



**Universiteit  
Leiden**  
The Netherlands

## **Beoordeling van lesgeefgedrag van leerkrachten lichamelijke ontwikkeling**

Kemper, H.C.G.; Verschuur, R.; Ras, J.G.A.; Snel, J.; Splinter, P.G.; Tavecchio, L.W.C.

### **Citation**

Kemper, H. C. G., Verschuur, R., Ras, J. G. A., Snel, J., Splinter, P. G., & Tavecchio, L. W. C. (1976). Beoordeling van lesgeefgedrag van leerkrachten lichamelijke ontwikkeling. *De Lichamelijke Opvoeding*, 64, 201-204. Retrieved from <https://hdl.handle.net/1887/10291>

Version: Not Applicable (or Unknown)

License:

Downloaded from: <https://hdl.handle.net/1887/10291>

**Note:** To cite this publication please use the final published version (if applicable).

# Beoordeling van lesgeefgedrag van leraren lichamelijke oefening

door  
Han C. G. Kemper  
Robbert Verschuur  
Koos G. A. Ras  
Jan Snel  
Paul G. Splinter  
Louis W. C. Tavecchio

Coronel Laboratorium en Laboratorium  
voor Psychofysiologie  
Jan Swammerdam Instituut  
Eerste Constantijn Huygensstraat 20  
Amsterdam

In dit artikel wordt de beoordeling van het lesgeefgedrag van leraren op basis van fysieke oefening besproken. De methode van de beoordeling wordt beschreven en de resultaten worden besproken. De resultaten laten zien dat de beoordeling van het lesgeefgedrag van leraren op basis van fysieke oefening mogelijk is.

## 2. Methodiek

2.1. De keuze van de beoordelingsmethode  
De keuze van de beoordelingsmethode is afhankelijk van de aard van de activiteit die wordt beoordeeld. In dit onderzoek is de keuze van de beoordelingsmethode afhankelijk van de aard van de activiteit die wordt beoordeeld. De keuze van de beoordelingsmethode is afhankelijk van de aard van de activiteit die wordt beoordeeld.

De derde les wordt door een andere leraar gegeven. De derde les wordt door een andere leraar gegeven. De derde les wordt door een andere leraar gegeven.

In dit artikel wordt de beoordeling van het lesgeefgedrag van leraren op basis van fysieke oefening besproken. De methode van de beoordeling wordt beschreven en de resultaten worden besproken. De resultaten laten zien dat de beoordeling van het lesgeefgedrag van leraren op basis van fysieke oefening mogelijk is.

In dit artikel wordt de beoordeling van het lesgeefgedrag van leraren op basis van fysieke oefening besproken. De methode van de beoordeling wordt beschreven en de resultaten worden besproken. De resultaten laten zien dat de beoordeling van het lesgeefgedrag van leraren op basis van fysieke oefening mogelijk is.



In dit artikel wordt de beoordeling van het lesgeefgedrag van leraren op basis van fysieke oefening besproken. De methode van de beoordeling wordt beschreven en de resultaten worden besproken. De resultaten laten zien dat de beoordeling van het lesgeefgedrag van leraren op basis van fysieke oefening mogelijk is.

In dit artikel wordt de beoordeling van het lesgeefgedrag van leraren op basis van fysieke oefening besproken. De methode van de beoordeling wordt beschreven en de resultaten worden besproken. De resultaten laten zien dat de beoordeling van het lesgeefgedrag van leraren op basis van fysieke oefening mogelijk is.

# Beoordeling van lesgeefgedrag van leraren lichamelijke oefening

door

Han C. G. Kemper

Robbert Verschuur

Koos G. A. Ras

Jan Snel

Paul G. Splinter

Louis W. C. Tavecchio

Coronel Laboratorium en Laboratorium voor Psychofysiologie  
Jan Swammerdam Instituut  
Eerste Constantijn Huygensstraat 20  
Amsterdam

## Summary

### Judgement of teaching behaviour of physical education teachers

This study gives an account of the construction and application of a system, developed to assess professional skill of teacher behaviour in physical education (p.e.).

Four male teachers took part in a multidisciplinary study into the effects of two extra lessons of p.e. a week during a schoolyear upon the physical and psychological development of 12 and 13 year old schoolboys.

Based on Ras (1971) 10 categories were

developed and three times during a lesson rated on a 5-point-scale.

To measure differences in teaching behaviour a sample of four series of videotaped lessons were taken and judged by four groups viz. experts (n=5), p.e.-students (n=62), p.e.-teachers (n=51) and the teachers themselves (n=4).

Reproducibility and objectivity of this system was checked. From the results it will be stressed that the reliability of judgement of p.e. teacher behaviour causes many problems.

## 1. Inleiding

In een onderzoek naar de effecten van extra lichamelijke oefening op de lichamelijke en geestelijke ontwikkeling van 12- en 13-jarige jongens uit brugklassen van een school voor VWO en HAVO (Kemper e.a. 1974), is het van groot belang de storende variabele leraar te leren kennen. Door het lesgeefgedrag van de leraren lichamelijke oefening meetbaar te maken kan men trachten meer te weten te komen over de invloed die dit lesgeefgedrag op de resultaten van het onderzoek zou kunnen uitoefenen.

In het experiment werden de lessen lichamelijke oefening aan vier brugklassen door vier verschillende leraren gegeven. In het jaar voorafgaande aan het experiment werden alle te geven lessen op een

kaartsysteem uitgewerkt om in ieder geval uniformiteit te krijgen wat betreft indeling en inhoud van de lessen. Vanuit experimenteel oogpunt is het ideaal wanneer leraren niet verschillen in hun lesgeefgedrag. Wanneer dit echter niet het geval is, moet getracht worden deze eventuele verschillen te meten.

Alle lessen lichamelijke oefening waren opgesteld volgens de richtlijnen van het Rijksleerplan Lichamelijke Opvoeding (1968). Op grond van de daarin vermelde doelstellingen betekent dit dat de lessen, indien deze tot resultaten willen leiden in ieder geval dienovereenkomstig aangeboden moeten worden en dat deze derhalve alleen maar van waarde kunnen zijn, als men in de vakwereld een positief oordeel kan uitspreken over de kwaliteit van de vaktechnische aspecten van de lessen zoals deze in feite werden gegeven.

