

BETROUWBAARHEID EN VALIDITEIT VAN CONSERVATIE-CRITERIA

[RELIABILITY AND VALIDITY OF CONSERVATION CRITERIA]

Johannes KINGMA & Willem KOOPS
Technische Hogeschool Twente Vrije Universiteit Amsterdam

Three types of scoring criteria for conservation tasks were distinguished: The judgment-only criterion and two judgment-plus-explanation criteria, respectively, the Goldschmid and Bentler criterion and the Piagetian criterion. Subjects were 462 children from kindergarten and primary school grades 1 to 6, who were given 13 conservation tasks. It was shown that the judgment-only criterion was less stringent than the Goldschmid and Bentler criterion, whereas the Goldschmid and Bentler criterion was less stringent than the Piagetian one. The conservation tasks scored according the Piagetian criterion did not show a clearcut factor structure, as was expected from the theory. In contrast, both the judgment-only criterion and the Goldschmid and Bentler criterion showed to produce a reliable test for conservation and a clearcut factor structure. Furthermore these two factor-structures showed to have a high degree of similarity.

In de standaard Piagetiaanse conservatie-taak wordt één van twee objecten, die equivalent zijn m.b.t. een kwantitatief kenmerk (lengte, gewicht, volume, enz.), perceptueel veranderd; de onderzoeker bepaalt of het kind begrijpt dat de kwantitatieve equivalentie in stand blijft («geconserveerd» blijft). De aanwezigheid van de cognitieve structuur, waarop conservatie geacht wordt te berusten, wordt bepaald met behulp van beoordeling («judgment») en/of uitleg («explanation»). Om een beoordeling aan het kind te ontlokken wordt een dichotome vraag gesteld, waarbij het aantal mogelijke antwoorden tot twee wordt teruggebracht: of de objecten zijn gelijk, of ze zijn verschillend. Een uitleg wordt verkregen door het kind te vragen: «Hoe weet je dat?», «Waarom?», of iets dergelijks.

De sinds de zestiger jaren exponentieel groeiende conservatieliteratuur (Brainerd, 1978) wordt helaas ontsierd door een methodologische controverse m.b.t. de vraag of beoordeling dan wel uitleg vereist is om te mogen besluiten tot de aanwezigheid van de betreffende cognitieve structuur (Brainerd, 1973a,b, Kingma, 1980, 1981). Er heerst helaas geen consensus m.b.t. de response-criteria voor conservatie.

Drie hoofd-typen criteria zijn in omloop: het beoordeling + uitleg criterium van Piaget c.s., dat uit de bekende «Conservation Assessment

Kit» (Goldschmid & Bentler, 1968) en het zgn. beoordeling-alleen («judgment-only») criterium. Het eerste criterium wordt zowel binnen als buiten (Elkind, 1961a,b) de Geneefse school gebruikt, de beide andere worden exclusief gehanteerd in Amerikaanse conservatie-experimenten.

Door Piaget c.s. wordt het beoordeling+ uitleg criterium zeer strikt gehanteerd; aan de uitleg zijn scherpe restricties opgelegd. Slechts drie soorten uitleg zijn volgens Piaget en Voyat (1968) theoretisch acceptabel: 1. *additie-subtractie argumenten* (bijv. «Er is niets afgenomen of bijgedaan en daarom zit in dit glas evenveel water als in dat andere»); 2. *inversie-argumenten* (zoals: «Je kunt het water weer *terug* schenken», e.d.); 3. *compensatie-argumenten* (zoals: «Het glas is lager en breder dan het andere en daarom zit er evenveel water in»). Deze argumenten weerspiegelen een operationele en reversibele denkwijze (Piaget, 1968; Piaget & Voyat, 1968).

Bruner (1966) vond dat 60% van de kleuterschoolkinderen een *identiteits-argument* («Het is hetzelfde water, het is alleen maar overgeschonken», e.d.) gebruikten om hun correcte oordelen uit te leggen. Bruner meende dat het identiteitsconcept de kern van conservatie uitmaakt en voegde derhalve het identiteitsargument toe aan het Piagetiaanse drietal. Dit tegen de zin van Piaget c.s. (Piaget, 1967; Piaget & Voyat, 1968), die menen dat het identiteits-argument niet getuigt van reversibiliteit. Piaget noemt Bruner's identiteit «kwalitatieve identiteit» en stelt die tegenover «operationele identiteit», die terug te vinden zou zijn in de additie-subtractie-argumenten. In Goldschmid en Bentler's «conservation kit» is het Bruneriaanse identiteitsargument overgenomen. De kit vereist evenals Piaget beoordeling+ uitleg. Met betrekking tot de uitleg onderscheiden Goldschmid en Bentler drie categorieën: identiteit, inversie en compensatie. De laatste twee zijn identiek aan die van Piaget c.s., de identiteitscategorie echter omvat zowel Bruners kwalitatieve argumenten als Piaget's operationele (additie-subtractie-) argumenten.

