hoofdeffecten: Hypothese 1

Volgens hypothese 1 zouden op de natoets en de retentietoets significante conditie-effecten moeten optreden. De resultaten in termen van percentages verklaarde variantie, F-waarden, ge-

middelen en standaarddeviaties zijn weergegeven in tabel 7.21. Noch op de natoets noch op de retentietoets waren er verschillen tussen de groepen. Er is dan ook onvoldoende steun voor hypothese 1.

Tabel 7.21 Overzicht van de percentages door het conditie-effect verklaarde variantie, de F-waarden en de gemiddelden en standaarddeviaties per condities op de verschillende afhankelijke variabelen (N=68)

	% var	F	\bar{X} exp	\bar{X} contr	SD exp	SD contr
Natoets (maximale score=30)	1.3	0.89	17.00	17.80	3.61	2.89
Retentietoets (maximale score=57)	0.0	0.00	25.59	25.79	7.23	6.08

Studietijden

In dit experiment was gekozen voor krappe studietijden (zie 7.3.1), maar zoals in 6.6 is uiteengezet, bestaat er geen garantie dat deze studietijden ook tijdens de uitvoering van het experiment worden gerealiseerd. Onverwacht bleken de studietijden in dit experiment toch ruim te zijn. Terwijl de maximale studietijd tijdens sessie I 50 minuten bedroeg, waren de effectieve studietijden gemiddeld slechts 32 minuten. Alle proefpersonen stakten de studie dan ook ruim voor tijd. Om toch nog krappe studietijden in de volgende twee sessies te bewerkstelligen werd de nominale studietijd in deze sessies teruggebracht tot 30 minuten. Hoewel de studiestof in de sessies II en III uitgebreider en moeilijker leek dan die in sessie I, had deze maatregel toch niet het gewenste effect. Weer bleken 65 procent (sessie II) en 67 procent (sessie III) van de effectieve studietijden korter dan de nominale studietijd.

In tabel 7.22 zijn de gemiddelde effectieve studietijden per conditie opgenomen. De leerlingen in de experimentele groep studeerden gemiddeld langer dan die in de controlegroep. In de

experimentele groep staakte in de tweede sessie 37 procent en in de derde sessie 47 procent van de leerlingen de studie voortijdig, terwijl deze percentages voor de controlegroep 97 procent en 88 procent bedroegen.

Tabel 7.22 Gemiddelde effectieve studietijden per conditie per sessie

	Nominale studietijd	Experimentele groep		Controlegroep	
		M	SD	M	SD
Sessie I	50	38.8	5.0	26.5	9.5
Sessie II	30	28.1	2.6	19.8	5.9
Sessie III	30	28.2	2.4	22.4	5.3
Totaal	110	95.0	7.1	68.8	18.7

Alle verschillen in effectieve studietijden tussen de experimentele en de controlegroep bleken significant op het .001 niveau. De F-waarden bij 1 en 66 vrijheidsgraden waren respectievelijk 55.2, 63.6, 31.5 en 59.9.

Interactie-effecten: Hypothese 2

De interacties tussen conditie en de verschillen op de zinnen en op de dia's konden niet voor de gehele steekproef worden getoetst omdat de scores op de ongestoorde zinnen en dia's van een relatief groot aantal leerlingen laag waren (zelfs onder de op basis van toeval verwachte score). Bij een lage score op de ongestoorde dia's of zinnen is het moeilijk om ten gevolge van de storing nog lager te scoren. De verschilcores (ongestoord min gestoord) zijn dan ambigu.

Daarom werden de gegevens van leerlingen die te laag scoorden op de ongestoorde zinnen en/of dia's afzonderlijk geanalyseerd. Als (willekeurig) criterium werd de score vijf genomen: De gegevens van alle leerlingen die op de zinnen en/of op de dia's op de ongestoorde versie lager dan vijf scoorden kwamen voor deze analyse in aanmerking. (Ook Boekaerts (1978, p. 215) analyseerde haar gegevens afzonderlijk voor hoog en laag scorenden op de primaire, ongestoorde taak.) De resterende leerlingen (N=44) bleken nogal ongelijk over de condities verdeeld

Tabel 7.23 Resultaten van de regressie-analyses met leerlingkenmerk, conditie en interactie als predictoren voor de 34 resterende leerlingen (df=30)

Predictor	df	Verschil op zinnen (Z)			Verschil op dia's (D)					
		Natoets % var	F	Retentietoets % var	F	Natoets % var	F	Retentietoets % var	F	
Totale model	3	27.5	3.80**	4.3	0.45	22.0	2.82*	2.2	0.22	
Leerlingkenmerk (Z of D)	1	0.3	0.12	3.2	1.00	0.0	0.00	0.5	0.15	
Conditie (C)	1	13.9	5.75**	1.0	0.31	12.5	4.81**	1.7	0.52	
Interactie (ZxC of DxC)	1	13.4	5.55**	0.1	0.02	9.4	3.63**	0.0	0.00	
Residu	30	72.5		95.7		78.0		97.8		

* .05 < p < .10

** .01 < p < .05

(17 in de experimentele groep en 27 in de controlegroep). De problemen met betrekking tot ongelijke frequenties werden vermeden door aselect tien proefpersonen uit de controlegroep te verwijderen. Deze manipulatie had geen invloed op de gemiddelde natoetsscore maar verlaagde de gemiddelde retentietoetsscore enigszins. Op deze 34 resterende gevallen werden de regressie-analyses met betrekking tot de verschillen op de zinnen en de dia's uitgevoerd (die opnieuw gestandaardiseerd werden voor deze groep). De resultaten hiervan zijn vermeld in tabel 7.23 en de regressiecoëfficiënten staan in tabel 7.24.

Tabel 7.24 Regressiecoëfficiënten in de experimentele en controlegroep voor de natoetsgegevens

Leerling-kenmerk	Constante (term) in experimentele conditie	Constante (term) in controleconditie	B voor leerling-kenmerk in experimentele conditie	B voor leerling-kenmerk in controleconditie
Vershil op de zinnen	16.09	18.49	1.85	-0.50
Vershil op de dia's	16.52	18.65	-0.86	1.01

Op de natoets werden significante interacties geconstateerd met betrekking tot het verschil op de zinnen ($F(1,30)=5.55$, $p<.05$) en met betrekking tot het verschil op de dia's ($F(1,30)=3.63$, $p<.10$). Deze interacties manifesteerden zich echter niet meer op de retentietoets.

Volgens de door Borich et al. (1976) voorgestelde procedure werd beproefd of de twee leerlingkenmerk-interacties hetzelfde deel of verschillende delen van de variantie verklaarden. Daartoe werden drie regressie-analyses gedraaid: Eén analyse waarin de interacties simultaan werden ingevoerd, één waarin eerst het verschil op de zinnen werd ingevoerd en pas daarna het verschil op de dia's en één waarin deze volgorde was omgedraaid.

In alle drie de analyses bleken beide interactietermen significant. Volgens de beslissingsregels van Borich et al.

(1976) kan uit deze analyses worden geconcludeerd, dat de interactietermen verschillende variantie verklaren. Daarom werden differentiatiegebieden voor de twee dimensies bepaald via het computerprogramma Hypati van De Leeuw (1977). De regressievlakken en de differentiatiegebieden zijn getekend in figuur 7.3. De regressievergelijking voor de controlegroep luidt: $Natoets = 18.52 - 0.50X + 1.16Y$ en voor de experimentele groep: $Natoets = 16.16 + 1.85X - 0.58Y$ (X =verschil op de zinnen; Y =verschil op de dia's). De projectie van de snijlijn op het X - Y vlak luidt: $Y = 1.35X - 1.36$ en de vergelijking van de grenscurve (een hyperbool) van de differentiatiegebieden:

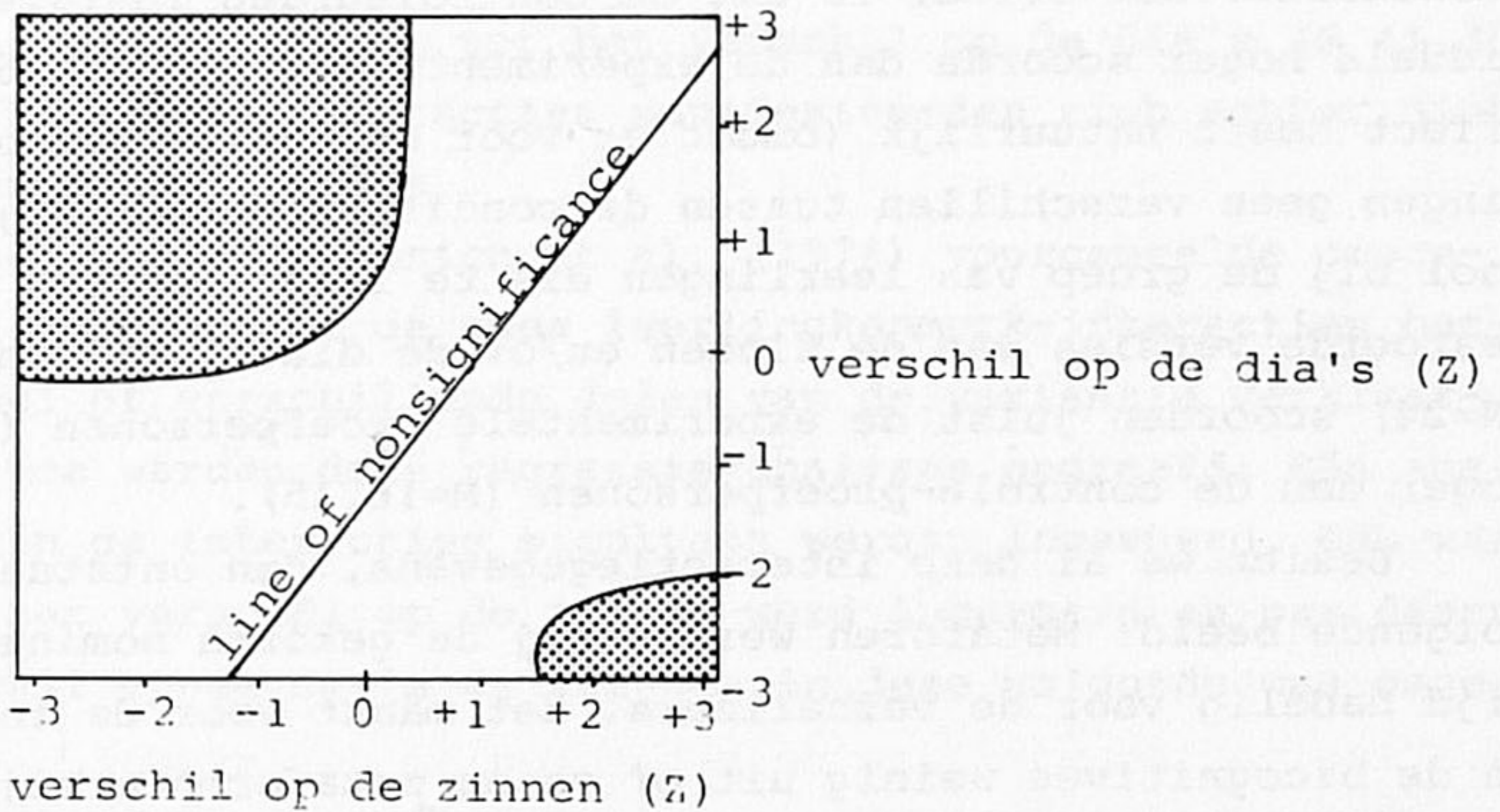
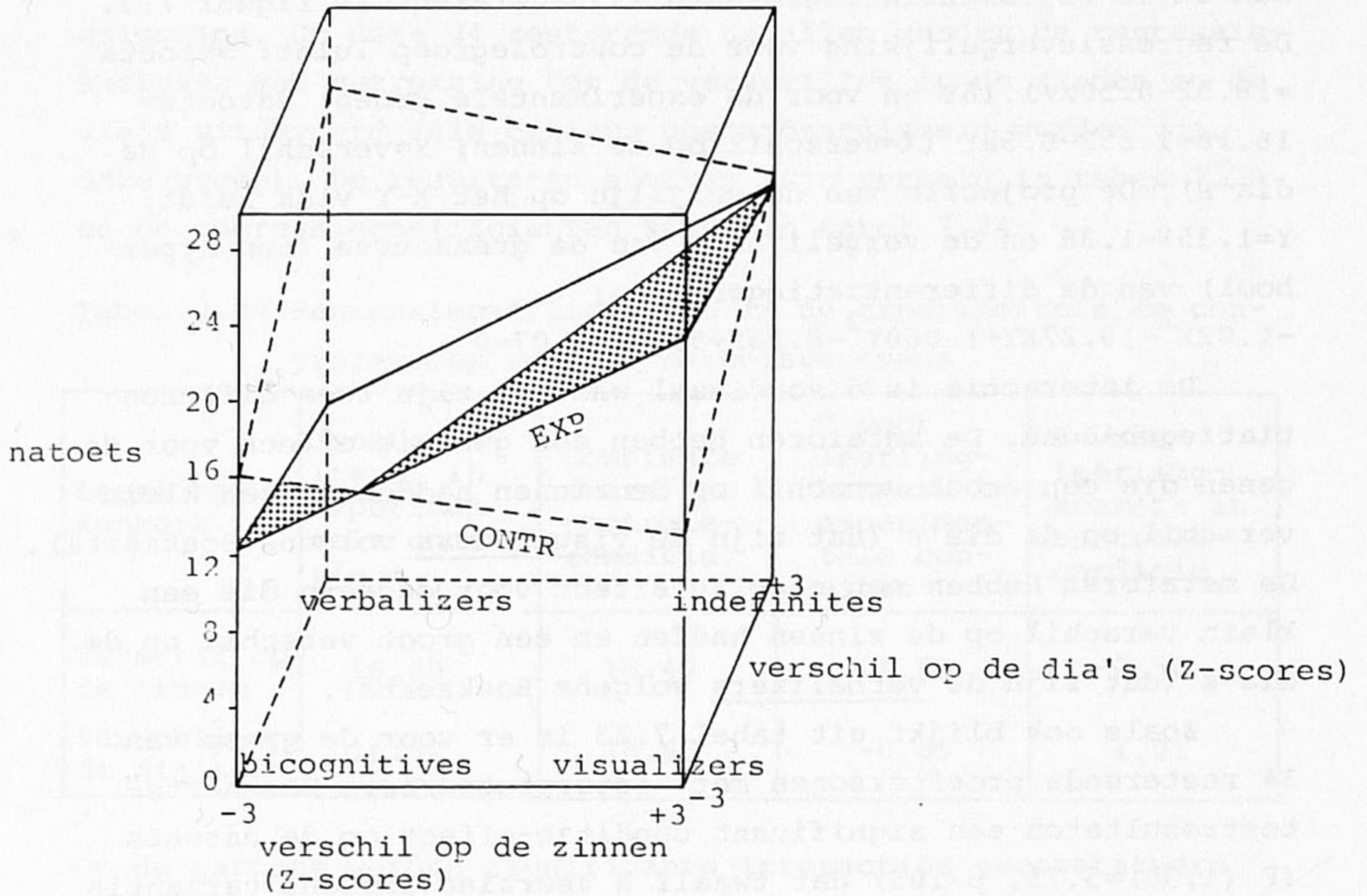
$$-1.02X^2 - 10.27XY + 1.060Y^2 - 8.28X + 9.20Y - 3.07 = 0$$

