

## Toetsing van procesvaardigheden bij natuurwetenschappen <sup>1)</sup>

I.de Bruyn  
Technische Hogeschool Twente  
Afdeling Technische Natuurkunde

### Summary

*In Dutch education the natural sciences are not implemented in integrated form, but as separated disciplines: biology, chemistry and physics. In this study some links and differences between the disciplines were investigated.*

*A test of science process skills was administered to 14 year old students. The abilities included reading graphs, reading tables and judgement of experimental design. Some additional tests were given to assess the knowledge of biology, chemistry and physics.*

*The analysis revealed no interpretable clustering of items, indicating that a scale for the abilities can not be distinguished. Furthermore the achievement of students in separate disciplines does not influence the response on process skill items.*

*The whole test showed a close relationship to a physics test, i.e. comparable sex-bias and dependence on mathematics ability were found. However, on the level of school type a difference with physics test performance was found.*

*The students who followed an experimental curriculum in physics (PLON) showed significant higher level of science process skills.*

### Inleiding

In het voorjaar van 1984 werd een survey-onderzoek verricht bij leerlingen in de 3e klas van het voortgezet onderwijs naar de prestaties bij natuurwetenschappen. Dit onderzoek maakte deel uit van een internationale studie (IEA). Als onderdeel van het Nederlandse onderzoek werd een studie gemaakt van de 'proces-vaardigheden' bij natuurwetenschappen.

Procesvaardigheden hebben te maken met de wijze waarop kennis - van de natuurwetenschappen - wordt vergaard. Deze vaardigheden worden verondersteld minder gebonden te zijn aan een van de natuurwetenschappen dan b.v. inhoudelijke kennis. (AAAS, 1965). Daarom krijgen procesvaar-

digheden veel aandacht bij onderzoek en leerplanontwikkeling voor b.v. natuuronderwijs.

Wij onderzochten de beheersing van de volgende vaardigheden:

- lezen van grafieken en het trekken van conclusies hieruit
- lezen en interpreteren van gegevens in de vorm van tabellen
- beoordelen van de opzet van proeven, o.a. het constant houden van variabelen.

Deze vaardigheden houden verband met de zogenaamde formele operaties van Piaget (Inhelder en Piaget, 1958). Het is lastig deze operaties klassikaal te onderwijzen of te testen. Een onderzoek van Padilla, Okey en Dillashaw (1983) toonde een duidelijke correlatie aan (ca .70) tussen de beheersing van Piaget-operaties en van procesvaardigheden. Vandaar dat systematische aandacht voor procesvaardigheden, zoals wij die b.v. vinden in leerplannen als PLON en WEI, mede de beheersing van Piaget-operaties kan dienen. Het aantonen van een relatie tussen leerwinst op formele operaties en het curriculum is echter nog niet gelukt. (Padilla, Okey en Garrard, 1984).

Het begrip 'procesvaardigheden' moet dus worden onderscheiden van de formele operaties van Piaget, hoewel er wel verband bestaat. De procesvaardigheden staan veel dicht bij concreet onderwijs. Wanneer toetsing van procesvaardigheden plaatsvindt, wordt steeds ook een natuurwetenschappelijke context aangeboden. Men kan dan verwachten, dat leerlingen die de betreffende context al kennen, gemiddeld hoger scoren op deze vragen: De inhoudelijke kennis is moeilijk te scheiden van de procesvaardigheden. Bij eerdere onderzoeken werden significante correlaties gevonden tussen de score op een procesvaardigheden-toets en scores op andere, inhoudelijke toetsen (o.a. Wood en Ferguson, 1975). In het algemeen wordt de hoogste correlatie, in de orde van .50 gevonden met natuurkunde-toetsen (Comber en Keeves, 1973).

#### Uitvoering van het onderzoek

Een toets van 26 items over de vaardigheden grafieken lezen, tabellen interpreteren en proefopzetten beoordelen werd voorgelegd aan leerlingen van derde klassen in het voortgezet onderwijs, bij havo/vwo (N=277), mavo (n=238) en lto (n=134).