Over het derde criterium kunnen we nu kort zijn. Volgens het beoordeling-alleen criterium (Brainerd, 1973a,b) kan worden volstaan met een beoordeling en wordt een uitleg niet noodzakelijk geacht.

De criteria van Goldschmid en Bentler en van Brainerd zijn niet gebaseerd op de Piagetiaanse theorie. Beilin (1975) wees er dan ook op dat de verschillende criteria niet alleen verschillende conservatiemeetwijzen impliceren, maar ook verschillende conceptuele definities

van conservatie. In het Piagetiaanse criterium komen twee zaken samen: conservatie en reversibiliteit. Met behulp van de beide andere criteria wordt alleen de aanwezigheid van conservatie bepaald.

Auteurs, die vóór Brainerd het beoordeling-alleen criterium hanterden (Bever et al., 1968; Braine, 1959; Calhoun, 1971; Mehler & Bever, 1967), gaven daarvoor geen theoretisch relevante argumenten: hun argumenten betroffen stimulus-aspecten en *niet* cognitieve structuren. Een voor Piagetianen relevant en lastig te bestrijden argument gaf Brainerd (1973a,b): in de theorie van Piaget zijn cognitieve structuren van logische en niet van linguïstische aard. Het criterium van een verbale uitleg is derhalve restrictief in oneigenlijke zin.

Beschikbare onderzoeksuitkomsten tonen dat men met behulp van het Piagetiaanse criterium relatief minder conserveerders vindt dan met behulp van het Goldschmid & Bentler criterium: Larsen en Flavell (1970) vinden zeer weinig kleuters en eerste klas basisschoolkinderen met compensatie-argumenten voor hun correcte conservatie-oordelen. Cathart (1971) en Tobin (1972) stelden vast dat zeer veel kinderen van 6 tot 16 jaar (tot 60%) kwalitatieve identiteitsargumenten geven. Voorts valt uit veel onderzoek (Brainerd, 1972; Brainerd & Brainerd, 1972; Gruen, 1966; Gruen & Vore, 1972; Rosenthal & Zimmerman, 1972, 1978; Vingh Bang, 1965) te concluderen dat met behulp van het beoordeling-alleen criterium veel meer conserveerders worden gevonden dan met behulp van enig beoordeling + uitleg criterium. Bovendien toonde Brainerd (1972) aan dat de uit de Piagetiaanse theorie te voorspellen ontogenetische opeenvolging van transitiviteit en conservatie gevonden wordt met behulp van het beoordeling-alleen criterium, maar *niet* met behulp van het beoordeling + uitleg criterium. Reese en Schack (1974) beweerden tenslotte dat met het beoordeling-alleen criterium een toename van «type I fouten» optreedt, terwijl Brainerd (1973a,b) stelt dat de toename van «type II fouten» te verwachten valt met het beoordeling + uitleg criterium.

Tegen de achtergrond van de geschetste theoretische discussie en de beschikbare empirische gegevens leek het ons van belang een bijdrage te leveren aan het onderzoek naar de functionele equivalentie van de drie typen conservatie-criteria. Tegelijk besteedden we daarbij aandacht aan de psychometrische kwaliteiten van de conservatietaken, voor elk van de drie soorten criteria.

METHODE

Proefpersonen

Proefpersonen ($N = 462$) waren Nederlandse kinderen uit twee kleuterscholen en uit twee basisscholen (klassen 1 t/m 6). De kinderen zijn ingedeeld in negen leeftijdsgroepen. Tabel 1 geeft een overzicht.

Tab. 1. — Een overzicht van de gemiddelde leeftijd in maanden en van het aantal proefpersonen per leeftijdsgroep (klas)

leeftijd	aantal ppn.	gem. leeftijd	sd.
1. 4 jr.	35	54	3.2
2. 5 jr.	37	62	2.0
3. tussen 5 en 6 jr.	41	68	2.7
4. 6 jr.	48	78	6.6
5. 8 jr.	58	91	5.3
6. tussen 8 en 9 jr.	53	103	5.2
7. 9 jr.	61	114	5.9
8. 10 jr.	66	126	5.2
9. 11 jr.	63	139	6.1

Materiaal

Op basis van publikaties van Piaget c.s. werd een groep van 13 taken ontworpen, die representatief geacht mogen worden voor de grote reeks taken, die in de ontwikkelingspsychologische conservatie-literatuur voorkomen. In eerder onderzoek werd aangetoond dat deze 13 taken een hoog betrouwbare conservatietest vormen (Kingma, 1983a,b; Kingma & Koops, 1983).

