

# Clusteranalyse in doelstellingenonderzoek\*

C. H. BEUK  
CITO, Arnhem

## Samenvatting

*In doelstellingenonderzoeken gaat het vaak om het inventariseren van doelen. Meestal wordt dan aan relevante respondenten gevraagd hun oordeel te geven over doelstellingen die door anderen zijn geformuleerd.*

*Door de gegevens met behulp van clusteranalyse-techniek op verschillende wijze te groeperen, kunnen interessante vergelijkingen worden gemaakt.*

*In een tweetal onderzoeken werden vergelijkingen gemaakt tussen zowel soorten doelstellingen als soorten respondenten. Doelstellingen werden zodanig gegroepeerd, dat onderscheid kon worden gemaakt tussen doelen die men wenst, doelen die men nastreeft en doelen die in feite waren bereikt. Ook is onderscheid gemaakt tussen communale en differentiële doelstellingen. Respondenten werden bij gegeven specifieke informatie onderzocht op consistent beoordelaarsgedrag. Vastgesteld is, dat binnen een gedefinieerde groep de ene respondent relevanter is dan de andere, en dat dit afhankelijk is van het materiaal dat beoordeeld wordt. Onderscheid is gemaakt tussen respondenten met inconsistent beoordelaarsgedrag en respondenten met duidelijk afwijkende oordelen. Geconcludeerd wordt, dat respondenten die op bepaalde kenmerken zijn gegroepeerd — bijvoorbeeld als docent of student aan een bepaalde opleiding — niet alle per definitie relevant zijn voor bepaalde informatie over onderwijsdoelstellingen. In veel onderzoeken en niet alleen in doelstellingenonderzoeken, worden ondervraagden beschouwd als consistente en onafhankelijke variabelen. Dit kan wijzen op onvoorzichtigheid van de onderzoeker.*

## 1. Inleiding

Eén van de grootste problemen in doelstellingenonderzoeken is het beschikken over geschikte methoden. In 1970 sprak De Corte (1970, 12) ten aanzien

\* Voor het schrijven van dit artikel ontving ik waardevolle aanwijzingen van de heer W. J. E. Knops, methodoloog van het CITO.

van het ontbreken van goede methoden voor het inventariseren van doelstellingen zelfs over een fundamentele lacune. Sederdien is er echter wel het een en ander gebeurd. Het aantal doelstellingenonderzoeken dat op dit moment loopt of opgezet wordt, is een veelvoud van de onderzoeken die toen werden gehouden. De redenen waarom deze onderzoeken plaatsvinden kunnen zeer verschillend zijn, maar in opzet en uitvoering komt steeds meer lijn. De begrippen die men hanteert, geven niet meer de verwarring van enkele jaren geleden. Aan de methoden die toegepast worden stelt men hogere eisen. Onder meer dat de methode betrouwbaar en objectief is. Bovendien worden de methoden, althans in een empirisch doelstellingenonderzoek, op hun waarde voor het onderzoek getoetst. Een methodologie voor doelstellingenonderzoeken is bezig te ontstaan. In dit licht gezien is uitwisseling van resultaten, ervaringen en ideeën voor theorieontwikkeling en onderzoek van doelstellingen waardevol. Dit artikel bedoelt hieraan een bijdrage te leveren. Getracht wordt de mogelijkheden van clusteranalyse in doelstellingenonderzoek in een ruim kader te plaatsen. Voor meer specifieke en concrete informatie wordt verwezen naar de onderzoeksverslagen zelf.

In 2 worden enkele veel gebruikte begrippen en onderscheidingen daartussen toegelicht. Deze worden in verband gebracht met een aantal vragen waarmee bij het opzetten van een doelstellingenonderzoek rekening moet worden gehouden.

In 3 worden de mogelijkheden bekeken van clusteranalyse bij het groeperen van onderzoeksgegevens, en in het bijzonder de mogelijkheden van de iteratieve clusteranalyse van Boon van Ostade. Aandacht wordt in dit hoofdstuk ook gegeven aan het probleem van de criteria, waaraan waardeoordelen over doelstellingen moeten voldoen.

## 2. Vragen voor een inventarisatieonderzoek

*Gaat het om doelstellingen die thans gelden, die men wenst, of die in feite bereikt zijn?*

Wanneer onderwijsdoelen worden geïnventariseerd, dient het gebruik dat men van de inventarisatie wil maken, vooraf duidelijk te zijn. Doelstellingen die men inventariseert voor bijvoorbeeld het construeren van studietoetsen voor een eindniveau, behoeven niet het volledig beeld te geven dat men mag verwachten van een inventarisatie voor het ontwikkelen van een onderwijsleerplan, waarin meestal ook niet-cognitieve doelstellingen zullen voorkomen. Bovendien zal ook een inventarisatie van cognitieve doelstellingen voor een onderwijsleerplan in het algemeen niet die specifieke leerdoelen bevatten die nodig zijn voor het ontwikkelen van schoolwerkplannen of het construeren van diagnostische toetsen.

Bij elk inventarisatieonderzoek dient vooraf bepaald te worden om welke soort doelstellingen het gaat en ook hoe specifiek en concreet deze moeten worden geformuleerd.

Uitgaande van een eerder door Velema (1966, 192) aanbevolen onderscheiding, maakt De Corte (1970, 24-28) verschil tussen:

actueel geldende doelstellingen, didactisch wenselijke doelstellingen en feitelijke nagestreefde doelstellingen.