Als voorbeeld zijn in fig.1 een drietal items gegeven. Teneinde de inhoudelijke kennis vast te stellen werden tevens toetsen afgenomen voor biologie, natuurkunde en scheikunde. De afname was gespreid over 57 scholen.

Aan het eind van de toets was een korte vragenlijst opgenomen, waarin werd gevraagd naar sexe, leeftijd, cijfers voor wiskunde en natuurwetenschappen, plezier in deze vakken en het voornemen bij komende pakketkeuze natuurwetenschappen op te nemen.

Fig.1a

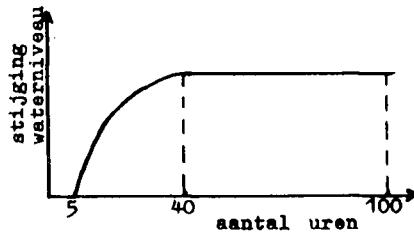
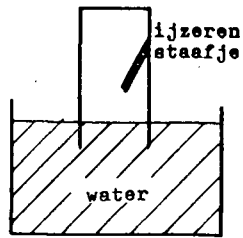


Fig.1b

Groep	Zuurstofgehalte ( $\text{cm}^3 / 100 \text{ cm}^3$ bloed)	
	slagaderen	aderen
gezond	19	15
ziekte I	15	12
ziekte II	9,5	6,5
ziekte III	20	13
ziekte IV	19	18

Fig.1c

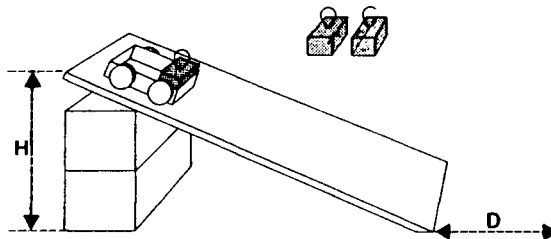


Fig.1a Een fles met lucht, waarin een ijzeren staafje is bevestigd, is omgekeerd in een bak met water geplaatst. (zie figuur). Het ijzeren staafje gaat roesten, waardoor de zuurstof in de fles wordt verbruikt. Elk uur wordt gemeten hoe ver het water in de fles is gestegen. Deze metingen zijn verwerkt in het erbij getekende diagram.

Na hoeveel uur is het roesten gestopt?

- A. na minder dan 5 uur
- B. na 5 uur
- C. na 10 uur
- D. na 100 uur
- E. het roesten stopt niet, volgens het diagram

Fig.1b Bij het verbruik van zuurstof door de mens zijn een aantal factoren belangrijk, zoals het gehalte aan zuurstof in de aderen en in de slagaderen.

Zoals bekend voeren de slagaderen het bloed naar de organen en voeren de aderen vandaar het bloed weer af.

In de volgende tabel worden gemiddelde waarden van het zuurstofgehalte gegeven voor mensen, die aan bepaalde ziekten lijden. Ook voor gezonde mensen zijn de gemiddelde waarden gegeven.

Welke van de volgende vier ziekten betreft een vergiftiging, waardoor in de organen nauwelijks zuurstof wordt verbruikt?

- A. I
- B. II
- C. III
- D. IV

Fig.1c Fred en Pauline meten welke afstand hun 'vrachtauto' aflegt, nadat deze de plank heeft verlaten (afstand=D). Zij kunnen de hoogte van de plank (H, in cm) veranderen. Zij kunnen ook verschillende massa's (gewichten) in de auto doen. Zij willen vooral graag weten, of de massa (het gewicht) invloed heeft op de afstand D.