Voordat de dertien conservatietaken werden afgenomen, werden de kinderen in een introductie-trial vertrouwd gemaakt met de vragen. Twee identieke parallelle rijen met zes blokken werden gepresenteerd. Aan het kind werd gevraagd of deze twee rijen hetzelfde aantal blokken bevatten. Daarna werden aan de ene rij twee blokken toegevoegd en werd opnieuw gevraagd of de beide rijen evenveel blokken bevatten. Na deze introductie werden de volgende 13 conservatietaken afgenomen: A. oppervlakte-conservatie (zie Piaget, Inhelder & Szeminska, 1948, hfdst. 11); B. conservatie van aantal (Piaget & Szeminska, 1941, hfdst. 2); C. conservatie van afstand (Piaget et al., 1948, hfst. 3); D. conservatie van substantie (Piaget & Inhelder, 1941, hfdst. 1); E. conservatie van omtrek (Vingh Bang, 1965); F. hoeveelheidsconservatie (Piaget &

Szeminska, hfdst. 1); G. conservatie van discontinue hoeveelheid (Piaget & Inhelder, 1941, hfdst. 2); H. conservatie van gewicht (Piaget & Inhelder, 1941); I. conservatie van bedekt oppervlak (Piaget et al., 1948, hfdst. 11 en 12); J. lengte-conservatie (Piaget et al., 1948); K. conservatie van discontinue lengte (Piaget et al., 1948, hfdst. 4 en 5); L. conservatie van volume (Piaget et al., 1948; hfdst. 13 en 14); en M. conservatie van volume (Piaget & Inhelder, 1941, hfdst. 2). Om een eventuele response set te vermijden, ten gevolge van het gebruik van twee identieke objecten in alle dertien conservatietaken, werden tevens twee conservatietaken voorgelegd (tussen E en F en tussen I en J), waarin de twee objecten ongelijk waren. Deze twee taken worden in de analyse van de resultaten niet betrokken. Een gedetailleerde beschrijving en verantwoording van de 13 conservatietaken is te vinden in Kingma (1977, 1981).

Scoring

Er wordt op drie manieren gescoord: volgens het Piagetiaanse criterium, volgens Goldschmid & Bentler en volgens het beoordeling-alleen criterium.

Volgens het Piagetiaanse criterium werd voor een taak een score 1 toegekend als het kind (na transformatie van één van beide objecten) zowel een correct oordeel gaf als een uitleg die past in één van de volgende drie categorieënargumenten: additie-subtractie; reversibiliteit en compensatie. Een score 0 werd gegeven als het kind of een onjuist oordeel gaf, of een onjuiste uitleg, of beide.

De scoring volgens Goldschmid en Bentler is identiek aan die volgens het Piagetiaanse criterium, met één afwijking: de additie-subtractie categorie is hier vervangen door de «identiteits-argumenten», die naast additie-subtractie ook Bruneriaanse kwalitatieve identiteits-argumenten omvatten.

Volgens het beoordeling-alleen criterium kreeg een kind voor een taak een score 1 als het een correct oordeel gaf, en een score 0 als het een incorrect oordeel gaf.

Samengevat: elk kind kreeg drie scores voor elke taak. De maximale totaalscore voor elk van de drie scoring-systemen was 13.

Procedure

De conservatietaken werden individueel afgenomen. Er werden 11 ervaren onderzoekers ingezet. Uit elke leeftijdsgroep werd ongeveer hetzelfde aantal kinderen random toegewezen aan elke onderzoeker.

De duur van de afname varieerde, afhankelijk van de motivatie en concentratie van het kind. Als het kind er blijk van gaf niet te willen doorgaan, dan werden de resterende taken afgemaakt in de volgende afnamezitting.

Tijdens de afname werd elk kind geobserveerd door twee onafhankelijke onderzoekers. De onderzoeker, die de afname verzorgde, testte het kind en noteerde de uitleg die het kind gaf. De tweede onderzoeker registreerde de responses van het kind eveneens; aldus kan achteraf een «interrater reliability» worden bepaald.

Alle onderzoekers werden vooraf een week lang getraind in het afnemen van de taken en in het coderen van de uitleg van de kinderen. Voordat de feitelijke afname begon, werden alle observatoren getraind tot een betrouwbaarheids criterium van .95 bereikt was.

RESULTATEN

Moeilijkheidsgraad

Inspectie van de gegevens toonde dat de frequentieverdelingen van de totaalscores op de dertien conservatie-opgaven voor elk van de drie scoringscriteria in de kleuterschoolklassen en in de lagere klassen van de basisschool U-vormige verdelingen opleverden. In de hogere klassen van de basisschool vertonen de frequentieverdelingen meer een mono-