Onder actueel geldende doelstellingen verstaat hij doelstellingen, die men op een bepaald moment in het onderwijs dient na te streven. Didactisch wenselijke doelstellingen zijn de doelen die men op datzelfde moment eigenlijk zou moeten nastreven. Actueel geldende en didactisch wenselijke doelstellingen kunnen met elkaar samenvallen, maar in veel gevallen moet rekening worden gehouden met verschillen tussen doelen die men wenst en doelen die men stelt. Allerlei factoren in het onderwijs of daarbuiten kunnen deze verschillen veroorzaken. Voor de feitelijke nagestreefde doelstellingen geldt hetzelfde. Door allerlei omstandigheden kunnen doelstellingen op het moment van het onderzoek feitelijk afwijken van doelstellingen die men dient of beweert na te streven (de actueel geldende doelstellingen) of eigenlijk zou moeten nastreven (de didactisch wenselijke doelstellingen).

Deze onderscheiding van De Corte is niet alleen van belang voor het gebruik dat men van een doelstellingsinventarisatie wil maken, maar ook in samenhang daarmee, met het bepalen van de personen die informatie moeten geven.

#### *Wie zijn onze relevante respondenten?*

Voor informatie over actueel geldende doelen kan men concreet denken aan personen die in het onderwijs een beleidsuitvoerende taak hebben, zoals: inspecties, directies en docenten.

Relevante informatie over didactisch wenselijke doelen mag men verwachten bij mensen met beleidsvoorbereidende en/of beleidsbepalende taken, zoals: vertegenwoordigers van maatschappelijke groeperingen, personen uit de beroepensector, pedagogische centra en ministerie.

Informatie over feitelijk nagestreefde doelstellingen kan vooral worden gegeven door degenen, die concreet met een bepaald vak of een bepaalde vorm van onderwijs dagelijks te maken hebben, zoals: docenten en leerlingen. Vooral de laatste categorie verdient bijzondere aandacht.

In veel onderzoeken worden leerlingen ten onrechte niet als 'relevante respondenten' beschouwd, hoewel kan blijken, dat zij over de feitelijk nagestreefde doelstellingen bruikbare informatie kunnen geven. Feiten, die wel eens afwijken van meningen die docenten hebben. Met name bij het construeren van studietoetsen is het verschijnsel van onder- en overschatting van leerprestaties door eigen docenten niet onbekend. Ook De Block (1975, 121) wijst op het onderscheid dat Werner maakt tussen 'Lernziele' (Endzustände von Lernprozessen) en 'Lehrziele' (Projektierungen für den Lehrer).

Het begrip 'feitelijk nagestreefde doelstellingen' kan daarom beter worden aangegeven met 'feitelijk bereikte doelstellingen'. Deze term ligt ook dichter bij de oorspronkelijke vraag van Velema (1966, 192), 'worden de nagestreefde doelen bereikt?'.

#### *Hoe relevant zijn onze respondenten?*

Behalve het bepalen van groepen relevante respondenten is het ook nuttig na te gaan hoe relevant een respondent in een bepaalde groep is. Het minste wat van een relevante respondent verwacht mag worden is, dat hij zichzelf niet tegenspreekt. Een methode voor onderzoek van doelstellingen dient het mogelijk te maken na te gaan hoe consistent een respondent is met zijn oordelen over bepaalde doelstellingen.

In een inventarisatieonderzoek naar actueel geldende doelstellingen stelt De Corte (1975, 30) de wenselijkheid om in onderzoek de hypothese te toetsen, dat afwijkingen van individuele oordelen elkaar in die mate zullen compenseren, dat het groepsresultaat per doelstelling een hoge stabiliteit zal bezitten.

Ons inziens is echter niet alleen nodig dat de stabiliteit van het groepsresultaat per doelstelling wordt nagegaan, maar ook de stabiliteit — consistentie — van het individuele resultaat over alle doelstellingen. Met cluster- of factoranalyse is dit aspect van het beoordelaarsgedrag te onderzoeken en te toetsen.

In het algemeen gezegd komt het erop neer, dat

een inventarisatiemethode betrouwbaar moet zijn (bij herhaling moet dezelfde resultaten verkregen worden) en objectief (de resultaten moeten meetbaar zijn).

*Hoe concreet en hoe specifiek moeten de doelstellingen zijn?*

Voor de meetbaarheid van de resultaten is een belangrijke vraag in welke mate doelstellingen concreet geformuleerd moeten zijn. Als iemand om een oordeel wordt gevraagd, mag er geen onduidelijkheid zijn over hetgeen men bedoelt. Hoe duidelijk men met doelstellingen moet zijn, is een zaak waar verschillend over wordt gedacht. Voor het formuleren van leerdoelen voor geprogrammeerde instructie geeft Mager (1966) een drietal kenmerken waaraan de vorm van een leerdoel moet voldoen. De Corte e.a. (1972, 38) voegen hieraan nog een vierde kenmerk toe om er zeker van te zijn dat vorm én inhoud van een doelstelling volledig zijn bepaald. In dit laatste geval is de kans op misverstanden wel bijzonder klein en zeker nodig wanneer een inventarisatie van lesdoelstellingen (leerdoelen) plaats vindt, bijvoorbeeld ten behoeve van geprogrammeerde instructie of de constructie van diagnostische toetsen. Dit wil echter niet zeggen dat men zich in alle doelstellingenonderzoeken, zoals dat soms bijna dwangmatig gebeurt, aan deze vier kenmerken zou moeten houden.

De vraag hoe concreet een doelstelling geformuleerd moet worden, heeft namelijk ook een praktische kant. Hoe erg is het als men in een onderzoek over een aantal doelstellingen ten gevolge van misverstanden geen bruikbare informatie krijgt? In sommige onderzoeken kan het efficiënter en zinvoller zijn om over een bepaald onderwerp meer doelstellingen te formuleren dan tijd te stoppen in het verder vastnagelen van een kleine verzameling doelen aan alle mogelijke eisen.

De vraag naar de concreetheid hangt ook samen met de vraag naar de deskundigheid van degenen die in een onderzoek de doelstellingen beoordelen.