De interactie is disordinaal want er zijn twee differentiatiegebieden. De metaforen hebben een gunstig effect voor degenen die een groot verschil op de zinnen hadden en een klein verschil op de dia's (dat zijn de visualizers volgens Boekaerts). De metaforen hebben een nadelig effect voor degenen die een klein verschil op de zinnen hadden en een groot verschil op de dia's (dat zijn de verbalizers volgens Boekaerts).

Zoals ook blijkt uit tabel 7.23 is er voor de groep van 34 resterende proefpersonen met interpreteerbare Boekaertstestresultaten een significant conditie-effect op de natoets ($F(1,30) = 5.75$, $p < .05$) dat twaalf à veertien procent variantie verklaarde. Dit effect is dat de controlegroep ($M = 18.49$) gemiddeld hoger scoorde dan de experimentele groep ($M = 16.89$). Het effect heeft natuurlijk (omdat er voor de totale groep leerlingen geen verschillen tussen de condities waren) zijn tegenpool bij de groep van leerlingen die te laag scoorden op de ongestoorde versies van de zinnen en/of de dia's. Bij deze groep ($N = 24$) scoorden juist de experimentele proefpersonen ($M = 17.53$) hoger dan de controle-proefpersonen ($M = 16.15$).

Bezien we al deze interactiegegevens, dan ontstaat het volgende beeld: Metaforen werken bij de gekozen nominale studietijd nadelig voor de verbalizers. Het maakt voor de indefinites en de bicognitives weinig uit of ze nu metaforen erbij gebruiken of niet (zie de -3, +3 en de +3, +3 gedeelten van figuur 7.3). Voordelig zijn metaforen voor de visualizers en voor de-

Figuur 7.3 De interacties van de condities met de beide Boekaerts-test-dimensies (N=34), alsmede de differentiatiegebieden



genen die oninterpreteerbare Boekaerts-testresultaten hadden. Opvallend is echter, dat deze interacties beperkt bleven tot de natoetsgegevens en bij de retentiemeting verdwenen.

Overige resultaten: Metafoortoetsen

In tabel 7.25 staan de resultaten op de verschillende metafoortoetsen. Evenals in de vorige twee experimenten waren de resultaten betreffende de kennis-van-de-metafoortoetsen acceptabel. Ook de resultaten op de vergelijkings- en analogietoetsen waren in dit experiment redelijk goed. Kennelijk hebben de verwijzingen in de tekst naar de metaforen en het beschikbaar zijn van de metaforen bij de studie tot gevolg gehad, dat de prestaties op de analogie- en vergelijkingsvragen hoger waren dan in de vorige experimenten. Uit de resultaten op de beoordelingschaal (zie tabel 7.26) bleek dat het studeren via metaforen minder gunstig werd beoordeeld dan in de experimenten 1 en 2.

Tabel 7.25 Overzicht van de resultaten op de metafoortoetsen
(N=36)

Toets	Aantal items	Mediaan
Kennis-van-de-metaforen	15	11.9
Vergelijkingsvragen	12	7.8
Analogievragen	12	7.5

Tabel 7.26 De beoordeling van het studeren via metaforen door de leerlingen uit de experimentele groep (N=36)
(1=ja, erg; 5=nee, niet)

Vraag ¹⁵⁾	Mediaan
1. Zijn de metaforen een hulpmiddel bij het studeren?	1.8
2. Lijkt deze manier van studeren op de manier waarop je normaal studeert?	4.2
3. Vind je het leuk op deze manier iets te leren?	2.6
4. Vind je het moeilijk de metaforen en de lesstof te vergelijken?	3.3

¹⁵⁾ De vragen werden aan de leerlingen in iets andere vorm voorgelegd.

7.3.5 Discussie

De resultaten van experiment 3 kunnen als volgt worden samengevat:

- a) Bij de gehanteerde nominale studietijden waren er geen verschillen in studieprestaties tussen leerlingen die met en leerlingen die zonder metaforen studeerden.
- b) De effectieve studietijden bleken onverwacht toch korter dan de nominale studietijd, vooral bij de leerlingen in de controlegroep. De effectieve studietijden van de leerlingen in de controlegroep waren significant korter dan die van de leerlingen in de experimentele groep.
- c) De metaforen bleken wat de natoetsgegevens betreft in verschillende mate (in)efficiënt voor verschillende typen leerlingen: Voordelig voor visualizers en nadelig voor verbalizers.

In hoofdstuk 6 werden twee verschillende efficiëntievragen onderscheiden: Vraag c (zijn er prestatieverschillen bij een vaste nominale studietijd die voor alle leerlingen gelijk is?) en vraag d (zorgt extra materiaal voor efficiënter leren?). De antwoorden op deze twee vragen die op basis van de resultaten van experiment 3 kunnen worden gegeven, zijn verschillend.

Vraag c moet ontkennend worden beantwoord: Bij de gekozen nominale studietijden was het leren via metaforen niet efficiënter dan het leren zonder metaforen (en ook niet inefficiënter). Interpretatieprobleem B uit paragraaf 6.4 (het ontbreken van prestatieverschillen kan zijn ontstaan door een verkorting van de effectieve studietijden onder invloed van de metaforen) kan een kleine rol hebben gespeeld (namelijk bij de leerlingen in de experimentele groep, die voortijdig ophielden), maar erg waarschijnlijk is dat niet (gezien de tijdsverschillen tussen de groepen).

Vraag d lijkt op het eerste gezicht niet alleen ontkennend, maar zelfs in tegengestelde, ongunstige zin te moeten worden beantwoord: Het leren via metaforen lijkt in experiment 3 minder

Hoofdstuk 9 Algemene discussie

In dit hoofdstuk worden de resultaten van de zes empirische onderzoeken bediscussieerd in het licht van theoretische, methodologische en praktische implicaties. Bovendien worden de resultaten in verband gebracht met voorgaand onderzoek dat in hoofdstuk 4 werd besproken, en worden enkele suggesties voor vervolgonderzoek gedaan.

9.1 Theoretische implicaties

In hoofdstuk 3 werden drie theoretische functies onderscheiden die metaforen in een onderwijsleerproces kunnen vervullen: Het structureren en concretiseren van de informatie, en het stimuleren van een actiever verwerkings- en integratieproces. Verschillende van onze onderzoeksresultaten kunnen worden geïnterpreteerd in termen van deze functies.

Dat metaforen een structurerende functie vervullen, kan vooral worden afgeleid uit de resultaten op de relatietoets in experiment 2. Onder invloed van de metaforen bleken proefpersonen significant beter in staat relaties tussen begrippen te leggen die niet expliciet in de tekst aan bod waren geweest.

Dat metaforen een concretiserende functie vervullen, kan worden afgeleid uit de aptitude-treatment-interacties die in de experimenten 3 en 4 werden gevonden. In deze beide experimenten bleken de metaforen immers andere effecten te hebben bij visualizers dan bij verbalizers.

Dat metaforen een actiever verwerkings- en integratieproces kunnen bewerkstelligen, kan uit verschillende resultaten worden afgeleid. Ten eerste uit het resultaat dat metaforen juist tot hogere scores leidden op een begripstoets (experiment 1), een transfertoets (experiment 2) en een relatietoets (experiment 2). Ten tweede uit het resultaat dat metaforen tot langere leestijden leidden bij de eerste verwerking van de lesstof (experiment 1 en 2). Kennelijk was er bij die eerste verwerking extra tijd nodig om de lesstof en de metaforen met elkaar in verband te brengen.

De bovenbesproken resultaten met betrekking tot de theoretische functies van metaforen, betekenen steun voor de theorieën (van Ausubel en van Mayer) en hypothesen (van Ortony, Wittrock en Davidson) waaruit deze functies werden afgeleid. De superioriteit van het vergelijkend leren die in de experimenten 1, 2, 5 en 6 werd gevonden, betekent ook in algemene zin steun voor deze theorieën en hypothesen.

Onze bevindingen dat ook concrete organizers een gunstig effect op het leren kunnen hebben, betekenen dat de theorie van Ausubel uitbreiding behoeft in de zin van de in hoofdstuk 3 beschreven extrapolaties. Organizers behoeven niet altijd abstract te zijn, zoals Ausubel veronderstelde; ook concrete organizers kunnen nuttige functies vervullen.