Vier, op papier identieke lessen, welke door de vier leraren aan hun brugklas werden gegeven werden beoordeeld op een aantal vaktechnische aspecten van het lesgeefgedrag.

## 2. Methoden

### 2.1. De keuze van de te beoordelen lessen

Uit het totale lessenbestand werden van iedere leraar gedurende het schooljaar 1971/1972 per trimester 5 lessen op videoband opgenomen, dus in totaal 15 series van vier lessen. De eerste serie van 3 lessen werd opgenomen om enerzijds de leraar en zijn klas vertrouwd te maken met de situatie en anderzijds om de filmers te laten wennen aan de apparatuur. Door storingen van technische aard waren uiteindelijk van elke leraar 8 lesopnamen bruikbaar voor beoordeling. Door een onafhankelijke persoon werden vervolgens vier series lessen uitgekozen die wat betreft de leerstof verschilden en bovendien gespreid lagen over het betreffende schooljaar:

1. Ringenles (vòòr de herfstvakantie).
2. Rekstokles (vòòr de kerstvakantie).
3. Bok- en kopstandles (vòòr de paasvakantie).
4. Basketballes (vòòr de zomervakantie).

De derde les werd door één groep beoordelaars 2 maal beoordeeld.



## 2.2. Het vastleggen van de lessen op videoband

De opnames geschieden met behulp van een vaste camera (Sony AV-3670-CE) in de toestelruimte, gelegen aan de korte zijde van het gymnastieklokaal, van waaruit met een groothoeklens bijna de gehele zaal bestreken kon worden. Het spreken van de leraar werd telemetrisch geregistreerd via een loopmicrofoon, die onder zijn trainingspak met een koordje om de hals op de borst hing. Het geluid van de leerlingen kon daardoor diffuus worden opgevangen.

## 2.3. Keuze van het beoordelings-systeem

Bij de keuze van het beoordelings-systeem werd uitgegaan van het systeem dat gebruikt wordt bij het afnemen van examens in het lesgeven aan schoolklassen. Dit lesgeef-cijfer bestaat in wezen uit de som van een groot aantal gewogen eisen, die ten grondslag liggen aan de totale lesgeef-

beoordeling. Voor het experiment werd het echter van belang geacht om de afzonderlijke cijfers voor die verschillende eisen te hanteren, opdat een genuanceerder inzicht in de overeenkomsten en verschillen tussen de leraren kon worden verkregen.

Ras (1971) stelde voor om een analyse-schema te gebruiken, waarbij over korte perioden van vijf minuten het lesgeef-gedrag telkens op een groot aantal verschillende aspecten wordt beoordeeld. De complexe lesgeef-situatie kan daarmee meer gestructureerd worden, men komt tot een meer genuanceerd oordeel over het lesgeven. Daarmee wordt vermeden:

- dat een laatste indruk doorslaggevend wordt,
- dat één aspect, dat een 'hobby' vormt van de beoordelaar, een overheersende invloed heeft op de totale beoordeling.

Bij het door ons gehanteerde beoordelings-schema zijn wij uitgegaan van Ras (1971) en wel van dat deel dat betrek-

king heeft op de specifieke eisen van de lessen lichamelijke oefening. Bovengenoemd schema bestaat uit de volgende hoofdcategorieën: 1. leerstofkeuze, 2. organisatie en 3. houding en optreden.

In ons onderzoek noemden wij deze: 1. leerstof, oefenstof, 2. organisatie en 3. docent. Voorts hebben wij een vierde hoofdcategorie toegevoegd 4. leerlingen.

Het definitieve beoordelings-schema is weergegeven in schema 1. Elke hoofdcategorie werd onderverdeeld in een aantal subcategorieën. De in totaal 10 subcategorieën werden per les door de beoordelaars drie maal gescoord. Dit aantal was in overeenstemming met de door de leraren gebruikte lesindeling:

1. inleiding ( $\pm 5$  minuten)
2. eerste deel ( $\pm 15-20$  minuten)
3. tweede deel ( $\pm 15-20$  minuten)

Elke subcategorie werd gewaardeerd door een kruisje te zetten op een 5-puntsschaal.

## Beoordelingsformulier les lichamelijke oefening

Waarderingscijfer:

- 1 = slecht
- 2 = onvoldoende
- 3 = voldoende of niet van toepassing
- 4 = goed
- 5 = uitstekend

(S.V.O.-Projekt 0185)

- datum :
- lesgever :
- lesnummer :

	Inleiding					Eerste deel					Tweede deel				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
<b>1. Leerstof / Oefenstof</b>															
1.1. Aangepast aan vaardigheid															
1.2. Verantwoorde opbouw															
1.3. Korrektes (individueel, klassikaal)															
<b>2. Organisatie</b>															
2.1. Opstellingen (toestel, leerling)															
2.2. Veiligheidsmaatregelen															
<b>3. Leerlingen</b>															
3.1 Arbeidsintensiteit (m.b.t. uithoudingsvermogen)															
3.2 Arbeidsvreugde															
<b>4. Docent</b>															
4.1. Houding en Optreden															
4.2. Spreken															
4.3. Uitleg en Voordoen															

- naam beoordelaar :
- student(jaar) / docent :
- lesgeefervaring op school (exclusief hospiteren) : JA/NEE, zo ja,

- welk schooltype :
- aantal lesuren :
- aantal jaren :

- opmerkingen/suggesties:

Schema 1

#### 2.4. Keuze van de beoordelaars

Ten behoeve van het experiment werden middels een onafhankelijke aanwijzing drie inspecteurs en één inspectrice uitgenodigd aan de beoordeling hun medewerking te verlenen. Tevens verleende een methodiek docent aan een academie voor lichamelijke opvoeding zijn medewerking. Genoemde functionarissen mogen in de vakwereld van de lichamelijke oefening gezien worden als het kader dat als groep kan beoordelen of een les lichamelijke oefening aan de impliciete normen voldoet en of deze les als voldoende of onvoldoende kan worden beoordeeld. Deze groep van 5 beoordelaars zal door ons verder worden betiteld als 'experts'.

In een later stadium is één serie van vier lessen beoordeeld door drie andere groepen: een groep van 62 vierde jaars studenten van de Academie voor Lichamelijke Opvoeding in Amsterdam, een groep van 51 leraren lichamelijke oefening van de A.G.O.V., werkzaam in de regio Amsterdam en tenslotte door de vier leraren die de lessen zelf hadden gegeven. Zij beoordeelden echter niet hun eigen les.