Tab. 2. — Een overzicht van de p-waarden voor de 13 conservatie-taken volgens het Goldschmid en Bentler criterium (N = 462)

conservatietaak	leeftijdsgroep								
	1	2	3	4	5	6	7	8	9
A oppervlakte	.07	.03	.03	.30	.49	.75	.88	.91	.98
B aantal	.06	.14	.36	.80	.88	1.00	1.00	1.00	1.00
C afstand	.00	.09	.14	.15	.42	.65	.63	.86	.95
D substantie	.03	.12	.18	.59	.77	.98	.95	1.00	1.00
E omtrek	.00	.09	.05	.35	.56	.79	.80	.95	.95
F hoeveelheid	.00	.09	.15	.57	.73	.87	.93	1.00	1.00
G discontinue hoeveelheid	.00	.09	.15	.52	.73	.87	.93	.98	1.00
H gewicht	.03	.18	.15	.46	.63	.77	.89	.95	.98
I bedekt oppervlak	.03	.06	.05	.24	.51	.73	.86	.92	.95
J lengte	.00	.06	.12	.48	.56	.90	.86	.89	.95
K discontinue lengte	.00	.06	.15	.41	.53	.83	.86	.89	.86
L volume (geometrisch)	.00	.09	.15	.41	.56	.81	.86	.95	.95
M volume (fysisch)	.00	.24	.21	.48	.54	.77	.85	.94	.92

toon stijgende verdelingsvorm. Naar aanleiding van deze verdelingen wordt bij toetsing gebruik gemaakt van non-parametrische toetsen.

Vervolgens werd voor elk paar proefleiders een interrater-betrouwbaarheid voor de observaties berekend. De product-moment correlaties voor de zes paren proefleiders varieerden van .94 tot .99 voor de kwalitatieve identiteitsargumenten. De overeenstemming voor additie-subtractie, reversibiliteit en compensatie-argumenten varieerde tussen respectievelijk .98 en .99; .96 en .98, .98 en 1.00.

Voor elk type scoringscriterium werd per leeftijdsniveau een totaalscore berekend. De frequentieverdelingen van deze totaalscores waren U-vormig voor de leeftijdsniveaus 1 t/m 5, terwijl voor de leeftijdsgroepen 6 t/m 9 monotoon stijgende verdelingsvormen voor elk van de scoringssystemen werden gevonden. Vanwege deze verdelingsvormen werd bij toetsing gebruik gemaakt van non-parametrische toetsen.

Voor elke leeftijdsklasse werd het percentage juiste oordelen (p-waarden) berekend volgens elk scoringscriterium voor alle 13 conservatietaken. Tabel 2 geeft de p-waarden van de conservatietaken, berekend volgens het Goldschmid en Bentler criterium. Het blijkt dat alle conservatietaken in klas 1 een bodemeffect (zie Kingma, 1983a,b, Kingma & Koops, 1983: p-waarde $\geq .10$) opleverden, terwijl in klas 3 een bodemeffect is waar te nemen voor slechts drie taken. Voor de klassen 4 t/m 7 kan een geleidelijke toename in de p-waarden van alle dertien taken worden waargenomen, terwijl er in klas 8 en 9 voornamelijk plafond-effecten (p-waarden $\geq .90$) optreden. De toename in de p-waarden voor de taken in de rijen van Tabel 2 — over de klassen 1 t/m 9 — kan worden opgevat als een transversale (ontwikkelings-)trend in conservatieverwerving, volgens het Goldschmid en Bentler-criterium.

Wat het beoordeling-alleen criterium betreft: dit criterium blijkt minder stringent dan het Goldschmid en Bentler (+ uitleg) criterium (zie Tabel 3). Voor de jongste kinderen uit klas 1 hadden vier taken p-waarden, die groter zijn dan het bodemeffect van .10, terwijl volgens het Goldschmid en Bentler criterium bij deze kinderen uitsluitend bodem-effecten werden gevonden (zie Tabel 2). Verder blijken alle p-waarden in de klassen 1 t/m 7 volgens het beoordeling-alleen criterium hoger te zijn dan volgens het Goldschmid en Bentler criterium. Volgens beide criteria worden voor bijna alle taken plafondeffecten gevonden in de klassen 8 en 9. Het is dan ook niet verwonderlijk dat de totaalscores voor de 13 conservatietaken in de klassen 1 t/m 7 volgens het beoordeling-alleen criterium significant hoger zijn dan volgens het

Tab. 3. — Een overzicht van de p-waarden voor de 13 conservatietaken volgens het beoordeling-alleen criterium (N = 462)

conservatietaak	leeftijdsgroep								
	1	2	3	4	5	6	7	8	9
A oppervlakte	.10	.06	.05	.33	.51	.77	.92	.91	.98
B aantal	.19	.31	.41	.85	.88	1.00	.98	1.00	1.00
C afstand	.16	.15	.18	.24	.47	.65	.66	.86	.95
D substantie	.03	.15	.21	.61	.79	.98	.95	1.00	1.00
E omtrek	.13	.15	.10	.39	.58	.81	.81	.95	1.00
F hoeveelheid	.03	.12	.13	.57	.77	.87	.95	1.00	1.00
G discontinue hoeveelheid	.00	.09	.15	.59	.75	.90	.95	1.00	1.00
H gewicht	.03	.24	.15	.50	.63	.79	.90	.95	.98
I bedekt oppervlak	.03	.06	.08	.24	.53	.77	.88	.95	.95
J lengte	.03	.06	.13	.50	.60	.92	.86	.91	.98
K discontinue lengte	.00	.15	.18	.44	.56	.85	.86	.89	.86
L volume (geometrisch)	.03	.12	.15	.48	.65	.83	.88	.95	.95
M volume (fysisch)	.16	.36	.26	.57	.61	.77	.85	.95	.94