De eenvoudigste manier om die invloed na te gaan is het meten van D:

- A. bij gelijke hoogte en enkele verschillende massa's
- B. bij gelijke massa en enkele verschillende hoogten
- C. bij gelijke hoogte en gelijke massa, enkele malen herhaald
- D. bij steeds verschillende massa's en ook verschillende hoogten

Bij een survey-onderzoek is het van belang iets te weten over de mate, waarin leerlingen onderwijs hebben gehad in de getoetste leerstof. Daartoe wordt aan de docent een vragenlijst voorgelegd, waarin de 'Gelegenheid Tot Leren' (GTL) wordt aangegeven. Voor elk item wordt gevraagd aan te geven, of de betreffende leerstof in een vorig leerjaar, dit leerjaar, in een later jaar of nooit wordt behandeld. Op klasniveau kan men dan correlaties bepalen tussen GTL en toetsscore (Ten Bruggencate en Pelgrum, 1985). Ook voor de onderhavige toets werd door de docent zo'n vragenlijst ingevuld. Het merendeel van de items werd tevens afgenomen op scholen, die bij natuurkunde met PLON werken (alleen havo/vwo en mavo, N=510). Op de volgende vragen dachten wij met dit onderzoek antwoorden te verkrijgen:

- a. Vertoont de verzameling items een structuur (clusters), zo ja komt deze overeen met de indeling in grafieken, tabellen en proefopzetten?
- b. Hoe verschilt de beheersing van deze vaardigheden per schooltype?
- c. Hoe is de invloed van, c.q. samenhang met resultaten op inhoudelijke toetsen?
- d. Welke samenhang is er met variabelen als sexe, motivatie, wiskundecijfer?
- e. Hoe is het verband met Gelegenheid Tot Leren?
- f. Onderscheiden scholen met het PLON-curriculum (voor natuurkunde) zich van andere scholen?

### Resultaten

Verscheidene analyses werden op de toetsresultaten uitgevoerd. Wij bespreken dit afzonderlijk voor de verschillende vraagstellingen van het onderzoek.

#### a. Clusters van items

Indien er een vaardigheid als 'grafieken lezen' onderscheiden kan worden, dan moet men dat in de itemresponsen terug kunnen vinden: de samenhang van items met grafieken onderling zal dan groter moeten zijn dan de samenhang met andere items. Aangezien alle items in de toets wel enige samenhang hebben, moeten wij hier spreken over een specifieke samenhang binnen een bepaalde itemgroep. Een methode om dat na te gaan is de bepaling van de betrouwbaarheid van de subtoets, gevormd door bijvoorbeeld de 9 vragen over grafieken. Gegeven de betrouwbaarheid  $r$  van de totale toets (26 vragen) geeft de formule van Spearman-Brown de verwachte betrouwbaarheid op de subtoets:

$$r' = n \cdot r / (1 + (n-1) \cdot r) \quad \text{met } n = 9/26$$

indien de items in de subtoets geen specifieke samenhang vertonen. Een hogere  $r'$  dan door de formule voorspeld, zou ons vermoeden bevestigen.

	totaal		havo/vwo		mavo		lto		aantal items
	r	r'	r	r'	r	r'	r	r'	
gehele toets	.78		.74		.73		.73		26
grafieken	.44	.55	.41	.50	.38	.48	.36	.48	9
tabellen	.60	.52	.53	.46	.54	.45	.47	.45	8
proefopzetten	.60	.55	.56	.50	.57	.48	.53	.48	9

Fig.2 Betrouwbaarheid (KR20) van toetsen en subtoetsen, voor totale populatie en per schooltype.

r' = verwachte betrouwbaarheid voor deeltoets na toepassing van de formule van Spearman-Brown.

In figuur 2 zijn de betrouwbaarheden van subtoetsen gegeven, alsmede de voorspelde r' volgens de formule. De conclusies zijn:

- de subtoets grafieken bestaat niet uit specifiek samenhangende items
- de subtoets tabellen vertoont enige specifieke samenhang
- de subtoets proefopzetten vertoont enige specifieke samenhang, maar minder dan tabellen.

Dergelijke conclusies worden ook bereikt, als men items groepeert per discipline: De biologie-items, de natuurkunde-items of de scheikunde-items hangen niet specifiek samen.