De resultaten van experiment 5 met betrekking tot samenhangende en niet-samenhangende metaforen bleken niet in overeenstemming met onze hypothese dat samenhangende metaforen een betere "retrieval aid" zouden vormen dan niet-samenhangende. Dit resultaat kan men ook interpreteren als steun voor de "encoding hypothese" van Mayer (zie hoofdstuk 3). Metaforen lijken hun effect vooral ten tijde van de "encoding" uit te oefenen en niet ten tijde van de "retrieval". Deze "encoding hypothese" bleek in experiment 4 echter niet van toepassing bij comparative organizers, die, conform de voorspellingen van Ausubel, met name op een retentietoets op langere termijn tot hogere prestaties leidden.

In hoofdstuk 3 werden ook mogelijke nadelen van vergelijkend leren besproken. Ten aanzien van twee van deze nadelen werd in de experimenten informatie verzameld, namelijk ten aanzien van het eerste nadeel (metaforen brengen leerlingen op een dwaalspoor en in de war) en het vierde nadeel (metaforen kosten tijd die beter anders kan worden besteed).

(In verschillende experimenten (1, 2 en 3) bleek een aantal leerlingen onvoldoende in staat (of onvoldoende bereid) de juiste relaties te leggen tussen de metaforen en de lesstof. Dat wil zeggen dat in ieder geval een aantal leerlingen door de metaforen op een dwaalspoor werd gebracht.) Uit deze gegevens

kan echter niet worden afgeleid, dat ook met betrekking tot de lesstof door de metaforen extra verwarring bij de leerlingen is ontstaan.

Ten aanzien van het tweede argument tegen metaforen (metaforen kosten studietijd die beter anders kan worden besteed) werden in de experimenten vele gegevens verzameld. De resultaten met betrekking tot studietijden waren complex. Enerzijds werden er aanwijzingen gevonden voor het feit dat metaforen inderdaad extra studietijd vergen; niet alleen omdat ze op zich moeten worden bestudeerd (experimenten 1, 2, 3, 4 en 6), maar ook omdat ze een ander (actiever) verwerkingsproces induceren (experimenten 1, 2 en 3). Anderzijds waren er echter aanwijzingen dat dit tijdsverlies in latere fasen van de bestudering van de lesstof in ieder geval gedeeltelijk weer werd gecompenseerd (experimenten 1, 2 en 5) en dat de extra tijd die aan metaforen en hun vergelijking met de lesstof werd besteed geen verloren tijd was, in die zin dat de in controlegroepen extra ter beschikking gestelde tijd ofwel niet werd gebruikt (experiment 3) ofwel niet tot hogere prestaties leidde (experiment 4).

Concluderend kunnen we met betrekking tot dit argument stellen, dat metaforen inderdaad tijd kunnen kosten, maar dat er geen aanwijzingen zijn gevonden, dat die tijd op een andere wijze nuttiger zou kunnen worden besteed.

Tot slot van deze paragraaf nog enige opmerkingen over de correspondentie-analysestrategie (zie hoofdstuk 2). Deze strategie werd gepresenteerd als een heuristisch om zinvolle en inzichtelijke ATI-hypothesen te beredeneren. In die zin heeft de strategie aan zijn doel beantwoord. Dat wil echter niet zeggen, dat - via de strategie beredeneerde - zinvolle, inzichtelijke ATI-hypothesen te allen tijde empirisch zullen worden ondersteund. Daarvoor zijn de met ATI-onderzoek samenhangende methodologische problemen te groot (zie hoofdstuk 2). In onze experimenten werd dan ook soms geen steun gevonden voor volgens de strategie beredeneerde ATI-hypothesen (bijvoorbeeld experimenten 2 en 5). Vaker werd een veronderstelde interactie echter wel in de data aangetroffen (experimenten 1, 3 en 4). Deze was

dan echter niet altijd in de voorspelde richting (in experiment 1 bijvoorbeeld). In plaats van een activatie van een vaardigheid of stijl bleek er dan supplantatie op te treden. Het is ons gebleken, dat het erg moeilijk is om bij leerlingkenmerken als leerstijlen en coderingsstijlen te voorspellen of er bij bepaalde lesstof activatie dan wel supplantatie van die stijlen zal optreden (zie ook Simons et al., in voorbereiding).

Het verdient dan ook aanbeveling dat onderzoek wordt opgezet naar de vraag onder welke omstandigheden (met betrekking tot bijvoorbeeld leerstofkenmerken) activatie dan wel supplantatie (en/of interferentie) van deze of soortgelijke leerlingkenmerken zal optreden.

9.2 Vergelijking van de onderzoeksresultaten met die van voorgaande onderzoeken

In hoofdstuk 4 werden de resultaten van het voorgaande onderzoek samengevat in een achttal generalisaties. Enkele daarvan zijn in dit kader van belang, namelijk:

- A. Comparative organizers leidden tot hogere prestaties bij collegestudenten en middelbare scholieren.
- B. Concrete modellen leidden bij collegestudenten en bij vierde en zevende klassers soms tot hogere totaalresultaten, soms alleen tot hogere "far-transfer" en tot lagere "near-transfer" resultaten.
- C. In teksten verweven verbale analogieën leidden soms tot hogere prestaties bij basisschool-leerlingen en niet bij oudere proefpersonen.
- D. Het vergelijken van twee formeel vergelijkbare begrippensystemen leidde bij collegestudenten en middelbare scholieren tot transfer mits aan de volgende voorwaarden was voldaan:
 - a) De eerste serie gepresenteerde begrippen moest een assimilatieve context bieden voor de tweede, b) de eerste serie begrippen moest gemakkelijker worden geleerd dan de tweede (door concrete analogieën, illustraties of abstracte analogieën).
- E. Soms bleek het effect van analogieën afhankelijk van lesmaterialefactoren en specifiek voor bepaalde afhankelijke variabelen.

F. Soms traden ATI-effecten op.

Onze onderzoeksresultaten zijn over het algemeen in overeenstemming met deze generalisaties; soms vormen ze een uitbreiding van deze conclusies uit hoofdstuk 4 en een enkele maal zijn er tegenstrijdigheden.

In overeenstemming met generalisatie A (betreffende comparative organizers) werd in experiment 4 (bij psychologiestudenten) een significant verschil gevonden tussen een comparative organizer enerzijds en een biografische en een metaforische organizer anderzijds. In overeenstemming met generalisatie C (door een tekst verweven verbale analogieën) werden in experiment 5 (met studenten) geen significante verschillen gevonden tussen de metaforengroepen en een controlegroep. In overeenstemming met generalisatie E (lesmateriaalfactoren en afhankelijke variabelen) werden in experiment 2 alleen significante verschillen gevonden bij de transfertoets en de relatietoets en niet bij de kennis- en begripstoets.

De verschillen in resultaten tussen de experimenten 1 en 2 kunnen worden verklaard in termen van lesmateriaalfactoren: Voor experiment 2 werd de natuurkundecursus vereenvoudigd en beter gestructureerd, wat - in overeenstemming met generalisatie E - tot het verdwijnen van de kennis- en begripstoetseffecten aanleiding kan hebben gegeven.

De volgende onderzoeksresultaten vormen uitbreidingen op de eerder genoemde generalisaties:

- Ook organizers die bestaan uit vergelijkingen van begrippen uit de lesstof met niet-samenhangende potentiëel verwarrende dagelijkse betekenissen van deze begrippen leidden tot hogere retentie in experiment 4 (uitbreiding op generalisatie A).
- Concrete modellen leidden ook bij vijfde en zesde klassers van basisscholen tot hogere prestaties in experiment 6 (uitbreiding op generalisatie B).
- Transfer van een gemakkelijker te leren begrippensysteem naar een moeilijker begrippensysteem trad in de experimenten 1 en 2 ook op bij middelbare scholieren (2-MAVO, 1-VWO) (uitbreiding op generalisatie D).

- Deze transfer trad op van een qua aard gemakkelijker te leren begrippensysteem naar een moeilijker begrippensysteem (uitbreiding op generalisatie D).
- Er werden AT-interacties gevonden met leerstijlkenmerken en met coderingsstijlkenmerken (uitbreiding op generalisatie F).

De voornaamste verschillen tussen onze bevindingen en de genoemde generalisaties betreffen de volgende resultaten:

- Concrete modellen en begrippensystemen leidden niet tot hogere resultaten bij middelbare scholieren en studenten wanneer voor de extra tijd die daaraan en tengevolge daarvan werd besteed, werd gecontroleerd (experimenten 3 en 4) (verschillen met generalisaties B en D).
- Er werden geen ATI's gevonden met voorkennis of "verbal ability" zoals in de voorgaande onderzoeken het geval was (verschil met generalisatie F).

9.3 Methodologische implicaties

In de eerste drie experimenten werd aangetoond, dat directe en indirecte studietijdinvloeden een belangrijke - in voorgaande experimenten ongecontroleerde - rol kunnen spelen bij het tot standkomen van prestatie-effecten. In volgende onderzoeken naar vergelijkend leren zal dan ook aandacht moeten worden besteed aan studietijden. Omdat indirecte studietijdinvloeden van groot belang bleken te zijn (zie experimenten 1 en 2) is het bij het gebruik van ruime nominale studietijden te allen tijde aanbevelenswaard de effectieve studietijden te meten. Omdat gebleken is dat prestatie-effecten alleen optreden indien de toegestane studietijd ruim genoeg is om directe en indirecte studietijdsinvloeden een kans te geven, lijken onderzoeken waarin krappe nominale studietijden worden gebruikt weinig praktisch relevante informatie op te leveren. (Een uitzondering op deze aanbeveling moet worden gemaakt voor die omstandigheden waarin er gegevens bekend zijn over maximale of gebruikelijke studietijden.) De vast-aantal-trials methoden die in hoofdstuk 6 werden geïntroduceerd, vormen zinvolle aanvullingen op het arsenaal studietijdcontrole methoden dat in de literatuur werd aangetroffen. Men moet zich echter realiseren, dat ook aan deze

methoden nadelen kleven (bijvoorbeeld met betrekking tot ecologische validiteit en met betrekking tot testeffecten, zie hoofdstuk 6).

Een ernstig probleem dat naar voren kwam in de eerste drie experimenten, was dat proefpersonen niet bereid bleken erg lang achtereen te blijven studeren. Het gevolg daarvan was dat de gemiddelde studiestatistieken laag moesten worden genoemd. Overigens deed ditzelfde probleem zich ook voor in eerdere onderzoeken. In het overzichtsartikel van Mayer (1979b) worden bijvoorbeeld gemiddelde prestaties van slechts 35 à 45 procent goede antwoorden vermeld. Het probleem kan worden opgelost door het lesmateriaal in verschillende korte studietoelagen te laten bestuderen (vergelijk het verschil tussen "massed" en "distributed practice"). Dezelfde lesstof kan dan in verschillende toelagen meermalen worden herhaald. Een te goede bestudering kan natuurlijk aanleiding geven tot plafondefecten op eventuele toelagen. Bij retentiemetingen zullen deze zich echter niet zo snel voordoen.

Een ander probleem bij studietoelagencontroletoelagen vormt de ecologische validiteit. In de hoofdstukken 4 en 6 werd al gewezen op de gebrekkige ecologische validiteit van de "post organizer methode" en van de neutrale vergelijkingmethode. Maar ook de vast-aantal-trialsmethode (er worden waarschijnlijk zelden opdrachten gegeven lesmateriaal een bepaald aantal keren door te werken), de efficiëntiemethode en de ongelijke studietoelagenmethode (betreffende de keuze van de nominale studietoelagen) zijn niet altijd ecologisch valide. Het verdient dan ook aanbeveling in volgende onderzoeken nog andere methoden te beproeven, bijvoorbeeld de in hoofdstuk 6 besproken varianten op de onbeperkte studietoelagenmethode die een ecologisch tijds-, herhalings- of prestatiecriteria stellen.