#### 2.5. Beoordelingsprocedure

Bij iedere beoordeling werd steeds één serie van vier dezelfde lessen afgedraaid in een willekeurige volgorde wat betreft de leraren. Elke groep beoordelaars zat in één ruimte tegelijkertijd te beoordelen. Aan de beoordelaars werd vooraf meegegeeld wat de inhoud en indeling van de les was, zoals dit op de leskaart stond aangegeven.

Aan het einde van elk lesonderdeel werd gevraagd om een beoordeling te geven op de tien subcategorieën, zonder dat de mogelijkheid werd gegeven om onderling overleg te plegen. Ook werd tegengegaan dat de beoordelaars door spontane reacties van goed- of afkeuring elkaar zouden beïnvloeden.

Op verzoek van de beoordelaars, werd niet de gehele les afgedraaid maar slechts een gedeelte van ieder onderdeel. De observatietijd was afhankelijk van de aard van het lesonderdeel en was in ieder geval zo lang dat van iedere beoordelaar de mening vaststond.

### 3. Experimentele opzet

#### 3.1. Proefpersonen, proefklassen

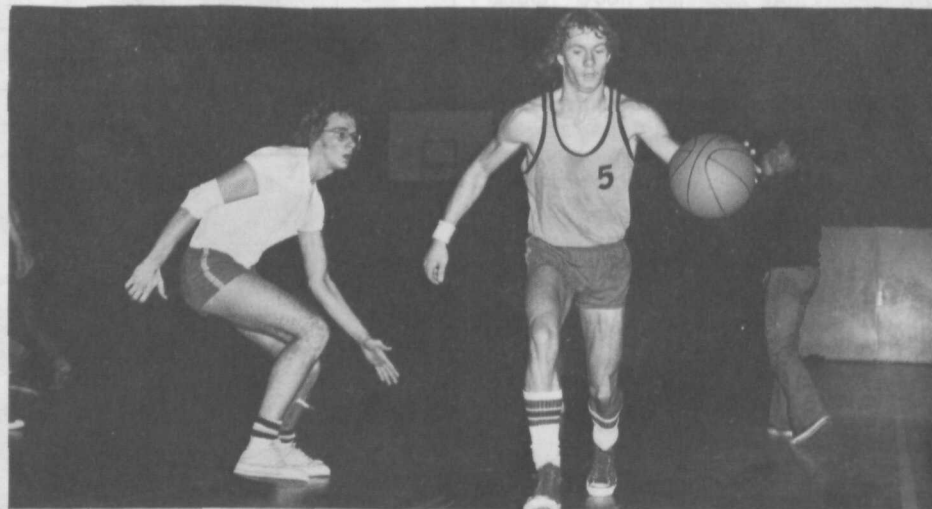
De proefpersonen waren jongens uit de vier brugklassen van het St. Ignatius College, een school voor VWO en HAVO te Amsterdam. De vier klassen bevatten in totaal 82 jongens en worden in het vervolg gecodeerd als klas 1, 2, 3 en 4. (zie tabel 2)

#### 3.2. De leraren lichamelijke oefening

De vier leraren die ten tijde van het onderzoek aan de school waren verbonden kregen ieder één brugklas toegewezen, waaraan zij het gehele schooljaar 1971/1972 lichamelijke oefening doceerden. Alle vier waren zij in het bezit van de akte M.O.-P en hadden vijf of meer jaren lesgeefervaring. Hun leeftijd varieerde van 28 tot 32 jaar ten tijde van het onderzoek. In het vervolg zullen zij gecodeerd worden als resp. A, B, C en D (tabel 2).

	age	class	number of boys
Teacher A	28	1	19
Teacher B	30	2	19
Teacher C	32	3	22
Teacher D	30	4	22

Tabel 2: Verdeling van de vier leraren over de vier brugklassen.



# Beoordeling van lesgeefgedrag

**Tabel 3:** Gemiddelde score (X) en standaard afwijking (s.d.) van de beoordeling door de experts van dezelfde vier lessen, elk gegeven door de 4 leraren.

Teacher Category	A		B		C		D	
	X	(s.d.)	X	(s.d.)	X	(s.d.)	X	(s.d.)
<b>1. Exercises</b>								
1.1. adjustment	3.53	(.09)	3.70	(.41)	3.18	(.39)	3.71	(.42)
1.2. methods	3.13	(.15)	3.26	(.38)	2.84	(.35)	3.45	(.18)
1.3. correction	3.00	(.42)	3.17	(.40)	2.53	(.48)	3.47	(.33)
<b>2. Organisation</b>								
2.1. formation	3.60	(.36)	3.56	(.42)	3.10	(.32)	3.64	(.32)
2.2. safety	3.00	(.05)	3.19	(.35)	2.80	(.24)	3.29	(.46)
<b>3. Pupils</b>								
3.1. work intensity	3.65	(.29)	3.58	(.43)	3.14	(.52)	3.43	(.30)
3.2. joy	3.60	(.19)	3.60	(.28)	3.17	(.33)	3.58	(.22)
<b>4. Teacher</b>								
4.1. attitude	3.51	(.13)	3.70	(.29)	2.58	(.57)	3.77	(.31)
4.2. speech	3.08	(.10)	3.41	(.24)	2.51	(.31)	3.28	(.23)
4.3. explanation/demonstration	3.09	(.18)	3.06	(.48)	2.51	(.38)	3.51	(.37)
<b>Total</b>	3.32		3.42		2.84		3.51	

## 4. Resultaten

### 4.1. Beoordeling door de experts

De scores van de experts is voor iedere leraar per categorie gemiddeld over de vier series door hen beoordeelde lessen. De gegevens zijn samengevat in tabel 3. Uit deze gegevens blijkt dat leraar A, B en D op alle tien categorieën gemiddeld over de vier lessen door de experts als 'ruim voldoende' (> 3) worden beoordeeld. Leraar C wordt op 4 van de 10 categorieën als 'ruim voldoende' beoordeeld. Uit de totale score, gemiddeld over de tien categorieën komt duidelijk naar voren dat de verschillen in beoordeling tussen de leraren A, B en D uiterst gering zijn. De beoordeling van leraar C is echter aanzienlijk lager dan van de overige leraren. Met een gemiddelde score van 2.84 wordt deze leraar op zes van de 10 categorieën lager dan voldoende (< 3) beoordeeld.