Waar goede verstaanders aan een half woord genoeg hebben, zal men in andere gevallen veel concreter moeten zijn met het omschrijven van leerdoelen. Behalve de vraag hoe concreet een doelstelling geformuleerd moet worden, is het in doelstellingenonderzoek ook belangrijk te bepalen hoe specifiek een doelstelling moet zijn.

Doelstellingen voor eindniveaus en leerplanconstructie behoeven meestal niet zo specifiek te zijn als nodig is bij geprogrammeerde instructie of diagnostische toetsen (1). In het eerste geval kan dit nodig zijn, om de overzichtelijkheid van het geheel te be-

waren en in het leerplan plekken open te laten, die de docent in zijn specifieke situatie het beste zelf kan invullen (De Block, 1975, 143-149). Voor de meeste niet-cognitieve doelstellingen is het trouwens ook niet mogelijk om in een leerplan aan te geven met welke resultaten elk lesuur moet worden afgesloten. Wij zijn overigens wel van mening dat ook niet-cognitieve doelen in termen van duidelijk waarneembare gedragingen of gedragsveranderingen geformuleerd moeten worden (2).

In ieder geval geldt: de soort doelstellingen die men gaat inventariseren en de deskundigheid van de respondenten bepalen de mate waarin doelstellingen concreet geformuleerd en specifiek moeten zijn.

*Is er in onze verzameling doelstellingen een ordening aanwezig of te ontdekken?*

Een volgend probleem is het ordenen van de doelstellingen bij het inventariseren.

Soms kan een ordening op vakdidactische gronden van tevoren zijn aangebracht, zoals in het hiervoor genoemde inventarisatieonderzoek met reken-doelstellingen (De Corte, 1975), en moet met deze ordening rekening worden gehouden.

In andere gevallen is het ook mogelijk dat bij het inventariseren naar een ordening wordt gezocht. Dit kan zin hebben bij doelstellingen die niet vanuit een bestaand vak bepaald kunnen worden. Voorbeelden hiervan zijn het onderzoek van Meuwese (1970, 53-70) naar de didactische wenselijke doelstellingen voor een onderwijsprogramma bedrijfskunde aan de T.H. Delft en een inventarisatieonderzoek van meningen over de didactisch wenselijke doelstellingen van het nieuwe vak algemene technieken (Kommissie . . ., 1974). Ook kan het zijn, dat door allerlei ingrepen in een opleiding, het zicht op de doelstellingen verdwenen is. Vooral wanneer in een dergelijke situatie toetsen moeten worden gemaakt, kan het nodig zijn de doelen en de bestaande ordening daartussen opnieuw te bepalen. Een voorbeeld hiervan is een onderzoek naar de nagestreefde einddoelstellingen van de N20-opleiding voor de vakken pedagogiek en psychologie (Beuk, 1974).

In dit onderzoek is een variant op de empirische constructiemethode van Meuwese ontwikkeld en beproefd. In het onderzoek voor het vak algemene technieken is deze methode nogmaals toegepast.

Bij het beoordelen van de geschiktheid van deze methode voor doelstellingenonderzoek dient men in het oog te houden dat deze gebruikt is voor het bepalen van einddoelen. Het zoeken naar enige ordening daarin vond plaats op basis van beoordelingen van hiervoor deskundig geachte personen.



Hieruit dienden aanwijzingen te komen voor het (opnieuw) classificeren van de doelen.

*Hoe snel kunnen de gegevens verzameld en verwerkt worden?*

Een praktisch punt in veel onderzoeken is verder nog, dat grote hoeveelheden informatie snel verwerkt moeten kunnen worden. Dit is vooral van belang wanneer doelstellingen voor continue evaluatie worden geïnventariseerd. Gewenst is daarbij bovendien dat dit kan zonder al te specifieke kennis. Dit betekent dat informatie machinaal (computer) verwerkt moet kunnen worden.

*Samenvattend* menen wij dat een inventarisatiemethode rekening moet houden met de volgende vragen:

- Om welke soort doelstellingen gaat het?
- Welke groepen relevante respondenten kunnen informatie geven?
- Kan het gedrag van respondenten op relevantie en consistentie worden onderzocht en getoetst?
- Hoe concreet en hoe specifiek moeten de doelstellingen worden geformuleerd?
- Wordt bij het inventariseren rekening gehouden met een bepaalde ordening van de doelstellingen?
- Kunnen grote hoeveelheden informatie snel worden verwerkt?

### 3. De iteratieve clusteranalyse

Gelet op de hiervoor geformuleerde vragen lijkt de empirische constructiemethode van Meuwese zeer geschikt. De methode kan voor vele situaties worden aangepast en maakt het in principe mogelijk het gedrag van respondenten op relevantie en consistentie te onderzoeken en te toetsen.

Bij deze methode wordt uitgegaan van vooraf, door een aantal deskundigen opgestelde uitspraken, die mogelijke doelstellingen kunnen bevatten. Door middel van een groeperingstechniek worden uitspraken geordend, daarna opnieuw geformuleerd en geherordend tot een hiërarchisch geheel van bij elkaar behorende doelstellingen wordt verkregen.

Meuwese (1970, 59 e.v.) gebruikt voor het ordenen twee groeperingstechnieken: de hiërarchische clusteranalyse van Johnson en factoranalyse. Niet duidelijk is waarom beide technieken worden gebruikt. Aangetoond is, dat zowel factoranalyse als cluster-

analyse tot hetzelfde resultaat leiden met de gegevens waar het hier om gaat (Boon van Ostade, 1969).

Clusteranalyse is bij zeer grote aantallen variabelen bovendien efficiënter machinaal uit te voeren dan factoranalyse. De mogelijkheden van clusteranalyse en in het bijzonder van de iteratieve clusteranalyse van Boon van Ostade zullen daarom nader worden bekeken. Factoranalytische terminologie wordt zoveel mogelijk vermeden.