Goldschmid & Bentler criterium (Friedman's tweewegs-variantieanalyse $\chi^2_{(1)}$ varieerde tussen 8.61 en 11.30, $\chi^2_{\text{crit.}} = 6.64$, $\alpha = .01$). Echter, de totaalscores voor beide criteria verschilden niet significant in de klassen 8 en 9 ($z = 1.28$ en $z = 1.72$). Dit is vermoedelijk het gevolg van de in die klassen geconstateerde plafond-effecten. We kunnen concluderen dat deze resultaten in overeenstemming zijn met die van Brainerd's (1973, a, b) bewering dat het beoordeling + uitleg criterium stringenter is dan het beoordeling-alleen criterium. Onze uitkomsten ondersteunen

Tab. 4. — Een overzicht van de p-waarden voor de 13 conservatietaken volgens het Piagetiaanse criterium (N = 462)

conservatietaak	leeftijdsgroep								
	1	2	3	4	5	6	7	8	9
A oppervlakte	.00	.00	.00	.11	.12	.27	.42	.31	.44
B aantal	.00	.00	.03	.22	.19	.39	.37	.50	.52
C afstand	.00	.00	.00	.02	.04	.14	.19	.25	.24
D substantie	.00	.06	.05	.24	.14	.37	.42	.55	.60
E omtrek	.00	.00	.00	.13	.16	.15	.37	.45	.39
F hoeveelheid	.00	.00	.00	.17	.35	.37	.51	.56	.60
G discontinue hoeveelheid	.00	.06	.03	.20	.35	.35	.49	.56	.50
H gewicht	.00	.03	.00	.09	.12	.29	.46	.42	.47
I bedekt oppervlak	.00	.03	.00	.07	.04	.21	.44	.41	.47
J lengte	.00	.00	.00	.15	.09	.29	.44	.33	.42
K discontinue lengte	.00	.03	.03	.13	.09	.25	.42	.47	.34
L volume (geometrisch)	.00	.03	.05	.20	.18	.42	.54	.63	.58
M volume (fysisch)	.00	.03	.03	.13	.09	.25	.42	.47	.34

voorts die van eerdere studies (Brainerd, 1972; Brainerd & Brainerd, 1972; Gruen, 1966; Gruen & Vore, 1972; Rosenthal & Zimmerman, 1972, 1978; Zimmerman & Rosenthal, 1974).

Bij gebruik van het Piagetiaanse criterium blijken de p-waarden drastisch af te nemen ten opzichte van de beide andere criteria. Voor klas 1 t/m 3 worden uitsluitend bodem-effecten gevonden (Tabel 4), terwijl voor klas 8 en 9 geen plafondeffecten worden gevonden. Verder bleek dat voor alle klassen de totaalscores voor alle 13 conservatietaken volgens het Goldschmid en Bentler-criterium significant hoger zijn dan volgens het Piagetiaanse criterium (Friedman analyse $\chi^2_{(1)}$ tussen 9.64 en 21.53). We kunnen concluderen dat het Piagetiaanse beoordeling + uitleg-criterium het meest stringente criterium van de drie onderzochte criteria is.

Psychometrie

Ter bepaling van de betrouwbaarheid werden item-rest correlaties en Kuder Richardson's homogeniteits-indices (KR20) berekend. Een KR20 $\geq .80$ en item-rest correlaties van $\geq .50$ kunnen volgens Nunnally (1967) als bevredigend worden beschouwd. Met behulp van het beoordeling-alleen criterium werd voor de 13 conservatietaken een KR20 van .95 gevonden voor de gehele steekproef ($N = 462$) en item-rest correlaties tussen .66 en .86. Met het Goldschmid en Bentler-criterium werd een KR20 van .96 gevonden en item-rest correlaties tussen .62 en .86. Met het Piagetiaanse criterium tenslotte werd een KR20 van .89 en itemrest correlaties tussen .53 en .67 gevonden. Conclusie: met alle drie scoringscriteria vormen de 13 conservatietaken — volgens de criteria van Nunnally — een betrouwbare test.

Factoranalyse

Met behulp van factoranalyse kunnen de structurele relaties tussen de 13 conservatie-taken — gescoord volgens de drie criteria — worden onderzocht. Men zou kunnen stellen dat leeftijd daarbij uitgepartialiseerd moet worden (zie b.v. Humphrey & Parsons, 1979). Dit zou in het bijzonder aangewezen zijn als de steekproef bestaat uit proefpersonen met een brede variatie in leeftijden, zoals in deze studie het geval is. Echter, vanuit een ontwikkelingspsychologisch gezichtspunt kan gesteld worden dat het uitpartialiseren van chronologische leeftijd (CA) ongewenst is, aangezien dat inhoudt dat men «..... might throw out the baby of Piagetian development with the bath of CA» (Kohlberg & de

Vries, 1980, p. 176). In het algemeen heeft men dan ook bij het factor-analyseren van Piagetiaanse taken leeftijd niet uitgepartialiseerd (zie voor een overzicht: Horn, 1976; Modgil & Modgil, 1976). Voorts toonden Glass & Stephens (1980) aan dat de resulterende factorstructuur relatief ongevoelig is voor het verwijderen van leeftijd. Op grond van deze argumenten besloten wij leeftijd niet uit te partialiseren.