*Wordt vervolgd.*

#### Correspondentieadres:

dr. H. C. G. Kemper  
Jan Swammerdam Instituut  
1e C. Huygensstraat 20  
Amsterdam  
tel. 020 - 181313 tst. 32



#### 4.2. Reproduceerbaarheid van de beoordelingen

De reproduceerbaarheid van het beoordelingssysteem geeft de meetbetrouwbaarheid van dit systeem aan als dit herhaald wordt door dezelfde beoordelaars en onder dezelfde omstandigheden (Barrow e.a. 1968).

Om de produceerbaarheid van deze beoordelingen van het lesgeefgedrag vast te stellen is één serie van vier lessen voor een tweede maal ter beoordeling aangeboden aan de experts. Van te voren werd aan hen niet meegedeeld dat het ging om een tweede beoordeling van dezelfde les. Men bemerkte overigens snel dat het ging om een reeds eerder vertoonde les. De les die voor de tweede maal werd beoordeeld was de les waarin bokspringen en kopstand werden beoefend. Het interval tussen de eerste en tweede beoordeling bedroeg zes maanden.

Omdat van een 5-puntsschaal gebruik werd gemaakt konden de verschillen tussen eerste en tweede beoordeling niet groter zijn dan 4 punten, immers  $5 - 1 = +4$  of  $1 - 5 = -4$ .

De verschillen tussen eerste en tweede beoordelingen bleken echter een waarde te hebben van -3, -2, -1, 0, +1 of +2 (zie tabel 4). Volgens deze indeling werden de verschillen geturfd en gesommeerd over de tien categorieën, een - of + betekent dat de tweede beoordeling respectievelijk lager of hoger is dan de eerste beoordeling.

Uit de resultaten blijkt dat het aantal gelijkblijvende scores (verschilscore = 0) per beoordelaar en ook per docent verschillend is. Ten aanzien van de verschillen groter dan 0, valt op dat de meeste negatief zijn, dat wil zeggen, dat tijdens de tweede beoordeling de leraren als minder goed werden beoordeeld.

Van bijzonder belang zijn echter de verschillen scores ter grootte van nul. Hoe groter dit aantal, hoe hoger de mate van betrouwbaarheid. Wanneer wij ten aanzien van het aantal gelijkblijvende beoordelingen een percentage van 70% redelijk achten, kunnen wij, uitgaande van een betrouwbaarheidsniveau van 5%, links-eenzijdig de betrouwbaarheidsgrens berekenen. Het bleek nu dat waarden van 17 en hoger acceptabel genoemd konden worden. Op grond hiervan blijkt dat de beoordelaars 1, 2 en 3 ieder twee van de viermaal deze waarde niet bereiken, beoordelaar 5 slechts eenmaal niet, terwijl beoordelaar 4 geen enkele keer weet te voldoen. Over het totaal wordt dan gevonden dat elf van de twintig keer niet aan de door ons gestelde norm wordt voldaan (zie tabel 4).

Van het totale aantal verschillen scores groter dan nul nemen de negatieve 64% voor hun rekening. Er komen zelfs veranderingen voor van 3 punten. Beoordelaar 1 bijvoorbeeld gaf een waardering van 4 (= goed) op de eerste beoordeling en een 1 (= slecht) op de tweede beoordeling; veranderingen van 2 punten traden 32 en 3 maal op, dat is 12% van het totale aantal verschillen scores groter dan nul.

Ook kan de vraag gesteld worden of de reproduceerbaarheid van de beoordelaars bij bepaalde categorieën beter blijkt te zijn. Daardoor zijn van alle beoordelaars tezamen de beoordelingen ingedeeld per categorie volgens hetzelfde principe als in tabel 4. Als wij ook nu weer uitgaan van een percentage van 70% wordt de linker betrouwbaarheidsgrens gevormd door de waarde 35 bij een onbetrouwbaarheidsniveau van 5%.

Uit tabel 5 blijkt dan dat geen enkele categorie deze waarde bereikt. Het is dus niet mogelijk een of meer categorieën aan te wijzen, waarin de beoordelingen volgens de door ons gestelde normen betrouwbaar genoemd kunnen worden.

Tabel 4

Frekwentie van de verschillen scores gemiddeld over de 10 categorieën van de 5 experts tussen de 1e en 2e beoordeling (- of + d.w.z. lager of hoger op de tweede beoordeling).

Tabel 4		difference-scores					
Judges	Teachers	-3	-2	-1	0	+1	+2
1	A	1	3	4	17	6	
	B		5	11	11	2	
	C			4	18	6	2
	D	1	3	14	11		1
2	A		1	15	14		
	B		2	8	18	2	
	C		1	13	14	2	
	D		1	2	19	18	
3	A		4	22	4		
	B		1	14	12	3	
	C		1	5	24		
	D		1	5	18	6	
4	A		3	17	8	2	
	B		3	9	15	3	
	C		2	16	12		
	D			11	16	3	
5	A		1	7	18	4	
	B			2	22	6	
	C			4	17	9	
	D			6	16	8	
Total		2	32	189	304	70	3