#### *Het basismodel van clusteranalyse*

Het resultaat van clusteranalyse bestaat uit een aantal clusters, die elk afzonderlijk worden gekenmerkt door een bepaalde eigenschap of een verzameling van samenhangende eigenschappen. Het kenmerkende van een cluster kan worden gezien als een theoretisch geconstrueerd kenmerk (het 'construct'), dat dient om bepaalde 'gedragingen' (van bijvoorbeeld beoordelaars) te verklaren en te voorspellen.

Gedrag, opgevat als resultaat van interactie tussen doelstelling en het oordeel van de beoordelaar kan worden uitgedrukt in de interactieformule:

$$g'_{ij} = f_{i1} \cdot a_{1j} + f_{i2} \cdot a_{2j} + \dots + f_{ik} \cdot a_{kj}$$

Hierbij wordt aangenomen dat het geobserveerde en gecodeerde gedrag  $g$  (het oordeel van een beoordelaar over een bepaalde doelstelling) verklaard kan worden door  $k$  factoren of kenmerken. Elke term stelt een dergelijke interactie voor.

Het totaal aantal interacties van  $l$  t/m  $k$  factoren verklaart of voorspelt  $g'$ . De coëfficiënten  $a$  en  $f$  geven aan in hoeverre voor deze factoren respectievelijk doelstelling en beoordelaar relevant zijn. Met  $a$  wordt aangegeven in welke mate een bepaalde doelstelling correleert met andere doelstellingen die door dezelfde eigenschap worden gekenmerkt. De mate waarin het oordeel van een respondent samenhangt met de oordelen van andere respondenten over deze eigenschap wordt met  $f$  aangegeven.

Het gedrag van een doelstelling wordt volledig door één eigenschap gekenmerkt wanneer één van de  $f$ -waarden gelijk is aan 1 en alle andere nul, onder voorwaarde dat alle  $k$ -eigenschappen onafhankelijk zijn. Gedrag  $g'$  is dan uitsluitend een functie van de  $a$ -waarde van deze doelstelling voor die bepaalde eigenschap. In de praktijk zal zich deze situatie niet vaak voordoen. Meest voorkomend zal zijn, dat een doelstelling door meer dan één eigenschap wordt gekenmerkt, en bepaalde  $a$ -waarden hoger zijn dan andere.  $g'$  wordt dan hoofdzakelijk verklaard door een bepaalde  $a$ - en bijbehorende  $f$ -waarde en gesteund of afgezwakt door overige bij bepaalde  $a$ -waarden behorende  $f$ -waarden, die in respectievelijk dezelfde of tegengestelde richting werkzaam zijn.

Een voorbeeld kan dit wellicht illustreren. Uitgegaan wordt van doelstelling  $j$ : het onderscheiden van symptomen behorende bij normaal en afwijkend seksueel gedrag van pubers. Aangenomen wordt, dat doelstelling  $j$  door twee eigenschappen of factoren wordt gekenmerkt: psychopathologie 1 en ontwikkelingspsychologie 2. Doelstelling  $j$  wordt door respondent  $i$  beoordeeld. Respondent  $i$  is voor beide factoren een relevante beoordelaar wanneer zijn  $f$ -waarden voor 1 en 2 duidelijk positief zijn. De  $f$ -waarden kunnen natuurlijk verschillend zijn. Beoordelaar  $i$  kan bijvoorbeeld psychopathologie relevanter vinden dan ontwikkelingspsychologie als doelstelling voor een bepaalde opleiding. Ditzelfde kan ook het geval zijn voor de mate waarin doelstelling  $j$  samenhangt of lading heeft met 1 en 2.

In een docentenopleiding zal het accent meer op het ontwikkelingspsychologische aspect van doelstelling  $j$  liggen, dan in een opleiding voor psychiatrisch verpleegkundigen, waar vooral aan het psychopathologische aandacht zal worden gegeven. De  $a$ -waarden van doelstelling  $j$  zijn dus afhankelijk van het gebruik in een bepaalde situatie. Uit de interactie-formule volgt nu, dat de wijze waarop respondent  $i$  oordeelt over doelstelling  $j$  (in een onderzoek bijvoorbeeld een streepje zet in een Likertschaal bij deze doelstelling) afhankelijk is van de relevantie van zijn oordeel om over psychopathologie als leerdoel te kunnen oordelen ( $f_{i1}$ ) in relatie tot de mate waarin doelstelling  $j$  voor gebruik in een bepaalde opleiding door de eigenschap psychopathologie wordt gekenmerkt ( $a_{1j}$ ), alsmede door de relevantie van  $i$  voor ontwikkelingspsychologie ( $f_{i2}$ ) in relatie tot de mate waarin  $j$  ontwikkelingspsychologisch gezien als doelstelling belangrijk is ( $a_{2j}$ ).

Met behulp van factor- of clusteranalyse kan, zij het op verschillende wijze, worden berekend in hoeverre  $g'$  een functie is van de  $a$ - en  $f$ -waarden van de doelstellingen. Wanneer getracht wordt "doelstellingen" te vinden met overeenkomstige  $a$ -waarden, wordt gesproken over R-analyse. Wanneer getracht wordt "respondenten" te vinden met dezelfde  $f$ -waarden over Q-analyse.

Vooraf in een situatie waarin we te doen hebben met groepen beoordelaars die wel eens verschillend zouden kunnen oordelen is dit laatste interessant. Nagegaan kan worden in welk(e) opzicht(en) verschillende groepen respondenten verschillen of overeenkomen bij het beoordelen van bepaalde soorten doelstellingen. Ook eventuele verschillen binnen een bepaalde groep respondenten kunnen worden vastgesteld.