Vooraf werden enkele data-inspecties uitgevoerd. Ten Berge (1972) wees erop dat verschillen in p-waarden van items, bij factoranalyse kunnen resulteren in artificiële «difficulty factors»; dat geldt speciaal in het geval van dichotome gegevens. Vanwege deze waarschuwing werden de drie correlatiematrixes, waarin de 13 conservatie-items waren geordend volgens toenemende p-waarden, nader bestudeerd. Wanneer de rijtotalen van de correlatiematrix toenemen naarmate de items in p-waarden dichter .5 benaderen, is dit een aanwijzing dat er een verdelingseffect aanwezig is (Ten Berge, 1972). Dergelijke toenamen traden in geen van de drie correlatiematrixes op, zodat mag worden aangenomen dat de bedoelde verdelingseffecten in onze data niet voorkomen.

Na factoranalyse (principale componenten zonder iteratie) van de correlatiematrix van de 13 conservatietaken, voor alle proefpersonen ($N = 462$), volgens het beoordeling-alleen criterium, bleek de eerste component (eigenwaarde ≥ 1.0) 65% van de variantie te verdisconteren. De volgende componenten hadden eigenwaarden < 1.0 ; in overeenstemming met Kaiser's (1981) vuistregel werd slechts één component nader bestudeerd en werd geen rotatie uitgevoerd. Nagenoeg hetzelfde resultaat werd gevonden bij het factoranalyseren van de scores op de conservatietaken volgens het Goldschmid en Bentler criterium. De eerste component had een eigenwaarde ≥ 1.0 en verdisconteerde 68% van de variantie, terwijl de tweede en volgende componenten eigenwaarden < 1.0 hadden.

Een afwijkend beeld werd verkregen voor de factoranalyse van de data volgens het Piagetiaanse criterium. De eerste drie componenten hadden eigenwaarden ≥ 1.0 en verdisconteerden samen 47% van de variantie. Met behulp van Varimax rotatie kon echter geen interpreteerbare factorstructuur worden gevonden.

We kunnen uit de factoranalytische uitkomsten concluderen dat de 13 conservatietaken een zelfde onderliggend begrip vertegenwoordigen, als de scoring althans plaatsvindt volgens het Goldschmid en Bentler criterium of volgens het beoordeling-alleen criterium. Echter: als wordt gescoord volgens het Piagetiaanse criterium wordt een eenduidige

conceptuele samenhang tussen de 13 conservatietaken factoranalytisch *niet* ondersteund.

Tenslotte werd ook nog de mate van overeenkomst tussen de factorladingen matrices bepaald. Na orthogonale procustes rotatie werden daartoe de relevante waarden van de congruentie-coëfficiënt van Tucker (1951) berekend en wel volgens het kleinste kwadraten-criterium (ten Berge, 1972). Het bleek dat de componenten volgens beoordeling-alleen en volgens Goldschmid en Bentler verregaande overeenkomst vertonen: Tucker's congruentie-coëfficiënten varieerden van .99 tot .91 voor de eerste tot de 13de component. Deze uitkomst wijst op een uitzonderlijk grote overeenkomst (zie ten Berge, 1977). Daar staat tegenover dat de factoranalytische uitkomsten volgens het Piagetiaanse criterium slechts in geringe mate overeen bleken te komen met die volgens elk van de beide andere criteria (Tucker's coëfficiënten varieerden van .52 tot .56).

DISCUSSIE

In het hier gerapporteerde onderzoek werd een poging ondernomen het criteriumprobleem in conservatieonderzoek psychometrisch te analyseren. Via totaalscores en p-waarden werd aangetoond dat het beoordeling-alleen criterium significant minder stringent was dan het Goldschmid en Bentler beoordeling+ uitleg criterium. Met «minder stringent» wordt bedoeld dat meer kinderen als conserveerder werden beschouwd. Deze uitkomsten sluiten aan bij wat in eerdere studies werd gevonden (Brainerd, 1972, 1973a, b; Gruen, 1966; Gruen & Vore, 1972; Rosenthal & Zimmerman, 1972, 1978; Zimmerman & Rosenthal, 1974). Met Brainerd (1973a, b) kunnen we ons afvragen of bij gebruik van het beoordeling+ uitleg criterium het «ware» conservatie-niveau van het kind niet wordt verdoezeld.