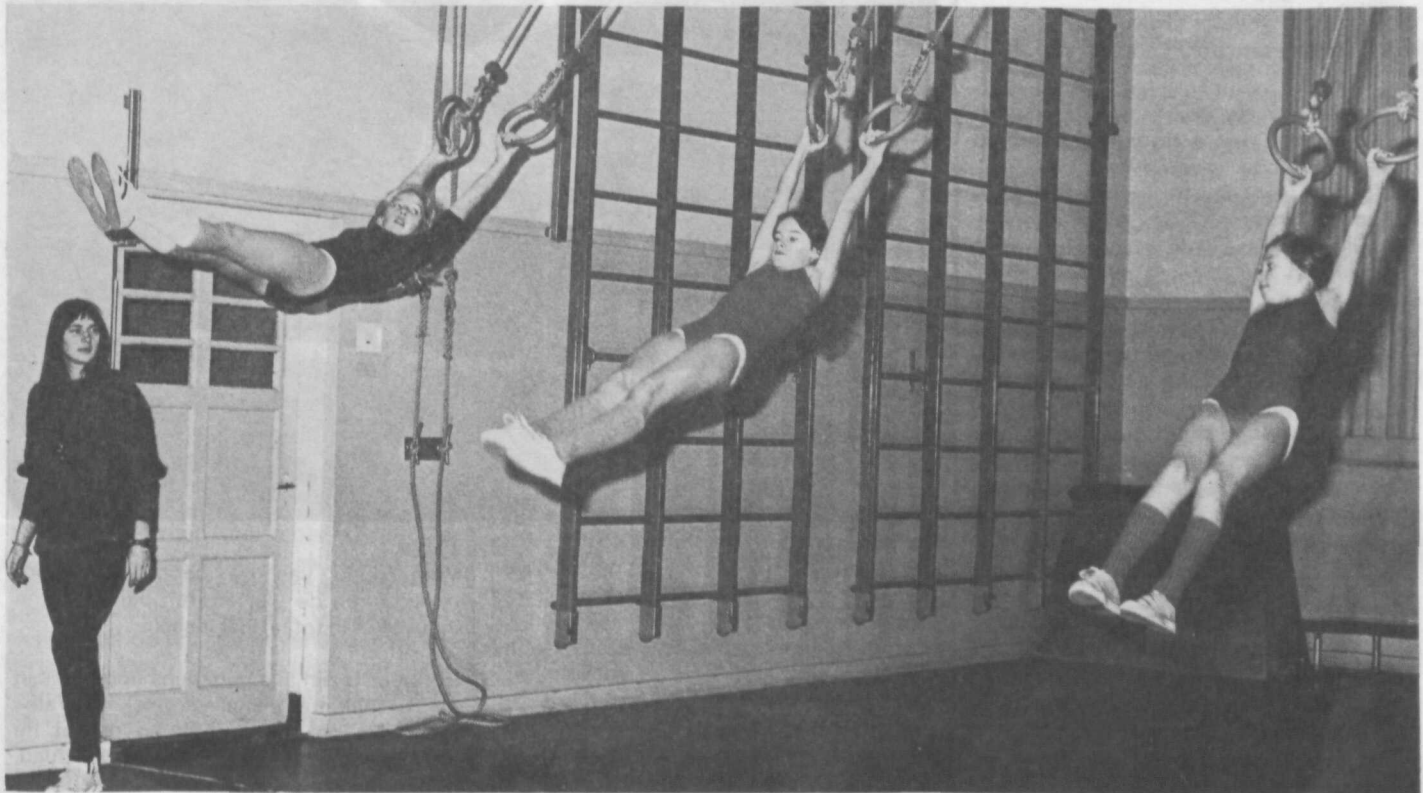
**Tabel 5**  
 Frekwentie van de verschillscores gemiddeld per categorie.

Tabel 5	Difference-scores					
	Categories	-3	-2	-1	0	+1
1.1 adjustment	1	4	13	29	10	3
1.2. methods		8	17	31	4	
1.3. corrections		4	20	29	7	
2.1. formation		7	22	29	2	
2.2. safety	1	3	24	31	1	
3.1. working intensity			23	32	5	
3.2. joy in working		2	22	29	7	
4.1. attitude			16	33	11	
4.2. speech			11	34	15	
4.3. explanation and demonstration		4	20	28	8	
Total	2	32	189	304	70	3

**4.3. Objectiviteit van de beoordelingen**

De objectiviteit van het beoordelings-systeem geeft de meetbetrouwbaarheid aan van het systeem, wanneer dit wordt gehanteerd door verschillende beoordelaars onder dezelfde omstandigheden. De objectiviteit heeft dus betrekking op de invloed van de proefleider op de meetresultaten.

In de jaren volgend op het experiment hebben wij één serie van vier lessen, te weten de derde les (bokspringen, kopstand), op identieke wijze voorgelegd aan 3 andere groepen. De bedoeling was na te gaan in hoeverre de resultaten van de experts overeenkwamen met andere groepen werkzaam op het terrein van de lichamelijke oefening. In tabel 6 zijn de gemiddelde scores van deze serie lessen naast elkaar gezet. Uit de gegevens blijkt dat zowel studenten, leraren als de collega's zelf de drie leraren A, B en D hoger beoordelen dan leraar C. In dit opzicht is er dus geen enkel verschil met de





experts. Wel zijn er verschillen in de absolute beoordelingsnorm: de leraren scoren de 4 leraren iets lager dan de studenten en de collega's, de collega's zelf beoordelen leraar C wat milder (d.w.z. hoger), de studenten en de collega's geven iets hogere beoordelingen en de leraren iets lagere dan de experts.

JUDGEMENT OF PROFESSIONAL SKILL OF TEACHING BEHAVIOUR OF 4 TEACHERS IN PHYSICAL EDUCATION

Tabel 6

TEACHER	A				B				C				D			
	Stu n=62	Tea n=51	Col n=3	Exp n=5	Stu	Tea	Col	Exp	Stu	Tea	Col	Exp	Stu	Tea	Col	Exp
JUDGEMENT OF 1 LESSON BY:																
MEAN	3,6	3,3	3,7	3,4	3,9	3,3	3,6	3,3	2,9	2,8	3,3	3,0	3,6	3,2	3,6	3,6
STAND. DEV.	0,4	0,4	0,1	0,4	0,3	0,4	0,1	0,1	0,3	0,5	0,3	0,6	0,4	0,3	0,2	0,4
MINIMUM	2,3	2,3	3,6	2,8	3,2	2,2	3,5	3,2	2,2	1,6	3,0	2,1	2,9	2,3	3,3	3,2
MAXIMUM	4,4	4,2	3,8	3,8	4,5	4,2	3,7	3,4	3,7	4,2	3,5	3,5	4,6	3,9	3,7	4,0
RANGE	2,1	2,9	0,2	0,9	1,3	2,1	0,2	0,3	1,6	2,5	0,4	1,4	1,7	1,6	0,4	0,8