Zowel R- als Q-analyse kunnen nog in twee situ-

aties worden toegepast; in exploratief onderzoek en in onderzoek waarin hypothesen worden getoetst.

In exploratief onderzoek wordt een gegeven dataset geanalyseerd op eventuele clusters. In hypothesetoetsend onderzoek worden enkele gegeven datasets met elkaar vergeleken. Getoetst wordt of deze datasets van elkaar verschillen dan wel tot dezelfde verzameling variabelen behoren. Voor doelstellingenonderzoek bieden deze mogelijkheden verschillende perspectieven. Bijvoorbeeld, wanneer verschillende groepen respondenten gevraagd wordt aan te geven in welke mate bepaalde doelstellingen didactisch wenselijk en tevens actueel geldend zijn. Wanneer op de scores van deze groepen Q-analyse toetsend wordt uitgevoerd, is het niet alleen mogelijk te onderzoeken in hoeverre deze groepen relevant zijn voor de doelstellingen waarover zij moeten oordelen, maar ook of de ene groep hiervoor wellicht anders oordeelt dan een andere groep. Ook exploratieve Q-analyse heeft aantrekkelijke mogelijkheden. Bijvoorbeeld, wanneer men werkt met één groep respondenten, die naast bekende kenmerken, als leeftijd, ervaring, bevoegdheid en dergelijke, ook een aantal onbekende kenmerken hebben.

Deze bekende en onbekende kenmerken of combinaties daarvan kunnen zoveel invloed hebben, dat de oordelen (scores) van sommige respondenten hierdoor niet consistent zijn.

Het kan dan verstandig zijn de instabiele informatie van deze irrelevante respondenten niet bij de verdere verwerking mee te nemen. Zijn er in een onderzoek veel van deze 'ruis-respondenten' (3), dan kan het erg zinvol zijn het onderzoek voorlopig te staken en te trachten de groep respondenten opnieuw te definiëren

#### *Het scoren van beoordelingen.*

Alle oordelen (scores) van de respondenten over de doelstellingen kunnen worden geplaatst in een scorematrix (figuur 1). In deze matrix staan de oordelen van 1 t/m  $N$  beoordelaars over 1 t/m  $n$  doelstellingen gecodeerd met bijvoorbeeld 1 of 0, respectievelijk met de betekenis wel of geen doelstelling voor of van een bepaalde opleiding. Het oordeel van beoordelaar  $i$  over doelstelling  $j$  is in figuur 1, nul ( $x_{ij} = 0$ ). Bij R-analyse worden deze enen en nullen over de kolommen met elkaar vergeleken en bij Q-analyse over de rijen. Voor beide analyses kan dit door berekening van productmoment correlaties. Waar dit bij R-analyse algemeen gebruikelijk en zonder beperkingen mogelijk is, dienen bij Q-analyse bepaalde voorwaarden te worden gesteld (Boon van Ostade, 1971, 91-93).

		Doelstellingen						
		1	.	.	.	j	.	n
Respondenten	1	.	.	.	.	0	.	.
	.	.	.	.	.	0	.	.
	.	.	.	.	.	1	.	.
	i	.	.	.	.	1	.	.
.	.	.	.	.	0	.	.	
N	.	.	.	.	.	.	.	.

Fig. 1 Scorematrix

De eerste voorwaarde is, dat de maateenheid geen kunstmatige intercorrelaties kan veroorzaken. Wanneer bijvoorbeeld met een vijf- of zevenpuntsschaal wordt gewerkt, kunnen de scores het beste worden gedichotomiseerd.

Bij dichotomiseren van scores is in veel gevallen aan te bevelen de mediaan als grens te nemen, omdat deze centrale maat van de scoreverdeling, in tegenstelling tot een gemiddelde, ongevoelig is voor extreme scores. Extreem hoge of lage scores geven in ons geval, waarin bovendien niet zeker is of de meet-schaal wel equidistant is, mogelijk een vertekend beeld van de aanwezige variantie. De modus als grens te nemen ligt niet voor de hand, omdat bij voorkomende multimodale verdelingen van de scores het probleem ontstaat welke modus dan moet worden genomen.

Behalve de vraag naar de centrale maat welke voor het dichotomiseren wordt gekozen, is er de vraag over welke mediaan de scores moeten worden verdeeld. Moet worden uitgegaan van de mediaan van de scores van alle beoordelaars — normatieve scores — ,of van de mediaan van de scores van elke beoordelaar afzonderlijk — ipsatieve scores — (Cattell, 1944)?. De keus tussen beide mogelijkheden hangt af van het schaalgebruik binnen de groep respondenten. Bestaat de groep uit personen met een gelijk schaalgebruik of uit individuele beoordelaars die verschillende schaalgedeelten gebruiken? In het eerste geval valt het individueel oordeel samen met het groepsoordeel en maakt het eigenlijk geen verschil welke mediaan wordt genomen. In het tweede geval bestaat de mogelijkheid dat variantie van bepaalde respondenten wegvalt wanneer zij alleen het linker of rechter schaalgedeelte gebruiken en de groepsmediaan ongeveer in het midden ligt. Omdat bij veel onderzoeken het schaalgebruik van respondenten vooraf niet bekend is, wordt in het kort

ingegaan op het verschil tussen normatieve en ipsatieve scores.

Figuur 2 laat zien, dat bij dichotomisering van de scores van een drietal beoordelaars (A, B en C) beschouwd als normatieve scores, er geen gedifferentieerde samenhang tussen de scores wordt gevonden. Uitgaande van ipsatieve scores blijken uit de gelijk en tegengestelde gerichte 'score-profielen' van de respondenten niettemin positieve en negatieve samenhangen. In heterogene groepen zouden dergelijke coherenties mogelijk toch tot clustervorming kunnen leiden.