Het meest kan mijns inziens echter worden verwacht van onderzoekopzetten met verschillende controlegroepen. Bijvoorbeeld één controlegroep die men evenveel tijd toestaat als de experimentele groep voor introductiemateriaal en lesmateriaal samen en één controlegroep die men evenveel tijd toestaat als de experimentele groep voor het lesmateriaal. Ook een opzet waar-

in twee experimentele groepen met twee verschillende nominale studietijden en twee controlegroepen zijn opgenomen, kan zinvolle nieuwe informatie opleveren.

Een andere belangrijke extra controlegroep is er één die het eigenlijke lesmateriaal niet hoeft te bestuderen, maar wel het extra introductiemateriaal (metaforen, organizers en dergelijke).

9.4 Praktische implicaties

Het uiteindelijke doel van onderwijspsychologisch onderzoek is te komen tot adviezen voor de praktijk. Dergelijke adviezen vloeien echter niet rechtstreeks voort uit een kleine serie experimenten. De experimenten hebben betrekking op bepaalde soorten lesmaterialen, bepaalde soorten metaforen, en een beperkt scala van leerlingen. Bovendien zijn er - ook al deden we nog zo ons best het onderzoek ecologisch valide te maken - verschillen tussen de experimentele- en onderwijssituaties. De onderstaande adviezen moeten dan ook met de grootst mogelijke voorzichtigheid geïnterpreteerd worden.

De praktische adviezen die in deze paragraaf worden geformuleerd, zijn voornamelijk gebaseerd op onze eigen experimenten, maar incidenteel wordt ook gebruik gemaakt van de resultaten van voorgaande onderzoeken (hoofdstuk 4). De adviezen worden gepresenteerd als antwoorden op een serie praktische vragen. De eerste vijf hebben betrekking op studietijden en werden eerder behandeld in hoofdstuk 6.

1) Hebben metaforen wel een effect op studieprestaties wanneer wordt afgezien van studietijdivloeden?

Het antwoord op deze vraag kan bevestigend zijn. In de experimenten waarin werd afgezien van tijdsinvloeden (experimenten 1, 2 en 6) werden significante verschillen ten gunste van de metaforenconditie gevonden.

2) Werken metaforen als introductie beter dan andere introducties?

Deze vraagstelling werd alleen onderzocht in experiment 4. Daar bleek één soort van metaforen niet tot betere prestaties te leiden, een andere (namelijk een comparative organizer) wel.

3) Wanneer een beperkte tijd beschikbaar is voor de bestudering van een onderwerp moeten er dan metaforen worden gebruikt?

Het antwoord op deze vraag is afhankelijk van de hoeveelheid beschikbare tijd. Als de tijd zeer krap is, kunnen waarschijnlijk beter geen metaforen worden gebruikt. Wanneer de tijd zo ruim is dat een deel van de proefpersonen wel en een ander deel niet in staat is de metaforen en de lesstof te bestuderen, maakt het wat betreft de studieprestaties niet uit of men al of niet metaforen gebruikt. Wanneer de studietijd nog ruimer is, verdient het misschien wel aanbeveling metaforen te gebruiken. Het geformuleerde antwoord op deze vraag is speculatief, omdat in het onderzoek (experiment 3) dat op deze vraag een antwoord had moeten bieden, de neiging van proefpersonen om al na korte tijd de studie te staken storend werkte.

4) Zijn metaforen efficiënte onderwijskundige hulpmiddelen en zijn er naast door tijdsinvloeden veroorzaakte effecten ook kwalitatieve effecten?

De meeste gegevens (experimenten 1, 3 en 6) wezen erop, dat vergelijkend leren niet efficiënt kan worden genoemd en dat wanneer de tijdseffecten en de prestatie-effecten tegen elkaar worden afgewogen, de balans uitslaat ten nadele van metaforen. Twee experimenten wezen echter toch in de richting van efficiëntie, namelijk de experimenten 2 en 5. In experiment 2 werden belangrijke kwalitatieve verschillen tussen de groepen gevonden bij niet zo erg grote tijdsverschillen. In experiment 5 hadden metaforen een gunstig (niet significant) effect op de prestaties en de tijdsbesteding. Overigens moet men zich realiseren dat het vergelijkend leren nieuw was voor veel proefpersonen en dat wellicht bij verdere training en oefening zal blijken dat vergelijkend leren wel een efficiënte vorm van leren is.

5) Welke invloeden hebben metaforen op de studietijdsbesteding?

Alle in hoofdstuk 5 onderscheid studietijdsinvloeden (directe verlenging, indirecte verlenging en indirecte verkorting) bleken in de experimenten naar voren te komen.

Metaforen moeten worden bestudeerd en dat kost soms (directe) tijd (experimenten 1, 2 en 6), soms ook niet (experiment 5). Soms zorgden metaforen ook voor een indirecte verlenging van de studietijden, met name bij de eerste lezing van de lesstof (experimenten 1 en 2) die soms ook weer werd gecompenseerd door een indirecte verkorting bij tweede en latere lezingen (experimenten 1 en 2). Over het totaal gezien, leken de metaforen de bereidheid om tijd te investeren in de studie te vergroten (experimenten 1, 2 en 3), maar het is natuurlijk de vraag in hoeverre die bereidheid ook bij regelmatig gebruik van metaforen zou blijven bestaan en vereist zou blijven.

6) Moeten er in het onderwijs nu zoveel mogelijk metaforen worden gebruikt bij het overdragen van informatie?

Deze vraag moet in algemene zin beslist ontkennend worden beantwoord. De empirische evidentie is nog beperkt en de onderzoeken hebben slechts betrekking gehad op bepaalde soorten lesstof. Bovendien zijn er nog onvoldoende gegevens bekend over de mogelijke nadelen van metaforen op korte en langere termijn (zie hoofdstuk 3) en ook over de studietijdeffecten (zie boven). Er zijn echter wel omstandigheden aan te geven waaronder een voorzichtig gebruik van metaforen zou kunnen worden aangeraden (zie onder).

7) Bij welke soorten leerstof is het gebruik van metaforen geïndiceerd?

Impliciet zijn veel onderzoekers er van uitgegaan, dat metaforen vooral van nut zijn bij meer exacte vakken als natuurkunde en het leren maken van computerprogramma's. De meeste onderzoeken hadden dan ook op dat soort leerstof betrekking. Dat wil echter niet zeggen, dat metaforen geen belangrijke functie

kunnen hebben bij andersoortige leerstof als vreemde talen, geschiedenis of psychologie (zie experiment 4).

De empirische evidentie beperkt zich echter tot de meer exacte vakken. Voorlopig kan alleen worden aangeraden metaforen te gebruiken bij vakken als natuurkunde, biologie, wiskunde en het leren maken van computerprogramma's.

Kenmerken van leerstof die aanleiding geven metaforen te bedenken, zijn blijkens onze experimenten de volgende: a) Leerstof die een moeilijk te doorgronden structuur heeft (bijvoorbeeld een ingewikkeld netwerk van begrippen en relaties daartussen) (zie experimenten 1 en 2); b) leerstof die dermate abstract is dat het moeilijk is zich daarbij mentale voorstellingen te vormen (experimenten 3 en 4); c) leerstof die nieuw is voor leerlingen (waarvoor zij nog geen assimilatieve context in hun cognitieve structuur hebben, waar de nieuwe leerstof bij kan aansluiten (experimenten 1 tot en met 6); d) leerstof waarbij het meer gaat om begrip, inzicht en transfer dan om kennis (experiment 2). Of metaforen ook bij andere leerstof(kenmerken) nuttig kunnen zijn, zal moeten blijken uit vervolgonderzoek.

8) Welke soort metaforen moet worden gebruikt?

De verschillende soorten metaforen werden nog maar zelden met elkaar vergeleken. In experiment 4 bleek een comparative organizer effectiever dan de (voetbalelftal)-metafoor. In experiment 5 bleken niet-samenhangende metaforen iets effectiever (maar niet significant) dan samenhangende metaforen. Wanneer we de robuustheid van de empirische evidentie ten gunste van de verschillende soorten metaforen bezien (hoofdstuk 4), dan bleken in teksten verweven analogieën meestal niet effectief, concrete modellen en systemen meestal wel en comparative organizers altijd. Deze gegevens leveren derhalve slechts weinig informatie waarmee de vraag kan worden beantwoord.

Misschien is de vraag ook wel niet goed en zou de volgende vraag beter zijn: Wanneer (dat wil zeggen bij welke leerstof(kenmerken) en met welke functies moeten welke soorten metaforen worden gebruikt? Op deze vraag kunnen tentatief bijvoorbeeld de volgende antwoorden worden gegeven: Comparative orga-

nizers kunnen het best worden gebruikt bij leerstof die sterk lijkt op reeds geleerde leerstof en daarmee kan worden verward. De organizer heeft dan de functie die verwarring in de cognitieve structuur vooral op lange termijn te voorkomen. Concrete modellen kunnen het best worden gebruikt bij leerstof die abstract en moeilijk voorstelbaar is. De concrete modellen vervullen dan de functie die voorstellingen toch mogelijk te maken. Begrippensysteemmetaforen en concrete modellen kunnen het best worden gebruikt bij leerstof die een ingewikkelde structuur heeft (om leerlingen te helpen die structuur te doorzien).

9) Welke typen leerlingen moeten wel metaforen worden aangeboden en welke niet?

Hoewel er - zowel in de voorgaande onderzoeken als in de onze - diverse aanwijzingen zijn gevonden dat metaforen voor bepaalde typen leerlingen meer van nut zijn dan voor andere, is voor geen van deze aanwijzingen voldoende empirische evidentie om de basis te vormen van een prescriptief antwoord op deze vraag. In voorgaande onderzoeken werden interacties gevonden met voorkennis en met verbale intelligentie. In onze experimenten werden deze interacties niet significant bevonden. De evidentie met betrekking tot de leerstijl- en coderingsstijlkenmerken van leerlingen die in de experimenten 1 tot en met 4 werd aangetroffen, was te beperkt en bleek onvoldoende herhaalbaar om daar nu al prescripties op te baseren. Het enige leerlingkenmerk dat relatief consistent van belang bleek in verschillende experimenten was het al dan niet beschikbaar hebben van een assimilatieve context in de cognitieve structuur. In onze experimenten werd dit kenmerk weliswaar niet gevariëerd, maar het was toch wel van belang: Immers de leerstof werd steeds zo gekozen, dat - naar veronderstelling - bij geen der leerlingen een assimilatieve context aanwezig was. Wellicht is het voorlopig het verstandigst om - wanneer men metaforen gebruikt in schriftelijk materiaal - de lezers te wijzen op de mogelijkheid deze metaforen over te slaan en aldus een keuzesituatie te creëren. Er is natuurlijk geen enkele garantie dat de vrijwil-

lige gebruikers ook degenen zijn die het meest van de metaforen profiteren. Ook ten aanzien van een ander leerlingkenmerk dat over de verschillende experimenten werd gevariëerd, namelijk de leeftijd, kunnen geen differentiatiemaatregelen worden voorgeschreven. De verschillende metaforen bleken zowel voor basisschool-leerlingen (experiment 6) als voor middelbare schoolieren (experimenten 1 en 2) als voor studenten effectief. Het minst was dit nog het geval voor de studenten (experimenten 4 en 5), maar voor de effectiviteit van verschillende soorten metaforen bij studenten was juist in de voorgaande experimenten (zie hoofdstuk 4) de meeste positieve evidentie verkregen.

10) Brengen metaforen leerlingen in de war?

Aan deze vraag werd reeds aandacht besteed in 9.2. Daar werd geconcludeerd, dat er wellicht enige zwakke aanwijzingen zijn dat metaforen voor verwarring bij sommige leerlingen zorgdragen. De gegevens waren echter poly-interpretabel.