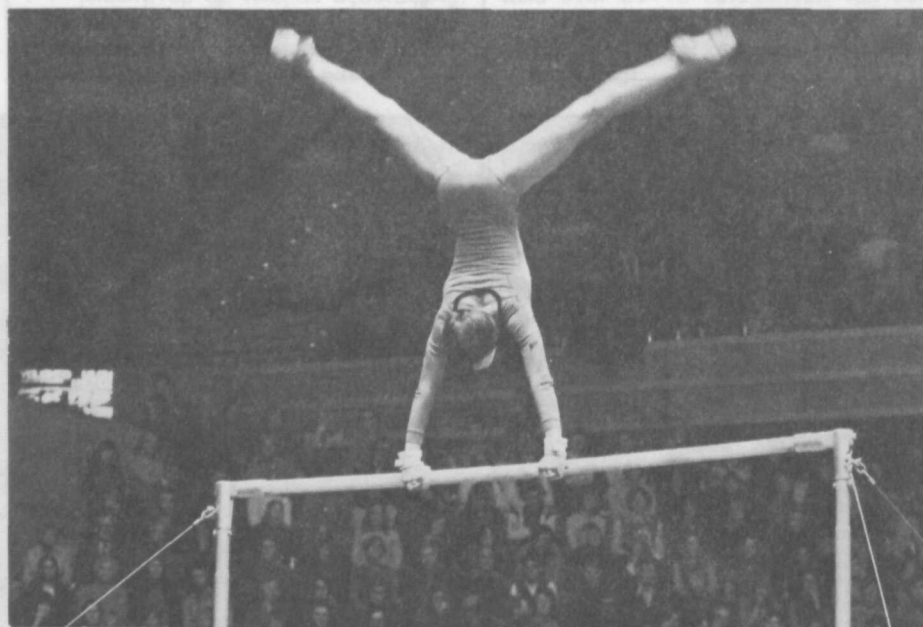
Tabel 6

Gemiddelde beoordeling door studenten, leraren, collega's zelf en experts, van één serie lessen gegeven door de vier leraren.

#### 4.4. Representativiteit van de lessen

Vervolgens is de vraag van belang in hoeverre deze ene serie van vier lessen representatief is geweest voor het lesgeefgedrag van de vier leraren. Daarom hebben wij in tabel 7 de gemiddelde beoordeling van deze serie lessen vergeleken met die van alle door de experts beoordeelde lessen.

De gemiddelde scores over de vier lessen van A, B en D komen bijna geheel overeen met die van die ene les: C wordt met 2,8 gemiddeld lager gescoord. Het beeld is dus in zijn geheel identiek en wij mogen aannemen dat de gekozen les een redelijke afspiegeling van het totale lesgeefgedrag vormt.



Tabel 7

TEACHER	EXPERTS JUDGEMENT OF 1 LESSON				EXPERTS JUDGEMENT OF 4 LESSONS			
	A	B	C	D	A	B	C	D
MEAN	3,4	3,3	3,0	3,5	3,3	3,4	2,8	3,5
STAND. DEV.	0,4	0,1	0,6	0,4	0,1	0,3	0,3	0,3
MINIMUM	2,8	3,2	2,1	3,2	3,2	3,2	2,4	3,2
MAXIMUM	3,8	3,4	3,5	4,0	3,5	3,9	3,2	3,8
RANGE	0,9	0,3	1,4	0,8	0,3	0,7	0,8	0,6

Tabel 7

Vergelijking van de beoordeling door experts van de vier leraren aan de hand van één of in totaal vier lessen.

In figuur 8 zijn alle gegevens nogmaals in een grafiek met elkaar vergeleken. Welke groep nu ook beoordeelt en of dat nu over vier series of over 1 serie is geschied, in alle gevallen worden de docenten A, B

en D tussen voldoende en goed beoordeeld en leraar C tussen onvoldoende en voldoende. Er is dus sprake van een vrij grote overeenstemming. Tevens blijkt duidelijk dat men het onderling het minste eens is over de lesgever C, hetwelk tot uiting komt in de grootste standaardafwijking.

tijdens de eerste beoordeling van hetzelfde lesdeel. Zo kwam het voor dat één van de experts de categorie veiligheidsmaatregelen de eerste keer voor een bepaald deel van de les als 5 d.w.z. uitstekend, beoordeelde en de tweede keer als 2, d.w.z. onvoldoende. Dit terwijl

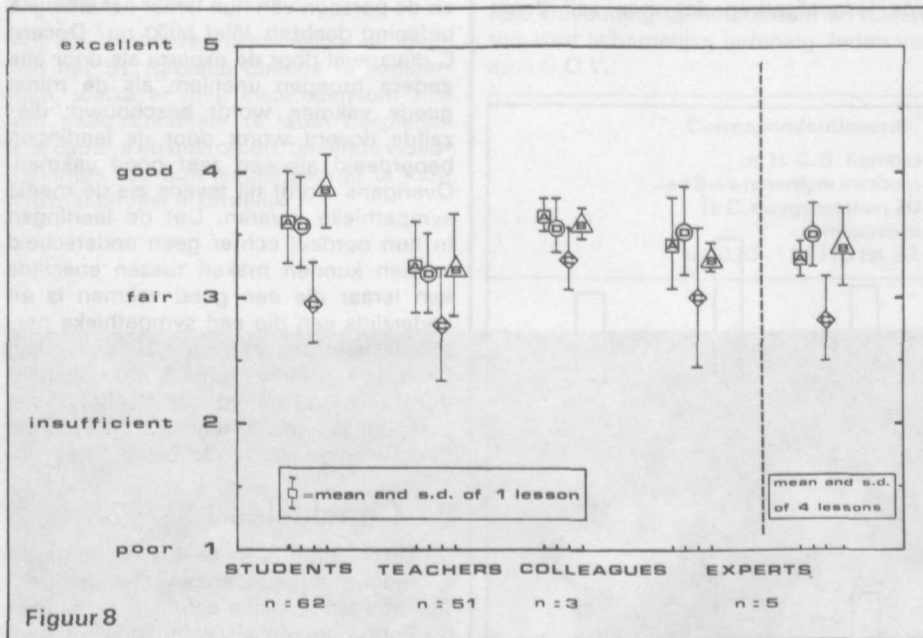
neer deze in het begin van een lesonderdeel een bijzonder goede of slechte indruk maakt. Deze vooronderstelling wordt bevestigd door het feit dat de beoordelaars het dikwijls niet nodig vonden eerst een lesonderdeel in zijn geheel uit te zien, alvorens hun scores vast te stellen. In dit opzicht zou men dus kunnen zeggen dat de opzet om de totale beoordeling van het totale lesgeefgebeuren beter te wegen is mislukt.