Boon van Ostade (1971, 95) vond bij toepassing van zowel normatieve als ipsatieve scores, dat ipsatieve scores onduidelijke resultaten opleverden. Hij wijt dit aan het mogelijk ontbreken van een dominerend algemeen cluster in het betreffende onderzoek. Hij sluit echter de mogelijkheid van goede resultaten niet uit, waar dat wel het geval is.

Onze ervaring (Beuk, 1974), waarbij doelstellingen

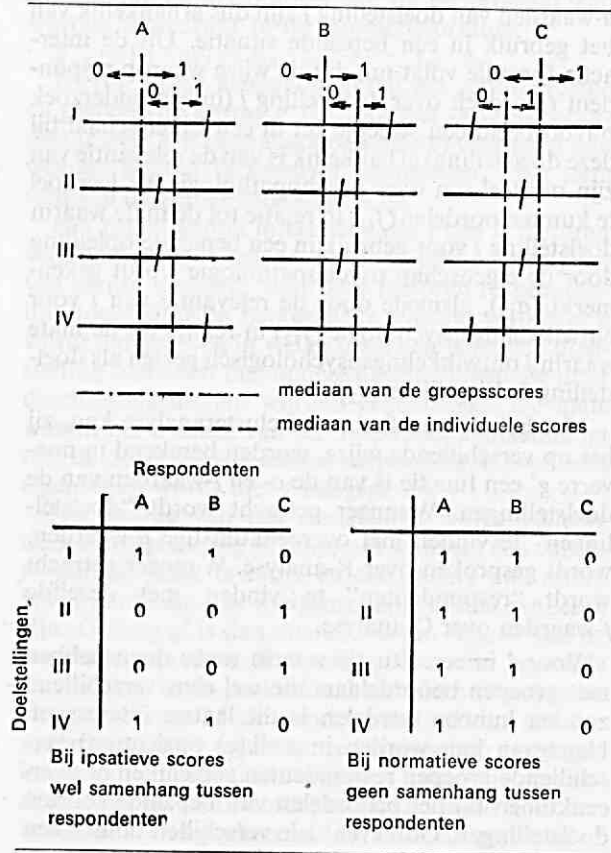


Fig. 2. Dichotomisering met ipsatieve en normatieve scores.



met behulp van trefwoorden werden geïnventariseerd, steunt deze veronderstelling. R-analyse met normatieve scores gaf geen resultaten van enige betekenis. Herhaling met ipsatieve scores leverde enkele zeer grote en een aantal kleine clusters doelstellingen op. Aangenomen mag worden, dat hetzelfde ook geldt bij Q-analyse. Dit betekent in feite dat ook een heterogene groep respondenten over doelstellingen een relevant oordeel geven.

#### *Hoe komen de clusters tot stand? (4)*

Uitgaande van de in figuur 1 weergegeven datamatrix worden allereerst de correlaties berekend van elke variabele met de somscore van alle variabelen. Gebleken is, dat hierdoor de kans om clusters te vinden het grootst is (Boon van Ostade, 1971, 94). In tegenstelling tot de hiërarchische clusteranalyse (Meuwese, 1970, 60 e.v.) worden daarom niet de correlaties berekend tussen elk paar variabelen. Bovendien wordt bij de iteratieve clusteranalyse de datamatrix aan het begin niet omgezet in een intercorrelatiematrix, zoals bij de hiërarchische clusteranalyse van Johnson (Meuwese, 1970, 61). Overigens verlopen beide technieken vrijwel gelijk. De variabele met de hoogste (point-biserial) correlatie vormt het beginpunt van een cluster, mits deze correlatie hoger is dan een vooraf opgegeven grenswaarde — het criterium. Alvorens deze variabele wordt gekozen, wordt eerst getracht door “spiegelen” zoveel mogelijk variabelen positief met de genoemde somscores te laten correleren. Onder “spiegelen” wordt verstaan dat de richting van de meting wordt omgedraaid: hoog wordt laag, laag hoog. Voor de overige variabelen worden de correlaties met deze variabele berekend. De variabele die het hoogst correleert wordt aan de eerste toegevoegd, als de correlatie van de tweede variabele boven het criterium ligt. De gemiddelde score van twee variabelen is de vervangingswaarde van de twee en wordt als een nieuwe variabele beschouwd waarmee de eerste twee kunnen worden voorspeld of verklaard. Het (iteratieve) proces begint nu opnieuw met deze nieuwe variabele als voorspeller van de eerste twee. Weer wordt gekeken naar de variabele die het hoogst correleert met de somscore of gemiddelde score van de eerste twee variabelen. Is de correlatie hoger dan het criterium, dan wordt deze variabele met de eerste twee variabelen samengenomen en wordt de nieuwe gemiddelde score beschouwd als de eerste voorspeller van een cluster van drie variabelen. Een cluster is voltooid, wanneer de gestelde grenswaarde — het criterium — niet meer wordt overschreden.

Het proces begint dan opnieuw met de overgeble-

ven variabelen tot er een tweede cluster is gevormd. Dit is althans het geval wanneer de clusters exploratief worden gevormd. Bij de hypothese-toetsende methode worden bij het bepalen van nieuwe clusters de reeds eerder geclusterde variabelen wel mee in beschouwing genomen. Nagegaan moet dan immers worden of de eerder in een cluster opgenomen variabelen mogelijk ook nog in een ander cluster thuishoren.

Wanneer er geen variabelen meer over zijn, óf de correlaties beneden het criterium komen, is de analyse geëindigd. Het aantal clusters dat gevormd zal worden hangt niet alleen van de betreffende data af, maar ook van het gekozen criterium. Door het kiezen van verschillende grenswaarden is het mogelijk clusters te vormen, die meer of minder algemeen zijn.