Meer verfijnde methoden van clusteranalyse (Mokken-analyse, Rasch-analyse) blijken tot dezelfde conclusie te leiden: De items over grafieken, tabellen of proefopzetten vormen niet of nauwelijks een 'schaal'. Zoekt men, met die technieken, de items bij elkaar die wel een schaal vormen, dan blijkt dat dwars door de genoemde vaardigheden en ook dwars door de drie disciplines heen te lopen (De Bruyn, 1985). De conclusie moet zijn, dat in de procesvaardighedentoets van 26 items statistisch geen substructuur naar vaardigheden of naar disciplines is aan te tonen. Waar in het volgende niettemin groepen items worden samengenomen, gebeurt dat op arbitraire criteria, zoals de percipieerde vakinhoud. Zoals zal blijken, levert dat een enkele keer extra informatie op.

#### b. Resultaten per schooltype

In fig.3 zijn de gemiddelde p-waarden van de items vermeld voor:

- de totale toets procesvaardigheden
- de subtoetsen grafieken, tabellen en proefopzetten
- de subtoetsen, bestaande uit items uit de drie disciplines
- de (apart afgenomen) inhoudelijke toetsen

	havo/vwo	mavo	lto
gehele toets	.68	.54	.48
subtoets grafieken	.53	.44	.39
subtoets tabellen	.71	.54	.47
subtoets proefopzetten	.74	.60	.57
items biologie	.64	.47	.39
items natuurkunde	.69	.56	.51
items scheikunde	.68	.54	.49
toets biologie	.62	.52	.43
toets natuurkunde	.71	.61	.63
toets scheikunde	.59	.43	.30

Fig.3 Gemiddelde p-waarde per (sub) toets en per schooltype

Op alle toetsen scoort havo+vwo het hoogste. Mavo scoort hoger dan LTO, behalve bij de inhoudelijke toets natuurkunde. Afgezien van die laatste toets, is voor de drie schooltypen de volgende 'vuistregel' te geven:

Stellen wij de gemiddelde score op een toets voor havo+vwo op 100%, dan scoort mavo gemiddelde ca. 80% en scoort LTO gemiddelde ca. 70%.

Deze waarden worden ook voor wiskunde-onderwijs in het 2e leerjaar gevonden (Pelgrum c.s., 1983).

#### c. Invloed inhoudelijke kennis op items over procesvaardigheden

Zoals reeds eerder opgemerkt, en zoals ook aan de voorbeeld-items in fig.1 is te zien, gaan de items over procesvaardigheden steeds over een probleem uit een der disciplines. Het is van belang te weten, of de inhoudelijke kennis van het betreffende onderwerp in belangrijke mate de kans op succes bij zo'n item bepaalt.

Op verschillende manieren is geprobeerd het antwoord op deze vraag te vinden:

1. Nagegaan is, of de items over één der disciplines specifieke samenhang vertonen, dus of deze items een schaal vormen (zie onder a.).
2. Van alle items is nagegaan, in hoeverre de respons samenhang vertoont met scores op inhoudelijke toetsen voor de drie disciplines.
3. Bij enkele items over procesvaardigheden werd een referentie-item geconstrueerd en afgenomen, dat niet de betreffende vaardigheid bevat maar overigens wel over hetzelfde onderwerp handelt. Zo werd bij het derde voorbeelditem een extra vraag geconstrueerd over een valbeweging, waarbij niet het probleem over de opzet van de proef aan de orde is.

Het resultaat van alle pogingen was negatief. De items over b.v. scheikundige situaties vormen samen geen schaal. De correlaties van deze items met een scheikunde-toets was niet systematisch hoger dan de correlatie met natuurkunde- of biologie-toetsen. De scores op items en bijbehorende referentie-items vertoonden geen specifieke samenhang. Wij moeten concluderen, dat de score op een item over procesvaardigheden weinig te maken heeft met inhoudelijke kennis, die bij dit item een rol speelt.