9.5 Suggesties voor verder onderzoek

Verder onderzoek zou zich kunnen richten op de in dit proefschrift onbeantwoord gebleven vragen. Suggesties daartoe werden op diverse plaatsen in dit hoofdstuk en vorige hoofdstukken gedaan. Belangrijkste onbeantwoord gebleven vragen betreffen de ATI's. Verder onderzoek naar interacties met de leerstijlen van Pask en de coderingsstijltests van Boekaerts en Richardson lijkt gewenst. Het verdient daarbij aanbeveling vooraf onderzoek te doen naar betrouwbaarheid en validiteit van de betreffende tests of gebruik te maken van elders verricht testverbeteringsonderzoek (bijvoorbeeld onderzoek naar leer- en coderingsstijltests dat aan de Vrije Universiteit (Amsterdam), afdeling Functioneleer wordt gedaan).

Een andere slechts gedeeltelijk beantwoorde vraag is die naar de efficiëntie van vergelijkend leren. Er werden met betrekking tot deze vraag tegenstrijdige resultaten gevonden. Het verdient daarbij aanbeveling om nieuwe methoden van studietijdcontrole te beproeven (zie 9.3) en om maatregelen te treffen die

zorgen voor een betere bestudering van de leerstof door de leerlingen (bijvoorbeeld door dezelfde leerstof in verschillende korte sessies verschillende malen te laten bestuderen (zie 9.3)). Het is ons gebleken, dat het belangrijk is dat leerlingen niet voortijdig stoppen met studeren en dat hogere studieprestaties worden geleverd.

Verder onderzoek zou zich ook kunnen richten op de volgende - in de loop van het project opgekomen - vragen:

- Hoe en in welke mate maken leerlingen gebruik van vergelijkingen met andere verschijnselen of andere leerstof wanneer er geen metaforen worden gepresenteerd? Onderzoek naar deze vraag (zie hoofdstuk 4.1.6) zou aanwijzingen op kunnen leveren over nieuwe relevante leerlingkenmerken.
- Welke leerlingen maken vrijwillig gebruik van gepresenteerde metaforen en welke niet? Zijn de vrijwilligers tevens degenen die het meest van de metaforen profiteren? Er zijn inmiddels voldoende aanwijzingen waaruit blijkt dat sommige leerlingen meer profiteren dan andere en dat sommige leerlingen zelfs nadeel ondervinden van metaforen. Onduidelijk is echter nog gebleven voor welke typen leerlingen die voordelen en voor welke de nadelen overwegen. Kunnen leerlingen dit zelf bepalen?
- Vormen metaforen een effectief feedbackmiddel? Tot nu toe werden metaforen steeds vooraf aan of tijdens de bestudering van een tekst gepresenteerd (een uitzondering op deze regel vormen de postorganizer studies). Welk effect hebben metaforen wanneer ze alleen worden gepresenteerd als leerlingen fouten hebben gemaakt?
- Moet vergelijkend leren worden aangeleerd? Heeft een trainingsprogramma in vergelijkend leren effect op de mate waarin metaforen effect uitoefenen op leerprestaties op kortere en langere termijn? In onze experimenten bleken nogal veel leer-

lingen onvoldoende in staat en/of bereid de juiste relaties te leggen tussen de leerstof en de metaforen of concrete modellen. Ondanks dit gegeven hadden de metaforen toch een significant effect op studieprestaties. Hoe zal dit effect zijn als meer leerlingen met behulp van een trainingsprogramma hebben geleerd de juiste relaties wél te leggen? Zijn er met betrekking tot zo'n trainingsprogramma ATI's te verwachten?

- Welk effect hebben metaforen bij een langduriger gebruik ervan? In hoofdstuk 3.3 werd een aantal bezwaren tegen het gebruik van metaforen besproken. Eén van die bezwaren was dat metaforen op de lange duur een concrete denkattitude en -wijze zouden bestendigen. Treedt zo'n bestendiging inderdaad op? Andere aspecten die bij een langduriger gebruik van metaforen aandacht zouden vragen, zijn: Hoe vergaat het op langere termijn de tijdsinvloeden en de satisfactie van de leerlingen? Hoe gaat het met vergelijkend leren wanneer het nieuwe er af is?

Een meer algemene aanbeveling voor verder onderzoek betreft de afhankelijke variabelen die in het onderzoek worden betrokken. Zoals in experiment 2 het kunnen structureren van de leerstof werd geoperationaliseerd met behulp van de relatietoets, zouden ook nog andere afhankelijke variabelen moeten worden geconstrueerd waarmee kan worden nagegaan of de theoretische functies van metaforen daadwerkelijk worden vervuld. Leiden metaforen inderdaad tot betere mentale voorstellingen van de leerstof? Zorgen metaforen voor een actiever verwerkings- en integratieproces? Zie voor een mogelijkheid om de laatstgenoemde variabelen te operationaliseren het artikel van Mayer en Bromage (1980).

In the first part of the paper, the author discusses the general principles of the theory of the structure of the atom. The author then proceeds to a detailed analysis of the experimental results obtained by Rutherford and his colleagues. The author concludes that the results of these experiments are in agreement with the theory of the structure of the atom.

The author then discusses the implications of these results for the theory of the structure of the atom. The author concludes that the results of these experiments are in agreement with the theory of the structure of the atom. The author then discusses the implications of these results for the theory of the structure of the atom.

The author then discusses the implications of these results for the theory of the structure of the atom. The author concludes that the results of these experiments are in agreement with the theory of the structure of the atom. The author then discusses the implications of these results for the theory of the structure of the atom.

The author then discusses the implications of these results for the theory of the structure of the atom. The author concludes that the results of these experiments are in agreement with the theory of the structure of the atom. The author then discusses the implications of these results for the theory of the structure of the atom.

The author then discusses the implications of these results for the theory of the structure of the atom. The author concludes that the results of these experiments are in agreement with the theory of the structure of the atom. The author then discusses the implications of these results for the theory of the structure of the atom.

SUMMARYBy way of comparison: Experimental research on influences of metaphors on learning

This dissertation deals with the influence of metaphors on learning. In the first chapter the study is situated in the context of an information processing approach to instructional psychology that also emphasizes aptitude-treatment-interactions (ATI). The second chapter discusses ATI-research and the problem of how to find relevant combinations of aptitudes, treatments and learning tasks. Disappointing results of prior ATI-studies are, among other things, imputed to a lack of theorizing on this problem. Some proposed starting points for an ATI-theory are discussed, integrated and extended into eight principles for ATI. These principles are concretized into a heuristic ten-steps-procedure. In this strategy a very important distinction is made between structural and functional descriptions of learning tasks, student-characteristics and treatments. A functional description is one in which is specified how a variable (for instance a treatment-variable) might influence or change other variables (for instance a student-variable or a learning task variable).

In the next chapter functional descriptions of metaphors (as treatment-variables) are deduced from several psychological theories. According to this theoretical analysis metaphors might serve three functions:

- a. to concretize abstract learning tasks,
- b. to structurize unstructured learning tasks, and
- c. to induce a more active processing of learning material by learners.

Chapter Three is concluded with an overview of possible drawbacks of metaphors, concerning study-time, cognitive distortion and cultivation of a concrete way of thinking. Chapter Four reviews prior instructional studies on metaphors. Results of these studies were rather consistent. The number of studies, however, was small. Moreover, often too little attention was paid to reading-time.

In chapter Five some methodological problems are discussed. In this context a flow-diagram is presented on which decisions on the analysis of pre-test-data might be based. Finally, it is argued that in ATI-studies data might best be analyzed by generalized regression methods. Because this technique is not yet very well known in psychology and especially not in ATI-applications, several possible decisions concerning alternative procedures in generalized regression-analysis are discussed. A flow-diagram summarizes this discussion.

The sixth chapter concerns several ways to control and manipulate reading-time. Two indirect reading-time-effects remained uncontrolled in prior studies: the time devoted to the text proper might, depending on the additional material, be lengthened or shortened. These two indirect effects may inhibit or strengthen performance differences in an uncontrolled way, depending on the kind of research-question posed. Some new reading-time control-methods are described that could circumvent or solve some of the discussed problems.

In chapters Seven and Eight six experiments are reported. In these experiments influences of metaphors on the learning of different types of students under different reading-time-conditions were investigated. Hypotheses were derived from the functional descriptions of metaphors. In the first experiment the expected main effects of metaphors on learning science concepts by secondary school-pupils were significant. Performance-effects, however, were accompanied by indirect reading-time-effects: experimental subjects devoted significantly more time to the science-text than did control subjects. Statistical control of these time-effects led to a disappearance of some short-term performance-effects, long-term effects remaining significant. Data-analysis also revealed aptitude-treatment-interaction.

In the second experiment the effects of metaphors proved to be restricted to some dependent variables: short-term recall and long-term transfer. Short-term comprehension and long-term recall-effects were not significant. Experimental subjects also performed significantly better on a relation-test ("which concepts relate to each other and which do not?") than control subjects. This finding supported the expected structurizing function of metaphors. However, no further evidence was found for the prior found interaction. In this experiment the performance differences were independent from reading-time-effects, because experimental subjects compensated the lengthening of their first readings by shorter second reading-times.

While the first two experiments dealt with influences of metaphors on both performances and reading-time, the third experiment studied the influence of metaphors only on learning performances under restricted time-conditions. Control-subjects were permitted as much reading-time for the text as experimental subjects had for the text and the metaphors together. Under these conditions, metaphors did not lead to significant better performances. Results were interpreted in terms of reading-time-strategies. An ATI-effect was found concerning a visualizer-verbalizer-dimension.

In the last three experiments different kinds of metaphors, subject-matter, presentation-forms and student populations were studied. In experiment four, previous results as to effects of metaphors under restricted time conditions were confirmed. Unexpectedly, however, comparative organizers led to significantly, better long-term retention than both concrete metaphors and historical passages.

Experiment five showed consistent but non-significant effects of metaphors on both learning performances and reading-times. This result was attributed to a shortening of reading-times caused by the metaphors. The expected treatment-attitude-interactions did not show up. Non-interrelated metaphors led to somewhat better performances than interrelated ones but the difference was far from significant.

In the last experiment metaphors and subject matter were orally presented to elementary pupils. Learning of these young pupils also was facilitated by metaphors. Only scarce evidence was found for the hypothesized ATI between the conditions and verbal ability.

In the last chapter some theoretical, methodological and practical implications of the experiments were described. Finally, suggestions for further research on metaphors were given.

LITERATUUR

Anderson, J.R., & Bower, G.H.

Human associative memory. Washington: Winston, 1973.

Anderson, R.C.

How to construct achievement tests to assess comprehension.

Review of Educational Research, 1972, 42, 145-170.

Anderson, R.C.

Schema-directed processes in language comprehension. In A.M. Lesgold, J.W. Pellegrino, S.D. Fokkema, & R. Glaser (Eds.), Cognitive psychology and instruction. New York: Plenum Press, 1978.

Anderson, R.C., & Ortony, A.

On putting apples into bottles - a problem of polysemy.

Cognitive Psychology, 1975, 7, 167-180.

Anderson, R.C., Spiro, R.J., & Montague, W.E. (Eds.)

Schooling and the acquisition of knowledge. Hillsdale: Erlbaum, 1977.

Andre, Th.

Does answering higher-level questions while reading facilitate productive learning? Review of Educational Research, 1979, 49, 280-318.

Ausubel, D.P.

Educational psychology: A cognitive view. New York: Holt, Rinehart & Winston, 1968.

Ausubel, D.P.

In defense of advance organizers: A reply to the critics. Review of Educational Research, 1978, 48, 251-257.

Ausubel, D.P., & Fitzgerald, D.

The role of discriminability in meaningful parallel learning and retention. Journal of Educational Psychology, 1961, 52, 266-274.