Echter, het gegeven dat het beoordeling-alleen criterium minder stringent is dan het Goldschmid en Bentler criterium, en die weer minder stringent dan het Piagetiaanse criterium, had kennelijk geen effect op de betrouwbaarheid. Voor alle drie scoringscriteria waren de KR20 en item-restcorrelaties voor de 13 conservatie-taken zeer bevredigend (volgens criteria van Nunnally, 1961).

Langs factoranalytische weg bleek dat de 13 conservatietaken, gescoord volgens het beoordeling-alleen criterium zowel als volgens het Goldschmid en Bentler-criterium, een evidente eenduidige conceptuele samenhang vertonen. Berekening van relevante overeenkomstmaten

toonde voorts aan dat er zeer grote overeenkomst bestaat tussen de beide betreffende factorladingen-structuren. Volgens het Piagetiaanse criterium werd een dergelijke factoranalytische structuur echter niet gevonden. Dit is zeer opmerkelijk, aangezien Piaget zelf op theoretische gronden bij uitstek conceptuele eenduidigheid claimt (berustend op de «reversibiliteit» van het denken). Hierbij valt verder nog aan te tekenen dat de uitkomsten van de, m.b.v. het Piagetiaanse criterium verkregen, principale componenten-analyse zeer weinig overeenkomst vertonen met die volgens de beide andere criteria.

Psychometrisch kan geconstateerd worden dat de 13 conservatietaken een betrouwbaar conservatie-instrument vormen — althans bij gebruik van het beoordeling-alleen of het Goldschmid en Bentler-criterium. Daarbij kan worden aangetekend dat het laatste criterium aanmerkelijk stringenter is dan het eerste. Vanwege Brainerd's (1973a, b) stelling dat een beoordeling + uitleg criterium theoretisch niet kan worden verantwoord, moet men wel bijzondere redenen hebben om het Goldschmid en Bentler-criterium te gebruiken. Anderzijds kan worden vermoed dat met het beoordeling-alleen criterium meer «type I fouten» worden gemaakt, waartegenover wellicht «type II fouten» resulteren uit de toepassing van het Goldschmid en Bentler-criterium. Hoe dan ook: deze fouten-typen hebben kennelijk geen invloed op de psychometrische kwaliteiten.

Per saldo kunnen we vaststellen dat het Piagetiaanse beoordeling + uitleg criterium ongeschikt is — zowel volgens een psychometrisch als een theoretisch gezichtspunt. Volgens de beide andere criteria beschikken we over een betrouwbare conservatietest; een keuze tussen deze twee vereist een nauwgezette theoretische argumentatie.

REFERENTIES

- Beilin, H. (1975). The training of the acquisition of logical operations. In M.F. Rosskopf, L.P. Steffe & S. Taback (Eds.), *Piagetian cognitive development research and mathematical education* (pp. 81-124). Washington DC: National Council of Teachers of Mathematics.
- Berge, J.F.M. Ten (1972). Difficulty factors, distribution effects and the least squares data matrix solution. *Educational and Psychological Measurement*, 32, 911-920.
- Berge, J.F.M. Ten (1977). *Optimizing factorial invariance*. Unpublished Ph.D. dissertation, State University of Groningen, Groningen.
- Bever, T.G., Mehler, J., & Epstein, J. (1968). What children do in spite of what they know. *Science*, 162, 921-924.

- Braine, M.D.S. (1959). The ontogeny of certain logical operations: Piaget's formulations examined by nonverbal methods. *Psychological Monographs*, 73 (5, Whole No. 475).
- Brainerd, C.J. (1972). The age stage issue in conservation acquisition. *Psychonomic Science*, 29, 115-117.
- Brainerd, C.J. (1973a). Judgments and explanations as criteria for the presence of cognitive structures. *Psychological Bulletin*, 79, 172-179.
- Brainerd, C.J. (1973b). Neo-Piagetian training experiments revisited: Is there any support for the cognitive-developmental stage hypothesis? *Cognition*, 2/3, 349-370.
- Brainerd, C.J. (1978). *Piaget's theory of intelligence*. Englewood Cliffs, NJ: Prentice Hall.
- Brainerd, C.J., & Brainerd, S.H. (1972). Order acquisition of number and quantity conservation. *Child Development*, 43, 1401-1406.
- Bruner, J.S. (1966). On the conservation of liquids. In J.S. Bruner, R.R. Olver & P.M. Greenfield (Eds.), *Studies in cognitive growth*. New York: Wiley.
- Calhoun, L.G. (1971). Number conservation in very young children: The effect of age and mode of responding. *Child Development*, 42, 561-572.
- Cathart, W.G. (1971). The relationship between primary's students rationalization of conservation and their mathematical achievement. *Child Development*, 42, 755-765.
- Elkind, D. (1961a). Quantity-conceptions in junior and senior highschool students. *Child Development*, 32, 551-560.
- Elkind, D. (1961b). The development of quantitative thinking: A systematic replication of Piaget's studies. *Journal of Genetic Psychology*, 98, 37-46.
- Glass, G.N., & Stephens, B. (1980). Reply to Humphrey's and Parsons' Piagetian tasks measure intelligence and intelligence tests assess cognitive development. *Intelligence*, 4, 171-174.
- Goldschmid, M.L., & Bentler, P.M. (1968). *Manual for the Conservation Assessment Kit*. San Diego: Educational and Industrial Testing Service.
- Gruen, G.E. (1966). Note on conservation: Methodological and definitional considerations. *Child Development*, 37, 977-983.
- Gruen, G.E., & Vore, D.A. (1972). Development of conservation in normal and retarded children. *Developmental Psychology*, 77, 302-316.
- Horn, J.L. (1976). Human abilities: A review of research and theory in the early 1970's: In M.R. Rosenzweig & L.W. Porter (Eds.), *Annual Review of Psychology*, 1976.
- Humphrey, L.G., & Parsons, C.K. (1979). Piagetian tasks measure intelligence and intelligence tests assess cognitive development: A reanalysis. *Intelligence*, 3, 369-382.
- Kaiser, H.F. (1961). A note on Guttman's lower bound for the number of common factors. *Journal of Statistical Psychology*, 14, 1-2.
- Kingma, J. (1977). *Een meetinstrument voor conservatie, bestemd voor kinderen van 4 tot 12 jaar*. Rijksuniversiteit Groningen: Instituut voor Persoonlijkeids- en Ontwikkelingspsychologie.
- Kingma, J. (1980). Criteriumprobleem in Neo-Piagetiaans onderzoek. *De Psycholoog*, 15, 384-388.