Bij een hoog criterium worden kleine clusters gevormd van specifieke aard, bij een laag criterium zullen relatief grote clusters ontstaan van meer algemene betekenis.

Hoe hoog het criterium gesteld moet worden, hangt bovendien af van het doel waarvoor de resultaten worden gebruikt. In ons trefwoordenonderzoek (Beuk, 1974) werd bij Q-analyse 0,30 beschouwd als een redelijke samenhang tussen de mening van een individuele beoordelaar en zijn groep. In het onderzoek Algemene Technieken (Kommissie . . ., 1974) is bij R-analyse van de doelstellingen de grens zelfs bij 0,50 gesteld om de kans op misverstanden nog kleiner te maken.

Een belangrijke vraag is, of de gevormde clusters voldoende betrouwbaar en valide zijn. De betrouwbaarheid kan worden afgeleid uit de mate waarin de in een cluster opgenomen variabelen intern consistent zijn. De interne consistentie kan worden berekend met de bekende formule 20 van Kuder-Richardson (KR 20), die de ondergrens van de betrouwbaarheid aangeeft. Hoe hoog de betrouwbaarheid van een cluster moet zijn, hangt ook weer af van de situatie waarin de gegevens worden gebruikt. Het maakt verschil of men de gegevens gebruikt om een doelstellingendiscussie steun te geven, dan wel zich op deze gegevens baseert bij het construeren van eindtoetsen. In het laatste geval dient de betrouwbaarheid hoger te zijn, dan in het eerste geval. In de hiervoor genoemde onderzoeken werden voor exploratieve analyses respectievelijk de waarden 0,70 en 0,80 aangehouden en bij Q-analyses-toetsend 0,50 en 0,70.

Met de validiteit ligt een en ander minder eenvoudig.

Voor het bepalen van de predictieve validiteit zou idealiter de samenhang moeten worden nagegaan

tussen de variabelen in een cluster en een extern criterium voor het begrip of construct waarmee het cluster aposteriori kan worden geduid. Dit is vaak niet mogelijk of zeer moeilijk. In sommige gevallen kan dit nog wel eens worden nagegaan. Dit is bijvoorbeeld bij Q-analyse mogelijk, wanneer van groepen respondenten bepaalde kenmerken met andere tests of schalen kunnen worden gemeten. De concurrent validiteit van een cluster is hoog, wanneer de scores van de respondenten op de test of schaal hoog correleren met hun ladingen in een cluster waarmee bedoeld is dezelfde kenmerken te verklaren of te voorspellen. Boon van Ostade vermeldt een onderzoek waarin is onderzocht of via exploratieve Q-analyse gevonden clusters concurrent valide zijn. Zijn voorlopige conclusie op grond van dat onderzoek is "... dat aan betrouwbare typen, die door middel van de cluster-analyse worden gevonden, reële betekenis gehecht kan worden" (Boon van Ostade, 1971, 97),

Het ligt voor de hand, dat deze conclusie niet exclusief geldt voor Q-analyse, maar ook van toepassing is op door R-analyse verkregen betrouwbare clusters van doelstellingen.

Het bepalen van de begripsvaliditeit (construct-validity) ligt iets gemakkelijker. Hiervoor kunnen betrekkelijk eenvoudige controles worden uitgevoerd, zoals: het berekenen van de divergente en convergente validiteit en het bepalen van de inter-beoordelaar betrouwbaarheid (face-validity).

De divergente validiteit wordt bepaald door het berekenen van de (inter-) correlaties tussen de gevonden clusters. Zijn deze correlaties laag, dan hebben de clusters ook verschillende betekenissen. De convergente validiteit hangt af van de (intra-)correlaties binnen een cluster. Deze correlaties dienen hoog te zijn. Bij hoge intra-correlaties is ook de betrouwbaarheid van het cluster hoog. Voor betrouwbare clusters is het berekenen van de convergente validiteit daarom niet nodig. De inter-beoordelaar betrouwbaarheid kan worden vastgesteld door de clusters ter duiding voor te leggen aan een aantal deskundigen en de mate van overeenstemming tussen hun oordelen te bepalen.

Tenslotte rest nog de vraag hoe representatief de clusters moeten zijn.

Wanneer alle in een dataset aanwezige variabelen in clusters worden opgenomen, is de representativiteit van deze clusters voor het domein variabelen dat onderzocht wordt 100%. Of echter een variabele wel of niet wordt opgenomen, hangt af van de gekozen grenswaarde. Bij lager stellen van het criterium worden weliswaar meer variabelen opgenomen en neemt

de representativiteit toe, maar zonder veel zin, omdat de bijdrage van die variabelen aan het construct van de cluster dan ook afneemt. Door de correlaties van alle aanwezige variabelen met een cluster te berekenen (de cluster- of factorladingen) kan overigens altijd achteraf nog worden bekeken hoe ver de grenswaarde omlaag kan. Of misschien wel omhoog moet. Op deze wijze kan een zeker optimum worden gevonden. Meestal zal de representativiteit dan ongeveer 50% zijn. Beter dan bijstellen achteraf op de gevonden waarden is er voor te zorgen, dat respondenten niet onzeker zijn over de bedoelingen van de onderzoeker. In ons onderzoek (Beuk, 1974) bleken trefwoorden ook bij relevantere respondenten misverstanden op te leveren. De resultaten waren zeker beter geweest wanneer de doelstellingen wat concreter waren aangegeven. In het onderzoek Algemene Technieken (Kommissie . . . , 1974) werden daarom doelstellingen gebruikt die vrijwel alle voldeden aan de twee belangrijkste criteria van Mager (1966): gebruik van werkwoorden die waarneembare gedragingen aangeven, en aanwijzen van de belangrijkste middelen en omstandigheden waarmee en waarin dit gedrag tot stand moet komen.