Ausubel, D.P., Novak, J.D., & Hanasian, H.

Educational psychology: A cognitive view. New York: Holt, Rinehart & Winston, 1978.

Ausubel, D.P., & Youssef, M.

Role of discriminability in meaningful parallel learning. Journal of Educational Psychology, 1963, 54, 331-336.

Barnes, B.R., & Clawson, E.U.

Do advance organizers facilitate learning? Review of Educational Research, 1975, 45, 637-659.

Bartlett, F.C.

Remembering. Cambridge: University Press, 1932.

Bell, M.S., & Gagné, E.D.

Individual differences and the use of analogies in technical text. Paper presented at the meeting of the American Educational Research Association, San Francisco, april 1979.

Bielinski, L.S.

The effect of analogy structured teaching on student achievement in ninth-grade physical science. Dissertation Abstracts International, 1974, 34, 4855A-4856A.

Billow, R.M.

Metaphor: A review of the psychological literature. Psychological Bulletin, 1977, 84, 81-92.

Biskop, J.

Effecten van zelfgekozen oefenpatronen en gespreid herhalen. Onderzoekscriptie. Katholieke Hogeschool Tilburg, vakgroep onderwijspsychologie, 1978.

Blumenthal, A.L.

The process of cognition. New York: Prentice Hall, 1977.

Boekaerts, M.

Towards a theory of learning based on individual differences. Proefschrift. Katholieke Hogeschool Tilburg, 1978.

Bond, L., & Glaser, R.

ATI, but mostly A and T with not much I. Applied Psychological Measurement, 1979, 3, 137-140.

Borich, G.D., Godbout, R.C., & Wunderlich, K.W.

The analysis of aptitude-treatment interactions: Computer programs and calculations. Chicago: International Educational Services, 1976.

Bracht, G.H.

Experimental factors related to aptitude-treatment interactions. Review of Educational Research, 1970, 40, 627-645.

Brainerd, C.J.

Piaget's theory of intelligence. Englewood Cliffs: Prentice Hall, 1978.

Bransford, J.D., Nitsch, K.E., & Franks, J.J.

Schooling and the facilitation of knowing. In R.C. Anderson et al. (Eds.), Schooling and the acquisition of knowledge. Hillsdale: Erlbaum, 1977.

Bronfenbrenner, U.

The experimental ecology of education. Educational Researcher, 1976, 5, 5-15.

Bugelski, B.R.

Presentation time, total time and mediation in paired-associate learning. Journal of Experimental Psychology, 1962, 63, 409-412.

Burgmans, W.H.P.M.

De visualizers-verbalizers test van M. Boekaerts: Coderingsvoorkeur of coderingsvermogen. Doctoraalscriptie. Katholieke Hogeschool Tilburg, vakgroep onderwijspsychologie, 1979.

Campbell, D.T., & Fiske, D.W.

Convergent and discriminant validation by the multi-trait-multi-method matrix. Psychological Bulletin, 1959, 56, 81-105.

Campbell, D.T., & Stanley, J.C.

Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Carver, R.P.

A critical review of mathemagenic behaviours and the effect of questions upon the retention of prose materials. Journal of Reading Behaviour, 1972, 4, 93-119.

Case, R.

A developmentally based theory and technology of instruction. Review of Educational Research, 1978, 48, 439-463.

Cohen, J., & Cohen, P.

Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale: Erlbaum, 1975.

Corno, L.

A hierarchical analysis of selected naturally occurring aptitude-treatment interactions in the third grade. American Educational Research Journal, 1979, 16, 391-409.

Crombag, H.F.M.

ATI; perhaps not such a good idea after all. Tijdschrift voor Onderwijsresearch, 1979, 4, 176-183.

Cronbach, L.J.

The logic of experiments on discovery. In L.S. Shulman, & E.R. Keislar (Eds.), Learning by discovery: A critical appraisal. Chicago: Rand McNally, 1966.

Cronbach, L.J.

How can instruction be adapted to individual differences. In R.M. Gagné (Ed.), Learning and individual differences. New York: MacMillan, 1967.

Cronbach, L.J.

Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30, 116-127.

Cronbach, L.J., & Furby, L.

How we should measure change - or should we? Psychological Bulletin, 1970, 74, 68-80.

Cronbach, L.J., & Snow, R.E.

Aptitudes and instructional methods. New York: Irvington, 1977.

Crowder, R.G.

Principles of learning and memory. Hillsdale: Erlbaum, 1976.

Daniel, J.S.

Conversations, individuals and knowables: Toward a theory of learning. Engineering Education, 1975, 2, 415-420.

Davidson, R.E.

The role of metaphor and analogy in learning. In J.R. Levin, & V.L. Allen (Eds.), Cognitive learning in children. New York: Academic Press, 1976, 135-162.

Davis, G.A.

Research and development in training creative thinking. In J.R. Levin, & V.L. Allen (Eds.), Cognitive learning in children. New York: Academic Press, 1976.

Dijk, T.A. van

Semantic macrostructures and knowledge frames in discourse comprehension. In P. Carpenter, & M. Just (Eds.), Cognitive processes in comprehension. Hillsdale: Erlbaum, 1977.

Di Vesta, F.J.

Trait-treatment interaction, cognitive processes, and research on communication media. AV-Communication Review, 1975, 23, 185-196.

Domen, P., & Kreutzer, H.

Een verhaaltje meer of meer dan een verhaaltje. Onderzoeks-scriptie. Katholieke Hogeschool Tilburg, vakgroep onderwijspsychologie, 1979.

Elshout, J.J.

Karakteristieke moeilijkheden in het denken. Proefschrift, Universiteit van Amsterdam, 1976.

Entwistle, N.J.

Knowledge structures and styles of learning: A summary of Pask's recent research. British Journal of Educational Psychology, 1978, 48, 255-265.

Erve, M.F.J. van, & Timmermans, A.L.J.

Het effect van samenhang tussen metaforen op begrip en retentie van lesmateriaal. Onderzoeksscriptie. Katholieke Hogeschool Tilburg, vakgroep onderwijspsychologie, 1980.

Faw, H.W., & Waller, T.G.

Mathemagenic behaviours and efficiency in learning from prose. Review of Educational Research, 1976, 46, 691-720.

Feijen, R., Kleinmoedig, R., & Tooten, J.

Werken metaforen in het onderwijs prestatieverhogend en is dit afhankelijk van leerlingkenmerk en schooltype? Onderzoeks-scriptie. Katholieke Hogeschool Tilburg, vakgroep onderwijspsychologie, 1980.

Fijten, A., & Koene, M.

Het belang van metaforen voor het onderwijs. Onderzoeksscriptie. Katholieke Hogeschool Tilburg, vakgroep onderwijspsychologie, 1979.

Fitzgerald, D., & Ausubel, D.P.

Cognitive versus affective factors in the learning and retention of controversial material. Journal of Educational Psychology, 1963, 54, 73-84.

Flammer, A.

Individuelle Unterschiede im Lernen. Weinheim: Beltz Verlag, 1975.

Flavell, J.H.

The developmental psychology of Jean Piaget. New York: Van Nostrand Reinhold, 1963.

Flavell, J.H.

Cognitive development. Englewood Cliffs: Prentice Hall, 1977.

Frijda, N.

Memory processes and instruction. In A.M. Lesgold, J.W. Pellegrino, S.D. Fokkema, & R. Glaser, Cognitive psychology and instruction. New York: Plenum Press, 1978.

Gabriëls-van de Berg, J., & Siao, N.

De wisselwerking tussen de persoonlijke stijl van informatieverwerking en de vorm van de leerstofpresentatie: Een pilot-study. Doctoraalscriptie. Katholieke Hogeschool Tilburg, vakgroep onderwijspsychologie, ORC-nota 79-01, 1979.

Gagné, R.M.

The conditions of learning. New York: Holt, Rinehart & Winston, 1977.

Gick, M.L., & Holyoak, K.J.

Analogical problem solving. Cognitive Psychology, 1980, 12, 306-355.

Glaser, R.

Individuals and learning: The new aptitudes. Educational Researcher, 1972, 1, 5-13.

Glaser, R.

Components of a psychology of instruction: Toward a science of design. Review of Educational Research, 1976, 46, 1-24.

Glaser, R., & Resnick, L.B.

Instructional psychology. Annual Review of Psychology, 1972, 23, 207-276.

Glass, G.

Discussion. In M.C. Wittrock, & D.C. Wiley (Eds.), The evaluation of instruction. New York: Holt, Rinehart & Winston, 1970.

Groot, A.D. de

De betekenis van significantie bij verschillende typen onderzoek. Nederlands Tijdschrift voor de Psychologie, 1956, 11, 398-409.

Hartley, J., & Davies, I.K.

Pre-instructional strategies: The role of pretests, behavioral objectives, overviews and advance organizers. Review of Educational Research, 1976, 46, 239-265.

Haynes, A.R.

Metaphor as interactive. Educational Theory, 1975, 25, 272-277.

Heidt, E.U.

Media and learner operations: The problem of a media taxonomy revisited. British Journal of Educational Technology, 1977, 8, 11-26.

Hendrix, L.J., Carter, M.W., & Hintze, J.L.

A comparison of five statistical methods for analyzing pretest-posttest designs. Journal of Experimental Education, 1978, 47, 96-102.

Hermans, H.J.M.

Prestatie motivatie test. Amsterdam: Swets & Zeitlinger, 1968.

Hiller, J.H.

Learning from prose text: Effects of readability level, inserted question difficulty, and individual differences. Journal of Educational Psychology, 1974, 66, 202-211.

Hout-Wolters, B. van

Verbetering van tekstbestudering; een analyse van onderzoeksartikelen. Paper ten behoeve van de Onderwijsresearchdagen, Leiden, 1980.

Hunt, D.E.

The new three R's in Person-Environment Interaction: Responsiveness, reciprocity and reflexivity. Tijdschrift voor Onderwijsresearch, 1979, 4, 184-190.

Hunt, D.E., & Sullivan, E.V.

Between psychology and education. Hinsdale: The Dryden Press, 1974.

Hunt, E.

We know who knows, but why? In R.C. Anderson, R.J. Spiro, & W.E. Montague, Schooling and the acquisition of knowledge. Hillsdale: Erlbaum, 1977.

Hyman, R.

Impression formation, discrepancy from stereotype, and recognition memory. In D. Klahr (Ed.), Cognition and instruction. New York: Wiley, 1976.

Janssen, H.G.R.M., & Wagemans, L.J.J.M.

De vergelijking van advance organizer en metafoor als onderwijskundige hulpmiddelen. Onderzoeksscriptie. Katholieke Hogeschool Tilburg, vakgroep onderwijspsychologie, 1980.

Jaspers, J.P.

Het leren van een studietekst zonder organizers, met vooraf geplaatste organizers en met achteraf geplaatste organizers. Doctoraalscriptie. Universiteit van Amsterdam, vakgroep onderwijspsychologie, 1977.

Kanselaar, G.

Over stabiliteit van kognitieve processen en gedrag. Stellingen ten behoeve van Differentiatiestudiedag, Rijksuniversiteit Utrecht, maart 1979.

Kerlinger, F.N., & Pedhazur, E.J.

Multiple regression in behavioral research. New York: Holt, Rinehart & Winston, 1973.

Klausmeier, H.J.

Learning and human abilities. New York: Harper & Row, 1975.

Klerk, L.F.W. de

Aptitude treatment interaction. In Cremers et al., Losbladig onderwijskundig lexicon. Alphen aan de Rijn: Samson, 1978.