Overigens vertoont de totaalscore op de vaardigheidentoets wèl de 'gebruikelijke' samenhang met de toetsscores voor biologie ( $r=.50$ ), natuurkunde ( $r=.62$ ) en scheikunde ( $r=.51$ ). Deze correlaties zijn zelfs hoog te noemen, gezien de betrouwbaarheid van de gecorreleerde scores; zowel voor de vaardigheden toets als voor de inhoudelijke toetsen ligt deze in de orde van .72.

#### d. Samenhang met 'externe' variabelen

Wij wilden nagaan, of de score op de vaardigheden-toets samenhangt met variabelen als sexe, schooltype, wiskunde-cijfer e.d. Daartoe werd een multipelè regressie-analyse uitgevoerd, met de toetsscore als afhankelijk variabele en de diverse gegevens als onafhankelijk variabelen. Op zichzelf zijn de gevonden regressie-coëfficiënten niet goed te interpreteren. Door de regressie ook uit te voeren voor de toetsscores voor de drie disciplines, werd een vergelijking van de invloeden mogelijk: men kan de zogenaamde beta-gewichten dan vergelijken. (Als bijvoorbeeld  $\beta=.25$  dan geeft een verandering van 1 standaarddeviatie in de onafhankelijk variabele een verandering van 0.25 standaarddeviatie in de afhankelijk variabele.)

Als eerste variabelen worden in de regressie-vergelijking steeds sexe en schooltype ingevoerd, als de meest 'onveranderlijke' grootheden. Bij het vergelijken van de beta-gewichten speelt die volgorde overigens geen rol, zolang voor alle afhankelijk variabelen dezelfde volgorde wordt aangehouden (De Bruyn, 1985). Voor de significantie speelt de volgorde wel mee.

De zogenaamde sexe-bias blijkt, waar aanwezig, een hogere score voor jongens op te leveren. Deze is het grootst voor de natuurkunde-toets ( $\beta=.24$ ), voor de procesvaardigheden is deze minder doch wel aantoonbaar ( $\beta=.17$ ). Voor scheikunde en biologie blijkt de sexe-bias gering te zijn.

Voor de samenhang met schooltype worden dezelfde effecten gevonden als onder b. beschreven. Het wiskunde-cijfer blijkt vrij sterk samen te hangen met de score op procesvaardigheden, overigens in dezelfde mate als natuurkunde ( $\beta=.30$ ). Voor biologie en scheikunde is dit globaal de helft. De motivatie voor ('leuk vinden van') natuurkunde hangt, zoals verwacht, met de score op natuurkunde samen, maar ook met de score op procesvaardigheden ( $\beta=.18$ ). Deze invloed is nihil voor biologie en scheikunde.



Het voornemen tot het opnemen van natuurwetenschappen in een eventueel keuzepakket heeft geen invloed op de toetsscores. Wij concluderen, dat bij de vergelijking van externe correlatie de toets over procesvaardigheden veel op een toets over natuurkunde lijkt.

#### e. Gelegenheid Tot Leren (GTL)

Van elke school werd een docent gevraagd de vragenlijst 'Gelegenheid Tot Leren' in te vullen. Voor elk toetsitem werd gevraagd aan te geven, in welk leerjaar de betreffende leerstof wordt behandeld. De antwoorden werden gehercodeerd tot 1 of 0, wel of niet behandeld.

Van de 57 scholen werden slechts 38 bruikbare GTL-lijsten ontvangen. De respons ligt daarbij beduidend lager dan bij vergelijkbare lijsten voor inhoudelijke toetsen. Dit kan erop wijzen, dat docenten moeite hebben met de vraag, of een procesvaardigheid in het onderwijs aan de orde is geweest. Een gevolg van de lage respons is onder andere, dat geen zinvolle analyse voor afzonderlijke schooltypen mogelijk is.