In het trefwoordenonderzoek en het onderzoek Algemene Technieken bleken de gevonden clusters zeer treffend en gemakkelijk te kunnen worden aangeduid.

Er werden grote en kleine, gelijkgerichte en tegengestelde clusters gevonden, waarin communale en differentiële doelstellingen konden worden onderscheiden.

Q-analyse in het trefwoordenonderzoek leverde geen verschillen op tussen actueel geldende, didactisch wenselijke en feitelijk bereikte doelstellingen, maar wel verschillen tussen bepaalde groepen respondenten wat betreft hun oordeel over de relevantie van bepaalde onderwerpen. Met name was er sprake van duidelijke verschillen tussen docenten en studenten.

Q-analyse in het onderzoek Algemene Technieken gaf duidelijk te onderscheiden, zelfs tegengestelde opvattingen te zien tussen groepen docenten over de didactisch wenselijke doelstellingen voor dit vak. Niet alleen wat de inhoud van dit vakgebied betreft, maar ook wat de accentuering naar de psychomotorische, de affectief-dynamische en cognitieve kant van de doelstellingen aangaat.

In het algemeen kan worden gezegd, dat met clusteranalyse belangrijke feiten over doelstellingen kunnen worden ontdekt en dat dit ook zeer snel gaat. Het verder uitwerken van de gegevens in leerplan en toetsen is een zaak van langer adem.



## Noten

1. De tegenstellingen abstract-concreet en algemeen-specifiek worden vaak door elkaar gehaald. Hierdoor wordt dan gemeend dat algemene doelstellingen ook altijd abstract moeten zijn. Een algemene doelstelling kan echter even concreet zijn als een specifieke doelstelling.
2. De belangrijkste reden hiervoor is het verwarren van doelen en middelen, c.q. activiteiten en condities om deze doelen te bereiken, wanneer niet-cognitieve doelstellingen in procestermen worden geformuleerd.
3. Wij noemen deze respondenten 'ruis-respondenten', omdat zij zorgen voor ongewenste verhoging van de onverklaarde variantie tussen de scores. Ruis-respondenten zijn dus geen respondenten met een duidelijk afwijkend oordeel! Respondenten met afwijkende oordelen maken de groep welheterogener, maar daarom beslist niet irrelevanter. Ruisrespondenten verlagen door hun inconsistent gedrag zowel de relevantie van een homogene als van een heterogene groep respondenten.
4. Voor dit gedeelte is onder meer gebruik gemaakt van Programma-Bulletin nr. 18 (1-4-'71) van het Psychologisch Laboratorium der Katholieke Universiteit van Nijmegen, waarin de beschrijving wordt gegeven van het programma ØSTER 3: Iteratieve Clusteranalyse volgens Boon van Ostade. Omdat de praktische betekenis daarvan te gering is, wordt hier niet ingegaan op verschillen met andere clusteranalyses.

## Literatuur

Beuk, C. H., *Het inventariseren van doelstellingen met trefwoorden*, CITO, Arnhem, 1974.

Boon van Ostade, A. H., *De iteratieve clusteranalyse — een klassifikatiemethode voor psychologische data*, (diss.) K.U., Nijmegen, 1969.

Boon van Ostade, A. H., *Over type-onderzoek*, Ned. Tijdschrift v. Psychologie, 1971 (26), 82-101.

Boon van Ostade, A. H., *Enige opmerkingen over de gebruiksmogelijkheden van de Hoofdassenanalyse, de Alpha-Kluster-Analyse en de Iteratieve Clusteranalyse*, in: *Gawein, Nijmeegs Tijdschrift v. Psychologie* 1972 (20), 218-221.

Block, A. de, *Taxonomie van Leerdoelen*, Antwerpen/Amsterdam, 1975.

Cattell, R. B., *Psychological measurement: ipsative, normative and interactive*, *Psychol. Rev.*, 1944 (51) 292-303.

Corte, E. de, *Actualiteit en analyse van de problematiek der onderwijsdoelstellingen*, in: *Tijdschrift voor Opvoedkunde* 1970, (16), 4-39.

Corte, E. de, e.a., *Beknopte Didaxologie* Groningen 1972

Corte, E. de, *Inventariseren van actueel geldende leerdoelen* Leuven, 1975.

Graauw, C. P. H. M. de, *Alpha-Kluster-analyse*, *Gawein, Nijmeegs Tijdschrift v. Psychologie* 1972a (20) 63-84.

Graauw, C. P. H. M. de, *Alpha-Kluster-analyse en andere groeperingstechnieken*, *Gawein Nijmeegs Tijdschrift v. Psychologie* 1972b (20), 203-217.

Kommissie Algemene Technieken, *Rapport van de studiekommissie Algemene Technieken*, Ver, De Samenwerkende Landelijke Ped. Centra, Amsterdam 1974.

Mager, R. R., *Het bepalen van doelstellingen voor geprogrammeerde instructie* (Vert.), Alphen a/d Rijn 1966.

Meuwese, W., *Onderwijsresearch*, Utrecht/Antwerpen, 1970, 53-70.

Veleva, E., *Vraagstukken bij het verzamelen en formuleren van onderwijsdoelen*, in: *Verslag Nationaal Congres Onderzoek van het Wetenschappelijk Onderwijs*, 1966.

## Curriculum vitae:

C. H. Beuk (2-1-'36) M.O. A/B Pedagogiek aan Nutsseminarium te Amsterdam, doctoraal Sociale Wetenschappen aan R.U. te Leiden (hoofdvak Onderwijskunde, bijvak Meth. Soc. Wet. Ond.). Tot 1970 werkzaam in bedrijfsleven en beroepsonderwijs, daarna CITO-Arnhem.

Adres CITO: Oeverstraat 65, Arnhem, tel. 085-455555. Privé: Tramstraat 9, Velp, tel. 085-619107.