- Klerk, L.F.W. de
Inleiding in de onderwijspsychologie. Deventer: Van Loghum Slaterus, 1979.
- Klerk, L.F.W. de
 De invloed van item-feedback op het leerproces: Open versus gesloten items. Gedrag, 1979, 3, 131-143.
- Klerk, L.F.W. de
 ATI, perhaps a good idea after all. Tijdschrift voor Onderwijsresearch, 1979, 4, 191-194.
- Klerk, L.F.W. de, & Klerk, L. de
 The effect of knowledge of correct results per item on verbal learning and retention. Instructional Science, 1978, 7, 347-358.
- Klerk, L.F.W. de, Lodewijks, J.G.L.C., & Schoot, F.J. van der
 Onderwijspsychologie. Nederlands Tijdschrift voor de Psychologie, 1978, 33, 63-83.
- Koning, P. de
 Interne differentiatie. Pedagogische Studiën, 1974, 51, 105-124.
- Kooreman, H.J.
 Een analyse naar de mogelijkheid van een algoritmisch oplossingschema voor de spelling van werkwoordsvormen. Pedagogische Studiën, 1976, 53, 265-282.
- Kosslyn, S.M., Holyoak, K.J., & Huffman, C.S.
 A processing approach to the dual coding hypothesis. Journal of Experimental Psychology, 1976, 2, 223-233.
- Kremers, E.J.J.
Affectieve doelstellingen in het onderwijs en een toepassing voor het vak wiskunde (Citomemo nr. 277). Arnhem: Centraal Instituut voor Toetsontwikkeling, 1978.
- Kremers, E.J.J.
 Affectieve doelstellingen in het onderwijs. Pedagogische Studiën, 1980, 57, 87-105.
- Kulhavy, R.W., Dyer, J.W., & Silver, L.
 The effects of notetaking and test expectancy on the learning of text material. Journal of Educational Research, 1974-1975, 68, 363-365.

Landa, L.N.

Instructional regulation and control: Cybernetics, algorithmization and heuristics in education. Englewood Cliffs: Educational Technology Publication, 1976.

Laurillard, D.

The processes of student learning. Higher Education, 1979, 8, 395-410.

Lawton, J.T., & Wanska, S.K.

Advance organizers as a teaching strategy. Review of Educational Research, 1977, 47, 233-244.

Leeuw, L. de

Multipiele regressie-analyse - een alternatief voor variantie-analyse: Voor- en nadelen van beide methoden, in het bijzonder met betrekking tot de problematiek der veranderingsscores en aptitude-treatment-interactie-onderzoek. Intern rapport. Vrije Universiteit van Amsterdam, vakgroep functie-leer, Amsterdam, 1977.

Leeuw, L. de

Leren probleemoplossen. Proefschrift. Vrije Universiteit van Amsterdam. Lisse: Swets & Zeitlinger, 1979.

Levin, J.R., Rohwer, W.D., & Cleary, T.A.

Individual differences in the learning of verbally and pictorially presented paired-associates. American Educational Research Journal, 1971, 8, 11-26.

Linn, R.L., & Slinde, J.A.

The determination of the significance of change between pre- and posttesting periods. Review of Educational Research, 1977, 47, 121-150.

Lodewijks, J.G.L.C.

Over het aanleren van conceptuele netwerken door middel van uiteenlopende leerstofstructuren. Nederlands Tijdschrift voor de Psychologie, 1978, 33, 85-104.

Lodewijks, J.G.L.C.

Over wenselijkheid, noodzaak en haalbaarheid van differentiatie in het onderwijs: Positiebepaling. In H. Tillema (Ed.), Discussies rondom interne differentiatie, Rijksuniversiteit Utrecht, maart 1979.

Lodewijks, J.G.L.C.

Theorieën en methoden van leertaakbeschrijving en leertaak-analyse. Intern rapport 585-80-01., Katholieke Hogeschool Tilburg, vakgroep onderwijspsychologie, 1980.

Lodewijks, J.G.L.C., & Simons, P.R.J.

Onderwijspsychologie. In K. Soudijn, & H. Bergman (Eds.), Ontwikkelingen in de psychologie. Amsterdam: Boom/Intermediair, 1977.

Lodewijks, J.G.L.C., & Simons, P.R.J.

Een heuristische strategie ten behoeve van aptitude-treatment-interactie-onderzoek: Correspondentie-analyse. In W.J. Nijhof, & J. van Hout (Eds.), Differentiatie in het onderwijs. 's Gravenhage: Staatsuitgeverij, 1979.

Mayer, R.E.

Different problem-solving competencies established in learning computer programming with and without meaningful models. Journal of Educational Psychology, 1975, 67, 725-734.

Mayer, R.E.

Some conditions of meaningful learning of computer programming: Advance organizers and subject control of frame sequencing. Journal of Educational Psychology, 1976, 68, 143-150.

Mayer, R.E.

The sequencing of instruction and the concept of assimilation-to-schema. Instructional Science, 1977, 6, 369-388.

Mayer, R.E.

Advance organizers that compensate for the organization of text. Journal of Educational Psychology, 1978, 70, 880-886.

Mayer, R.E.

Can advance organizers influence meaningful learning? Review of Educational Research, 1979, 49, 371-383.

Mayer, R.E.

Twenty years of research on advance organizers: Assimilation theory is still the best predictor of results. Instructional Science, 1979, 8, 133-167.

Mayer, R.E., & Bromage, B.K.

Different recall protocols for technical texts due to advance organizers. Journal of Educational Psychology, 1980, 72, 209-225.

Merrill, M.D.

Learner control: Beyond aptitude-treatment interactions. AV-Communication Review, 1975, 23, 217-226.

Miller, R.M.

The dubious case for metaphors in educational writing. Educational Theory, 1976, 26, 174-181.

Neisser, U.

Cognition and reality. San Francisco: Freeman, 1976.

Netelenbos, J.B., & Rossum, J.H.A. van

Ontwikkeling van motorisch handelen: Competentie. In J. de Wit, H. Bolle, & J.M. van Meel (Eds.), Psychologen over het kind, deel 5. Groningen: Wolters-Noordhoff, 1979.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., & Bent, D.H.

Statistical package for the social sciences (2nd.ed.). New York: McGraw-Hill, 1975.

Nijhof, W.

Differentiëren of niet? Pedagogisch Tijdschrift, 1980, 1, 6-18.

Norman, D.A.

Notes toward a theory of complex learning. In A.M. Lesgold, J.W. Pellegrino, S.D. Fokkema, & R. Glaser (Eds.), Cognitive psychology and instruction. New York: Plenum Press, 1978.

Norman, D.A., Gentner, D.R., & Stevens, A.L.

Comments on learning schemata and memory. In D. Klahr (Ed.), Cognition and instruction. Hillsdale: Erlbaum, 1976.

Ortony, A.

Why metaphors are necessary and not just nice. Educational Theory, 1975, 25, 43-53.

Ortony, A.

On the nature and value of metaphor: A reply to my critics. Educational Theory, 1976, 26, 395-398.

Ortony, A.

The role of similarity in similes and metaphors. In A. Ortony (Ed.), Metaphor and thought. Cambridge: University Press, 1979.

Ortony, A.

Beyond literal similarity. Psychological Review, 1979, 86, 161-180.

Ortony, A., Reynolds, R.E., & Arter, J.A.

Metaphor: Theoretical and empirical research. Psychological Bulletin, 1978, 85, 919-943.

Paivio, A.

Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

Paivio, A.

Imagery and verbal processes. New York: Holt, Rinehart & Winston, 1971.

Parkhurst, P.E.

Generating meaningful hypotheses with aptitude-treatment interactions. AV-Communication Review, 1975, 23, 171-183.

Parreren, C.F. van

Enkele grondgedachten van de Utrechtse onderwijsproceeskunde. In G. Kanselaar en J. Peeck (Eds.), Strategieën in leer- en onderwijsprocessen. 's Gravenhage: Staatsuitgeverij, 1978.

Pask, G.

Conversation theory: Applications in education and epistemology. Amsterdam: Elsevier, 1976.

Pask, G.

Conversational techniques in the study and practice of education. British Journal of Educational Psychology, 1976, 46, 12-25.

Pask, G.

Styles and strategies of learning, British Journal of Educational Psychology, 1976, 46, 128-148.

Pask, G., & Scott, B.C.E.

Learning strategies and individual competence. International Journal of Man-Machine Studies, 1972, 4, 217-253.

Pask, G. et al.

Third progress report on SSRC research programme HR2708. Richmond, Surrey: System Research (Ltd.), 1977.

Pask, G. et al.

Fourth progress report on SSRC research programme HR2708.
Richmond, Surrey: System Research (Ltd.), 1978.

Pearson, P.D., Raphael, T., TePaske, N., & Hyser, C.

The function of metaphor in children's recall of expository passages. Paper presented at the meeting of the American Educational Research Association, San Francisco, april 1979.

Peeck, J.

Effects of prequestions on delayed retention of prose material. Journal of Educational Psychology, 1970, 61, 241-246.

Peeck, J.

Plaatjes in leerprocessen. Dissertatie. Utrecht: Elinkwijk, 1972.

Peeck, J.

Preinstructional strategies and extra reading time in learning from text. Tijdschrift voor Onderwijsresearch, 1977, 2, 202-207.

Peterson, P.L.

Aptitude x treatment interaction effects of teacher structuring and student participation in college instruction. Journal of Educational Psychology, 1979, 71, 521-533.

Plomp, Tj.

Enkele methodologische en statistische aspecten van ATI-onderzoek. Amsterdam: VOR-publicatie no. 5, 1977.

Pylyshyn, Z.W.

What the mind's eye tells the mind's brain: A critique of mental imagery. Psychological Bulletin, 1973, 80, 1-24.

Reed, S.K., Ernst, G.W., & Banerji, R.

The role of analogy in transfer between similar problem states. Cognitive Psychology, 1974, 6, 436-450.

Reigeluth, L.M., Merrill, M.D., & Bunderson, C.V.

The structure of subject matter content and its instructional design implications. Instructional Science, 1978, 7, 107-126.

Resnick, L.B.

Task analysis in instructional design: Some cases from mathematics. In D. Klahr (Ed.), Cognition and instruction. Hillsdale: Erlbaum, 1976.

Resnick, L.B., & Glaser, R.

Problem solving and intelligence. In L.B. Resnick, The nature of intelligence. New York: Erlbaum, 1976, 205-230.

Rhetts, J.E.

Task, learner and treatment variables in instructional design. Journal of Educational Psychology, 1974, 66, 339-347.

Richards, J.M.

A simulation study of the use of change measures to compare educational programs. American Educational Research Journal, 1975, 12, 299-311.

Richardson, A.

Mental imagery. New York: Springer Publication Company, 1969.

Richardson, J.T.E.

Mental imagery and memory: Coding ability and coding preference? Journal of Mental Imagery, 1978, 2, 101-116.

Rickards, J.P.; & August, G.J.

Generative underlining strategies in prose recall. Journal of Educational Psychology, 1975, 67, 860-865.

Rigney, J.W., & Lutz, K.A.

Effect of graphic analogies of concepts in chemistry on learning and attitude. Journal of Educational Psychology, 1976, 68, 305-311.

Rosenkrantz-Shapiro, K.

An overview of problems encountered in aptitude treatment interaction (ATI) research. AV-Communication Review, 1975, 23, 227-241.

Rothkopf, E.Z.

Barbarism and mathemagenic activities: Comments on criticism by Carver. Journal of Reading Behaviour, 1974, 6, 3-8.

Rothkopf, E.Z.

The sound of one hand plowing. Contemporary Psychology, 1978, 23, 707-708.

Royer, J.M., & Cable, G.W.

Facilitated learning in connected discourse. Journal of Educational Psychology, 1975, 67, 116-123.

Royer, J.M., & Cable, G.W.

Illustrations, analogies and facilitative transfer in prose learning. Journal of Educational Psychology, 1976, 68, 205-209.

Rummelhart, D.E., & Ortony, A.

The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale: Erlbaum, 1977.

Salomon, G.

Heuristic models for the generation of aptitude-treatment interaction hypotheses. Review of Educational Research, 1972, 42, 327-343.