Gemiddeld over alle scholen liggen de GTL-waarden tussen 24% en 96%. Belangrijker is echter de samenhang tussen GTL-waarde en prestatie, waarvoor de p-waarde per klas is genomen. Voor slechts 3 items wordt een significant positieve correlatie gevonden ( $r > .30$ ), doch dit kan ook 'toevallig' worden gevonden als men reeksen correlaties uitrekent. Een extra reden om hier aan toeval te denken is het feit dat ook tweemaal een significant negatieve correlatie wordt gevonden.

Een andere analysemethode gebruikt de correlatie tussen de score op een toets, of subtoets, en de gemiddelde GTL-waarde over die serie items. Ten Bruggencate en Pelgrum (1985) vonden hiervoor waarden van maximaal .40 binnen schooltypen en maximaal .69 over schooltypen heen, voor inhoudelijke toetsen. De scheikunde-items blijken daarbij de hoogste correlatie te vertonen.

Wij vonden lagere correlaties, .22 voor de subtoetsen tabellen en proefopzetten en .25 voor de scheikunde-items in de toets procesvaardigheden. Deze waarden zijn bij het lage aantal waarnemingen niet significant op 10%-niveau. Voor andere subtoetsen werd een nog aanzienlijk lagere correlatie gevonden. Hierbij speelt de lage betrouwbaarheid van deze subschalen, beter gezegd het niet herkenbaar zijn van de schaal als zodanig, mede een rol.

#### f. Resultaten op PLON-scholen

Er werd een deelonderzoek uitgevoerd op scholen, die voor natuurkunde met de PLON-leergang werken. Dit beperkte zich tot havo/vwo en mavo-scholen. Twee vragen uit de toets procesvaardigheden werden door alle leerlingen gemaakt, 15 vragen in geroteerde vorm afgenomen bij een derde deel van de leerlingen. Door deze methode van afname kan de verzameling items niet op clusters worden onderzocht, wel kunnen p-waarden worden vergeleken. Op havo/vwo-scholen scoren PLON-scholen

over de 17 items een gemiddelde p-waarde van 70,5 tegen de niet-PLON scholen 67,4. Dit verschil blijkt op 2%-niveau significant te zijn (Student's  $t=2,82$  bij  $df=16$ ). Op mavo-scholen zijn de gemiddelde p-waarden voor PLON 59,1 en voor niet-PLON 53,8. Ook dit verschil is significant op 2% niveau ( $t = 2,58$  bij  $df = 16$ ).

Bij deze vergelijking valt het op, dat de PLON-scholen hoger scoren dan de niet-PLON-scholen op alle vragen over grafieken. Bij vragen over tabellen en proefopzetten was het in enkele gevallen andersom. Onderzocht werd, of er ook een verschil was voor een inhoudelijke natuurkunde-toets van 20 items. Dit bleek niet het geval: Voor havo/vwo was de gemiddelde p-waarde precies gelijk, bij mavo scoorde PLON iets hoger maar dit was niet significant ( $t = 1,62$  bij  $df = 19$ ).

### Conclusies en aanbevelingen

In de toets procesvaardigheden worden geen interpreteerbare clusters aangetroffen. Enerzijds onderscheiden items over de deelgebieden grafieken, tabellen en proefopzetten zich niet als schaal, anderzijds vertonen de items over een discipline geen specifieke samenhang. Wij kunnen concluderen, dat de deelgebieden veel met elkaar te maken hebben en dat er één vaardigheid aan ten grondslag ligt. Het is echter niet uitgesloten, dat hier ook een algemene factor, als bijvoorbeeld intelligentie, in het spel is.

Verder blijkt het niet uit te maken, of die vaardigheid in een biologische, een natuurkundige of een scheikundige context wordt getoetst. Hieraan moet worden toegevoegd, dat deze conclusie door nieuw onderzoek zou moeten worden bevestigd: In dit verband zouden de PLON-scholen interessant zijn, waar de situatie wellicht voor het onderwerp 'grafieken' anders zou kunnen liggen.