- Kingma, J. (1981). *De ontwikkeling van kwantitatieve en relationele begrippen bij kinderen van 4-12 jaar* (Academisch proefschrift). Rijksuniversiteit Groningen. (Handelseditie: Lisse: Swets & Zeitlinger, 1981.)
- Kingma, J. (1983a). Piagetian tasks and traditional intelligence as predictors of performance on addition and subtraction tasks in primary school grades one and two. *Journal of Psychology*, 115, 39-53.
- Kingma, J. (1983b). The development of seriation, conservation and multiple classification: A longitudinal study. *Genetic Psychology Monographs*, 43-69.
- Kingma, J., & Koops, W. (1983). Piagetian tasks, traditional intelligence and achievement tests. *British Journal of Educational Psychology*, 53, 278-290.
- Kohlberg, L., & De Vries, R. (1980). Don't throw out the Piagetian baby with the psychometric bath: Reply to Humphrey and Parsons. *Intelligence*, 4, 175-177.
- Larsen, G. Y., & Flavell, J. H. (1970). Verbal factors in compensation performance and the relation between conservation and compensation. *Child Development*, 41, 965-977.
- Mehler, J., & Bever, T. G. (1967). Cognitive capacity of very young children. *Science*, 158, 141-142.
- Modgil, S., & Modgil, C. (1976). *Piagetian research: Compilation and commentary* (Vol. 4). Windsor GB: NFER Publishing Co.
- Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.
- Piaget, J. (1967). Cognitions and conservations: Two views. *Contemporary Psychology*, 12, 530-533.
- Piaget, J. (1968). *On the development of memory and identity*. Worcester, MA: Clark University Press.
- Piaget, J., & Inhelder, G. (1941). *Le développement des quantités physiques chez l'enfant*. Neuchâtel: Delachaux & Niestlé.
- Piaget, J., Inhelder, G., & Szeminska, A. (1948). *La géométrie spontanée de l'enfant*. Paris: Presses Universitaires de France.
- Piaget, J., & Szeminska, A. *La genèse du nombre chez l'enfant*. Neuchâtel: Delachaux & Niestlé.
- Piaget, J., & Voyat, G. (1968). Recherche sur l'identité d'un corps en développement et celle du mouvement transitif. *Études d'épistémologie génétique* (Vol. 24). Paris: Presses Universitaires de France.
- Reese, H. W., & Schack, M. L. (1974). Comment on Brainerd's criteria for cognitive structures. *Psychological Bulletin*, 81, 67-69.
- Rosenthal, T. L., & Zimmerman, B. J. (1972). Modeling by exemplification and instruction in conservation. *Developmental Psychology*, 6, 392-401.
- Rosenthal, T. L., & Zimmerman, B. J. (1978). *Social learning and cognition*. New York: Academic Press.
- Tobin, M. J. (1972). Conservation of substance in the blind and partially sighted. *British Journal of Educational Psychology*, 41, 192-197.
- Tucker, L. R. (1951). A method for synthesis of factor analytic studies. *Personal Research Section No. 984*. Washington, DC: Department of the Army.
- Vingh Bang (1965). Intuition géométrique et déduction opératoire. *Études d'épistémologie génétique* (Vol. 19). Paris: Presses Universitaires de France.

- Zimmerman, B.J., & Rosenthal, T.L. (1974). Conserving and retaining inequalities through observation and correction. *Developmental Psychology*, 10, 260-268.

Vrije Universiteit Amsterdam
Koningslaan 22-24
1075 AD Amsterdam-Zuid

Received September 1984