Salomon, G.

Media and symbol systems as related to cognition and learning. Journal of Educational Psychology, 1979, 71, 131-148.

Salomon, G., & Clark, R.E.

Reexamining the methodology of research on media and technology in education. Review of Educational Research, 1977, 47, 99-120.

Scandura, J.M., & Wells, J.N.

Advance organizers in learning abstract mathematics. American Educational Research Journal, 1967, 4, 295-301.

Shibles, W.A.

The metaphorical method. Journal of Aesthetic Education, 1974, 8, 25-36.

Simons, P.R.J.

De potentiële betekenis van metaforen en analogieën in het onderwijs. Intern rapport. Katholieke Hogeschool Tilburg, vakgroep onderwijspsychologie, 1978.

Simons, P.R.J., Klerk, L.F.W. de, & Lodewijks, J.G.L.C.

Aptitude-treatment-interacties tussen veld(on)afhankelijkheid en instructiekenmerken. Nederlands Tijdschrift voor de Psychologie, in voorbereiding.

Simons, P.R.J., & Lodewijks, J.G.L.C.

ATI-onderzoek op basis van correspondentie-analyse. In W.J. Nijhof, & J. van Hout (Eds.), Differentiatie in het onderwijs. 's Gravenhage: Staatsuitgeverij, 1979.

Snow, R.E.

Representative and quasi-representative designs for research on teaching. Review of Educational Research, 1974, 44, 265-292.

Snow, R.E.

Research on aptitude for learning: A progress report. In L.S. Shulman (Ed.), Review of research in education IV. Itasca: Peacock, 1977.

Snow, R.E.

Eye fixation and strategy analyses of individual differences in cognitive aptitudes. In A.M. Lesgold, J.W. Pellegrino, S.D. Fokkema, & R. Glaser (Eds.), Cognitive psychology and instruction. New York: Plenum Press, 1978.

Span, P.

Interne differentiatie in het onderwijs gebaseerd op ATI: Uitgangspunten en realiseringkansen. Amsterdam: VOR-publicatie no. 5, 1977.

Sternberg, R.J.

Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, Erlbaum, 1977.

Sternberg, R.J.

Component processes in analogical reasoning. Psychological Review, 1977, 84, 353-378.

Sternberg, R.J., & Weil, E.M.

An aptitude x strategy interaction in linear syllogistic reasoning. Journal of Educational Psychology, 1980, 72, 226-239.

Truin, E.F., & Zagers, A.F.M.

Met de slagroomsput de klas in: Metaforen als didactische hulpmiddelen in het basisonderwijs. Onderzoeksscriptie. Katholieke Hogeschool Tilburg, vakgroep onderwijspsychologie, 1980.

Underwood, B.J.

Degree of learning and the measurement of forgetting. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 112-129.

Vastenhouw, J.

Blokken op wakken. In G.J. Mellenbergh, R.F. van Naersen, & H. Wesdorp (Eds.), Rede als richtsnoer. 's Gravenhage: Mouton, 1979.

West, L.T.T., & Fensham, P.J.

Prior knowledge or advance organizers as effective variables in chemical learning. Journal of Research in Science Teaching, 1976, 13, 297-306.

Wickelgren, W.A.

Learning and memory. Englewood Cliffs: Prentice Hall, 1977.

White, K., Sheehan, P.W., & Ashton, R.

Imagery assessment: A survey of self-report measures. Journal of Mental Imagery, 1977, 1, 145-170.

Wittrock, M.C.

The generative process of memory. In M.C. Wittrock (Ed.), The human brain. Englewood Cliffs: Prentice Hall, 1977.

Wittrock, M.C.

The cognitive movement in instruction. Educational Researcher, 1979, 8, 5-11.

Wittrock, M.C., & Lumsdaine, A.A.

Instructional psychology. Annual Review of Psychology, 1977, 28, 417-459.

Wolters, L.A.M.M.

Vragen over vragen. Intern rapport. Katholieke Universiteit Nijmegen, Instituut voor onderzoek van het wetenschappelijk onderwijs, 1975.

BIJLAGE 1

Descriptieve gegevens met betrekking tot experiment 1 (N=61)

Variabele	M ¹⁾	SD ¹⁾	Correlaties met							10		
			1	2	3	4	5	6	7		8	9
1. Operation- learning	35.51	13.40	-	.17	.15	.19	.20	.18	.41	.08	.13	.14
2. Comprehension- learning	42.00	15.14	-	-	.17	.06	.27	.16	.35	.07	.19	.18
3. HFT	3.61	2.33	-	-	-	.02	.02	-.11	.05	.01	.13	.10
4. Begripstoets 1x lezen	7.09	2.77	-	-	-	-	.50	.28	.48	.31	.05	.20
5. Begripstoets 3x lezen	7.90	2.77	-	-	-	-	-	.37	.50	.12	.17	.19
6. Begripstoets na 3 weken	6.25	2.57	-	-	-	-	-	-	.48	.29	.20	.30
7. Kennistoets na 3 weken	13.97	4.75	-	-	-	-	-	-	-	.33	.24	.36
8. Leestijd 1e keer	22.72	4.52	-	-	-	-	-	-	-	-	.19	.68
9. Leestijd 2e keer	35.08	6.28	-	-	-	-	-	-	-	-	-	.85
10. Leestijd totaal	57.80	8.39	-	-	-	-	-	-	-	-	-	-

325

1) De hier gerapporteerde gemiddelden en standaarddeviaties zijn de ongestandaardiseerde.
De gestandaardiseerde gemiddelden en standaarddeviaties van de leerlingkenmerken (1 t/m
3) zijn natuurlijk 0 en 1.

BIJLAGE III

De descriptieve gegevens met betrekking tot experiment 3 (N=34-68)

Variabele	N	M	SD	Correlaties met									
				1	2	3	4	5	6	7			
1. Natoets	68	17.40	3.27	-									
2. Retentietoets	68	25.69	6.63	.57	-								
3. Studietijd	68	82.00	19.35			-							
4. HFT	68	70.78	48.55				-						
5. Voortoets	68	8.82	2.71					-					
6. Verschil op zinnen	34	2.09	2.45						-				
7. Verschil op dia's	34	1.79	2.45							-			

BIJLAGE IV

De descriptieve gegevens met betrekking tot experiment 4 (N=53-69)

Variabele	N	M	SD	Correlaties met																	
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Voortoets	69	7.43	5.44	-	.46	.50	.27	-.06	.04	-.04	-.25	.20	.00	.02	-.07	-.02	-.10	.16	-.05	-.16	.02
2. Natoets	69	24.09	6.17		-	.75	.30	-.21	-.16	-.26	-.02	.03	.01	-.03	.02	.10	.03	.13	-.25	.01	
3. Retentie- toets	69	19.59	7.23			-	.32	-.17	-.17	-.27	-.14	.23	.03	.11	-.07	.02	.01	-.07	-.01	-.30	.01
4. Open vragen	69	6.14	2.93				-	.16	.10	-.29	-.18	-.04	.11	.04	.11	-.15	-.08	-.27	-.07	.04	-.15
5. Coding- test- woorden	53	0.23	0.25					-	.75	-.27	-.23	.30	.44	.25	.07	-.13	-.09	-.03	.07	.02	-.00
6. Coding- test- tekeningen	53	0.44	0.32						-	-.28	-.06	.27	.47	-.07	-.05	-.09	-.06	.14	.04	-.02	-.01
7. Boekaerts- test- zinnen	55	2.51	2.21							-	.35	.13	.00	.01	-.16	-.23	.08	.03	.16	-.05	.19
8. Boekaerts- test- dia's	55	1.60	2.42								-	-.00	-.01	-.26	-.01	-.14	.05	.05	-.22	-.15	-.04
9. Coderings- voorkeur	69	80.12	12.88									-	.53	.11	-.21	-.07	-.11	-.19	-.21	-.26	.17
10. Stimulus- voorkeur	69	33.84	5.74										-	.05	-.03	-.27	-.12	.10	-.06	-.11	.04
11. Parentest	53	3.85	3.24											-	-.15	-.01	-.15	-.01	-.09	-.08	-.16

12. TLWV	53	39.55	9.31	-	.17	.18	.14	-.02	.12	.09
13. Flags	53	54.05	21.56	-	-.03	.28	-.12	.09	-.10	
14. Identical forms	53	67.34	82.83	-		.05	.25	-.11	-.09	
15. Connected figures	53	66.47	19.52	-		-.14	.07	-.18		
16. Prestatie-motivatie	69	15.14	5.90	-			.03	.03		
17. F+plus	69	8.54	5.03	-					-.27	
18. F-min	69	12.48	6.04	-						

COLLECTIEF: WEF

DE QUANTITIEVE ZEKERHEID VAN DE TOEGANG TOT DE COLLECTIEF: WEF

BIJLAGE V

De descriptieve gegevens met betrekking tot experiment 5 (N=110)

Variabele	M	SD	Correlaties met										
			1	2	3	4	5	6	7	8	9	10	11
1. Natoets	10.84	3.34	-	.72	.63	.93	.55	.64	.02	.42	-.18	-.26	-.13
2. Retentietoets	10.29	3.34		-	.47	.66	.75	.90	.03	.36	-.25	-.24	-.15
3. N-fartransfer	1.73	1.29			-	.30	.55	.30	-.08	.25	-.19	-.12	-.11
4. N-neartransfer	9.11	2.71				-	.42	.65	.06	.40	-.14	-.27	-.10
5. R-fartransfer	2.34	1.56					-	.39	-.07	.28	-.28	-.17	-.22
6. R-neartransfer	7.95	2.40						-	.09	.32	-.16	-.22	-.06
7. Tijd	26.05	6.43							-	-.08	.15	.08	-.00
8. Wiskunde-opleiding	0.58	0.50								-	-.47	-.28	-.06
9. Angst voor wiskunde	2.49	0.69									-	.12	.49
10. Plezier in wiskunde	3.56	0.68										-	.25
11. Relevantie van wiskunde	2.79	0.46											-

BIJLAGE VI

De descriptieve gegevens met betrekking tot experiment 6 (N=145)

Variabele	M	SD	Correlaties									
			1	2	3	4	5	6	7	8	9	10
1. Analogietest	9.50	4.09	-	.26	.39	.46	.18	.21	.30	.32	.31	.38
2. Voortoets	9.17	2.45	-	-	.21	.38	.71	.78	.23	.10	.30	.27
3. Natoets	15.64	2.61	-	-	-	.46	.16	.16	.80	.79	.32	.37
4. Retentietoets	13.11	2.92	-	-	-	-	.25	.31	.33	.40	.70	.79
5. Voortoets 1	5.00	1.55	-	-	-	-	-	.10	.25	-.00	.30	.08
6. Voortoets 2	4.17	1.74	-	-	-	-	-	-	.10	.15	.15	.30
7. Natoets 1	7.02	1.66	-	-	-	-	-	-	-	.26	.28	.22
8. Natoets 2	8.62	1.63	-	-	-	-	-	-	-	-	.23	.37
9. Retentietoets 1	6.38	1.79	-	-	-	-	-	-	-	-	-	.12
10. Retentietoets 2	6.73	2.10	-	-	-	-	-	-	-	-	-	-

BIJLAGE VII

Lijst van gebruikte afkortingen en symbolen

df	= degrees of freedom, vrijheidsgraden
F	= de bekende toetsingsgrootte die bestaat uit een verhouding tussen twee varianties
\bar{x}	= gemiddelde van x
contr, cont	= controlegroep
exp	= experimentele groep
SD	= standaard deviatie
% var	= percentage variantie
p	= kans van voorkomen
M	= gemiddelde
cond	= conditie
B	= richtingscoëfficiënt
α	= significantieniveau
N	= number, aantal proefpersonen
dl	= deel
n.s.	= niet significant
Rep	= reproductie
Prod	= productie