De procesvaardigheden hebben bij gebruik van de huidige leerplannen het meeste te maken met het vak natuurkunde. Zo correleren de itemscores, in vergelijking met biologie en scheikunde, het meest met de score op een natuurkunde-toets (zelfs items die duidelijk over biologie of scheikunde gaan). Ook de samenhang met sexe en met het wiskunde-cijfer is vergelijkbaar. Dit resultaat is in overeenstemming met de bevindingen in eerder onderzoek (Comber en Keeves, 1973).

Op twee aspecten is de toets over procesvaardigheden afwijkend van de inhoudelijke natuurkunde-toets. Op lto-scholen wordt slechter gescoord op procesvaardigheden dan op mavo-scholen, terwijl het voor natuurkunde-toetsen andersom is. Verder scoren PLON-scholen significant hoger op de toets procesvaardigheden dan niet-PLON-scholen, maar voor de natuurkunde-toets is er geen significant verschil.

De afname van een vragenlijst Gelegenheid Tot Leren bij docenten is bij dit onderzoek matig geslukt. De lage respons en de lage correlaties van GTL met p-waarde suggereren, dat docenten bij dit type items moeite hebben met de invulling van de vragenlijst. Het is wellicht

onduidelijk wat de 'leerstof' bij een item is. Om verbetering te krijgen, zal een ander instrument moeten worden ontwikkeld. Hierbij kan worden overwogen de respondent meer in de richting van de procesvaardigheid in een item te leiden, b.v. te vragen of een grafiek van het onderhavige type is behandeld. Verder zou het antwoord van docenten voor alle natuurwetenschappelijke vakken aan een school verkregen moeten worden.

De belangrijkste functie van het verrichte onderzoek is, dat een goed beeld is verkregen van de stand van zaken anno 1984. Dit maakt het in de toekomst mogelijk de meerwaarde te bepalen van nieuwe curricula op het punt van procesvaardigheden. Het aangetoonde effect bij gebruik van het PLON-curriculum kan hierin een voorbeeld-functie vervullen.

#### Noot

1. De uitvoering van het hier beschreven onderzoek werd mogelijk gemaakt door een subsidie van de Stichting voor Onderwijsresearch (SVO)

#### Literatuur

- A.A.A.S. Commission on Science Education, *The psychological basis of Science - A process Approach*, Washington-DC: A.A.A.S., 1965.
- Bruggencate, G.C. ten en Pelgrum, W.J. *Eerste uitkomsten van het internationale deel van de Second IEA Study in Nederland*, Paper voor de Onderwijs Research Dagen, Enschede: I.H.-Twente, 1985.
- Bruyn, I.de. *Toetsing van enkele vaardigheden bij het interpreteren van data uit natuurwetenschappelijke experimenten*, Paper voor de Onderwijsresearchdagen, Enschede: I.H.-Twente, 1985.
- Comber, L.C. en Keeves, J.P. *Science education in nineteen countries. International Studie in Evaluation I*, Stockholm: Almqvist&Wiksel, 1973.
- Inhelder, B. en Piaget, J. *The growth of logical thinking: From childhood to adolescence* (trans. Parsons A & Milgram, S.), New York: Basic Books, Inc., 1958.
- Padilla, M.J., Okey, J.R. en Dillashaw, F. The relationships between science process skill and formal thinking abilities, *Journal of Research in Science Teaching*, 20, 239-246, 1983.
- Padilla, M.J., Okey, J.R. en Garrard, K. The effects of instruction on integrated process skill achievement, *Journal of Research in Science Teaching*, 21, 277-187, 1984.
- Pelgrum, W.J., Eggen, Th.J.H.M. en Plomp, T. *Het Tweede Wiskunde Project: Beschrijving van uitkomsten*, Enschede: I.H.-Twente, afd. Toegepaste Onderwijskunde, 1983.
- Wood, R. en Ferguson, C.H. Teacher assessment of practical skills in A-level chemistry, *School Science Review*, 56, 605-608, 1975.