Het verdient daarom aanbeveling er naar te streven dat de vaktechnische aspecten van het lesgeefgedrag kunnen worden gescoord met een 'low inference' systeem, dat wil zeggen een systeem waarvan de categorieën die gescoord moeten worden minder voor verschillende interpretaties vatbaar zijn, zodat de reproduceerbaarheid groter is.

Een dergelijk systeem is door ons wel gebruikt om het sociaal-emotionele klimaat tijdens de les te kunnen vaststellen. In dit systeem wordt het gedrag uitgedrukt in de tijdsduur dat een bepaald gedrag (instructie, aanmoediging, afkeuring, voordoen) wordt vertoond. Hoewel veel nauwkeuriger, moet men er voor waken dat de nauw omschreven gedragscategorieën zinvolle en voor het lesgeefgedrag relevante aspecten betreffen waarover later een uitspraak gedaan kan worden.

Met het gebruikte systeem, dat een modificatie is van dat van Flanders (1970) is het onder meer mogelijk om directief (autoritair, leraar geïntendeerd) en non-directief (democratisch, leerling geïntendeerd) lesgeefgedrag te onderscheiden. Verder is het interessant na te gaan of er een relatie bestond tussen de vakkundigheid en de mate van sociaal emotioneel gedrag, zoals door beide beoordelingssystemen wordt gemeten.

De 51 leraren welke wij de 5 lessen lieten beoordelen hebben wij ook gevraagd om de 4 leraren op een 5-puntsschaal te scoren, waarbij de uitersten 1 en 5 respectievelijk als zeer democratisch lesgeven en zeer autoritair lesgeven werden gedefinieerd. Het resultaat is in figuur 9 weergegeven. De 'betere leraren' op het vaktechnische systeem werden als meer autoritaire lesgevers beschouwd, maar leraar C bleek ook hier zich te onderscheiden van de overige drie, doordat hij zowel autoritair als democratisch werd gescoord: de frekwentieverdeling is over de hele 5-puntsschaal verspreid.



**Figuur 8**

Gemiddelde en standaardafwijking van de beoordeling van lesgeefgedrag van de vier leraren door de onderscheiden groepen beoordelaars.

## 5. Discussie

De reden om de lessen lichamelijke oefening op een aantal vaktechnische aspecten te beoordelen was een indruk te krijgen over de kwaliteit van deze lessen. Formulering van de 10 subcategorieën van het systeem was met opzet zo gekozen dat deze aansloot bij het taalgebruik in de vakwereld. In het gebruik bleek dit ook na een korte instructie bij geen enkele groep veel moeilijkheden te geven. Desalniettemin bleek de scoring van dezelfde les door dezelfde experts tot grote verschillen te leiden. Dit kan men afleiden uit de resultaten verkregen bij de bepaling van de reproduceerbaarheid (tabel 4). Daarbij bleek dat de beoordelaars tijdens de tweede keer lager scoorden, d.w.z. strenger waren dan

toch op de videoband exact dezelfde lesgeefsituatie werd vertoond. De verbazing van de onderzoekers werd nog groter, toen de experts desgevraagd verklaarden ten opzichte van de eerste keer wat milder te hebben gescoord! Dat in een dergelijke gestandaardiseerde lesgeefsituatie met een zeer gestructureerd beoordelingssysteem de reproduceerbaarheid van lesgeefgedrag zo veel te wensen overlaat mag een ieder, die bij de beoordeling van het lesgeven betrokken is, te denken geven.

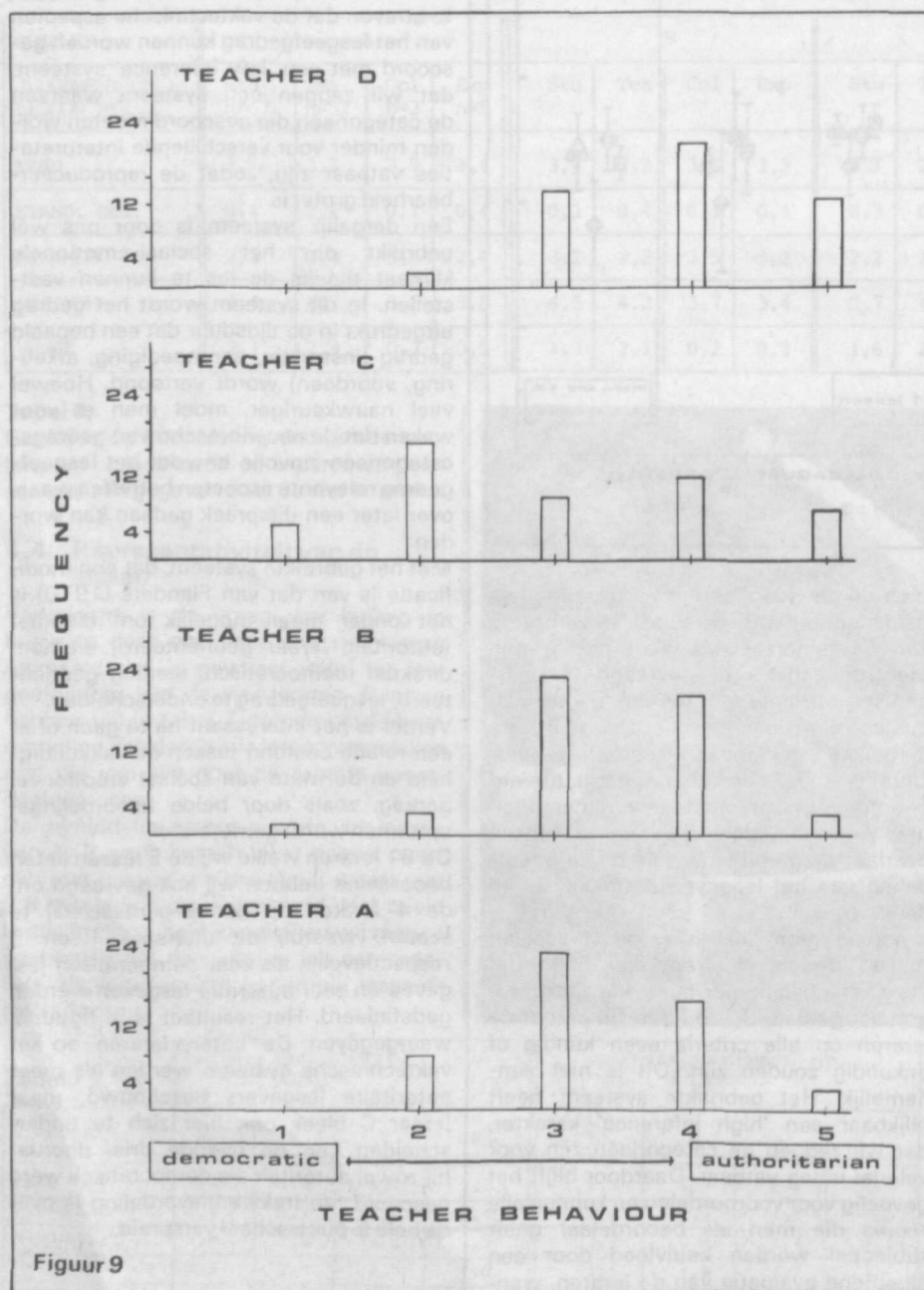
Wanneer men kijkt naar de verschillen tussen de vier leraren, dan blijkt dat deze verschillen voor bijna alle tien categorieën gelden. Dit zou betekenen dat de leraren op alle criteria even kundig of onkundig zouden zijn. Dit is niet aanneemelijk. Het gebruikte systeem heeft blijkbaar een 'high inference' karakter, dat wil zeggen de categorieën zijn voor velerlei uitleg vatbaar. Daardoor blijft het gevoelig voor vooroordelen en kunnen alle scores die men als beoordelaar geeft subjectief worden beïnvloed door een algemene evaluatie van de leraren, wan-

Overigens vroegen wij ook de beoordelende leraren zich zelf op de schaal autoritair (5) - democratisch (1) te scoren. De helft (50%) scoort zich dan in de middenpositie (3), van de 28% die zichzelf als meer autoritair beoordeelt kiest overigens niemand de extreme positie (5), terwijl van de 22% democratische leraren slechts 4% zichzelf extreem democratisch vindt (1).

Nadere analyse leerde dat de leraren, die zichzelf als zeer democratische lesgevers (score 1 en 2) beschouwden, leraar C als autoritair bestempelden en degenen die zich als zeer autoritair beschouwden hem als democratisch scoorden. Het lesgeefgedrag van leraar C wordt gescoord zover mogelijk van het type lesgeefgedrag dat men zelf denkt

te vertonen. Men refereert derhalve bij zijn beoordeling duidelijk aan zichzelf.

De verwarring over de beoordeling van het lesgeefgedrag van leraar C wordt groter wanneer wij de beoordeling van de leerlingen erbij betrekken. Wij vroegen alle leerlingen aan het einde van het schooljaar hoe zij over de deskundigheid en de persoon van hun leraar lichamelijke oefening dachten. Wat blijkt nu? Docent C die zowel door de experts als door alle andere groepen unaniem als de minst goede vakman wordt beschouwd, diezelfde docent wordt door de leerlingen beoordeeld als een zeer goed vakman. Overigens wordt hij tevens als de meest sympathieke ervaren. Dat de leerlingen in hun oordeel echter geen onderscheid zouden kunnen maken tussen enerzijds een leraar die een goed vakman is en anderzijds een die een sympathieke persoonlijkheid is, gaat niet op. Uit de leerlingbeoordeling bleek namelijk dat docent B door zijn klas gemiddeld als het minst sympathiek, maar tegelijkertijd als de beste vakman werd beschouwd.



### 6. Conclusies

1. De kwaliteit van de lessen, gegeven tijdens het onderzoek 'Invloed van extra lichamelijke oefening' werd over het algemeen als voldoende tot ruim voldoende gekwalificeerd door de diverse groepen beoordelaars.
2. De reproduceerbaarheid van de gehanteerde vaktechnische beoordeling, nagegaan bij de experts, bleek onvoldoende te zijn.
3. Hoewel de objectiviteit van het beoordelingssysteem door een gebrek aan reproduceerbaarheid niet meer goed bepaald kan worden, kan wel gesteld worden dat de diverse groepen beoordelaars gemiddeld de 'voldoende' leraar van de 'onvoldoende' leraar in dezelfde mate blijken te onderscheiden.

**Figuur 9**  
 Frekwentieverdeling van de beoordeling op een 5-puntsschaal democratisch-autoritair van de 4 leraren (A, B, C en D) door de groep leraren lichamelijke oefening (n=51).

4. Het feit dat in zulke gestandaardiseerde lesgeefsituaties de individuele beoordelingen zo ver uiteen lopen geeft aan hoe moeilijk het is om op een betrouwbare wijze lesgeefgedrag te beoordelen.

### 7. Slot

Wij zijn de leraren lichamelijke oefening van het St. Ignatius College te Amsterdam, welke er mee instemden om zich een heel schooljaar te laten bespieden met camera en microfoon, teneinde video-opnames van hun lesgeefgedrag te kunnen maken zeer erkentelijk.

Verder spreken wij onze dank uit aan de groep experts welke bereid bleken om gedurende een aantal middagen vele uren videoband te bekijken.

Tot slot gaat onze erkentelijkheid uit naar de andere groepen beoordelaars, te weten de groep vierde jaars studenten (cursus '72/'73) van de Academie voor Lichamelijke Opvoeding te Amsterdam en docenten voor lichamelijke oefening, leden van de A.G.O.V.

#### Correspondentieadres:

dr. H. C. G. Kemper  
Jan Swammerdam Instituut  
1e C. Huygensstraat 20  
Amsterdam  
tel. 020 - 181313 tst. 32

### 8. Literatuur

Barrow, H. M.; R. McGee (1968) A practical approach to measurement in physical education. Lea and Febiger, Philadelphia.

Flanders, N.A. (1970) Analysing teaching behavior. Addison Wesley, Massachusetts.

Kemper, H. C. G.; J. G. A. Ras; J. Snel; P. G. Splinter; L. W. C. Tavecchio, R. Verschuur. (1974) Invloed van extra lichamelijke oefening. Uitg. De Vrieseborch, Haarlem.

Ras, H. (1971) Het gebruik van een analyse model bij de observatie en evaluatie van lesgeefgedrag. De Lich. Opv. 23, 962-967.

