

# De ontwikkeling van de Functie Observatie Schalen Santpoort

door G. T. Koopmans

## Samenvatting

In dit artikel wordt de ontwikkeling van een observatieschaal waarmee het niveau van functioneren van psychiatrische patiënten kan worden vastgesteld, beschreven. Uiteengezet wordt hoe de inhoudelijke criteria zijn gespecificeerd, welke een leidraad vormden bij de samenstelling van de schaal in termen van items. Schaalvorming heeft plaatsgevonden, gebruik makend van gegevens ontleend aan een steekproef van 549 patiënten. De resultaten daarvan worden beschreven, alsmede een aantal psychometrische kwaliteiten van de gevormde schalen, i.e. de interne consistentie, de onderlinge afhankelijkheden en het onderscheidend vermogen.

## Inleiding

In dit artikel zal ingegaan worden op de ontwikkeling van de Functie Observatie Schalen Santpoort (FOSS), een instrument bedoeld om het niveau van functioneren van psychiatrische patiënten vast te stellen. Achtereenvolgens zullen worden belicht: het doel van het instrument, de wijze waarop het tot stand is gekomen, de geconstrueerde schalen en de kwaliteit daarvan, en enkele toepassingen. Zoals zal blijken hebben de schalen nog een experimenteel karakter. Voor een definitieve versie zal nog nader onderzoek nodig zijn. Enige gedachten worden aan het slot uiteengezet.

## Doel

De directe aanleiding tot de ontwikkeling van de FOSS-schalen was gelegen in de behoefte aan een instrument voor het meten van effecten van (psychiatrische) behandelingen in het kader van het Evaluatieproject Santpoort. Het primaire doel van de schalen is dan ook het meten van veranderingen in het functioneren van patiënten. Het meten van veranderingen houdt doorgaans in, dat de uitkomsten van twee metingen die op verschillende tijdstippen zijn uitgevoerd, met elkaar vergeleken

worden. Een enkelvoudige meting bestaat derhalve in feite uit het vaststellen van het niveau van functioneren van een patiënt. Wil het voor zijn doel geschikt zijn, dan zal het instrument gevoelig moeten zijn voor veranderingen die in het niveau van functioneren optreden. Behalve voor dit primaire doel is het instrument ook geschikt voor andere toepassingen waarbij het niveau van functioneren een rol speelt. Daarbij valt te denken aan het bepalen van de zorgbehoefte van patiëntengroepen, de vergelijking van patiëntengroepen, plaatsingsbeslissingen en dergelijke.

Het belangrijkste probleem bij de keuze van een instrument wordt gevormd door het criteriumprobleem: welke aspecten van het functioneren moeten door het instrument bestreken worden. In de strategie die uiteindelijk geleid heeft tot het construeren van de FOSS is met name veel aandacht besteed aan het ontwikkelen van deze criteria. In concreto hield dit in dat pas na een grondige kwalitatieve analyse van de behandelprogramma's waarbij het instrument toegepast zou worden, keuzes zijn gemaakt ten aanzien van de te meten functioneringsaspecten. Uitgangspunt daarbij was, dat de te hanteren en te operationaliseren effectmaten ontleend moesten worden aan de behandeldoelen van de diverse behandelprogramma's. Daarnaast diende het instrument ook geschikt te zijn voor het onderling vergelijken van patiënten en behandelingen. Om dit te bereiken zijn de behandeldoelen van een reeks van behandelprogramma's eerst afzonderlijk en vervolgens in hun onderlinge samenhang in kaart gebracht. Deze geordende lijst van behandeldoelen vormde het basismateriaal waaraan de meetbaar te maken functioneringsaspecten ontleend zijn.

### **Werkwijze bij de ontwikkeling van criteria en operationalisering daarvan**

Dit basismateriaal is verkregen door bij een reeks van behandelteams de verschillende leden ervan, zoals de psychiater of arts, psycholoog, enkele verpleegkundigen, maatschappelijk werker, bezigheidstherapeut, creatief therapeut en dergelijke, te interviewen over de doelstellingen van het programma waarbij zij betrokken waren. Dit is gebeurd bij de volgende behandelprogramma's:

- kortdurende behandelig (Dekkers et al. 1980);
- klinische psychotherapie (Dekkers en Koopmans 1980);
- resocialisatie (Swalen en Tack 1981);
- enkele programma's binnen de sectoren chronische psychiatrie en gerontopsychiatrie (Koopmans 1981).

Op deze wijze is een breed spectrum van behandeldoelen lopende van chronische psychiatrie en gerontopsychiatrie tot en met klinische psychotherapie in kaart gebracht. Deze definieerden tezamen, zij het nog in ruwe vorm, de criteria voor het evalueren van de behandelresultaten. Dit materiaal is vervolgens geordend in een viertal categorieën, waar-

van drie verder uitgesplitst waren in een aantal subcategorieën, waarin de diverse specifieke functioneringsaspecten werden ondergebracht. Deze indeling zag er als volgt uit:

*A. Zelfstandigheid en ADL-functies*

A1. Fysieke ADL-functies (wassen, aankleden, eten)

A2. Huishoudelijke ADL-functies (maaltijden verzorgen, schoonhouden)

A3. Sociaal maatschappelijke ADL-functies (geldbeheer, werken)

A4. Dagvulling (dagritme, recreatie/ontspanning)

A5. Kritische zelfstandigheid (zelfstandig medicijnbeheer, signaleren van/hulp inroepen bij onraad)

*B. Sociale functies*

B1. Sociale contacten (bezoeken/ontvangen familie e.d., deelname aan verenigingen)

B2. Sociale vaardigheden (samenwerken, contact leggen)

*C. Somatische functies (motoriek, spraak, etc.)*

*D. Psychische functies*

D1. Gedragsbeheersing/-regulatie (regulatie van symptomen, bijv. 'acting out', alcoholgebruik, zich aan afspraken houden)

D2. Kennende functies (geheugen, waarneming, denken)

D3. Emotionele functies (angstniveau, stemming)

D4. Interpersoonlijke relaties (stabiliteit relaties, identiteit)

D5. Inzicht/zelfkennis (in eigen mogelijkheden en beperkingen, motieven, gevoelens, etc.)

Deze indeling, verder aan te duiden als *a priori*-indeling, was met name bedoeld als een soort hulpconstructie bij de operationalisering van de doelen in meetbare termen. Ze bood de mogelijkheid om te controleren, of beschikbare meetinstrumenten (dan wel het te construeren meetinstrument) inhoudelijk volledig waren, dat wil zeggen: alle relevante aspecten dekten.

In de literatuur is vervolgens gezocht naar instrumenten waarmee deze aspecten te meten zouden zijn<sup>1</sup>. Dit leverde in eerste instantie zo'n 100 instrumenten op. De voorkeur ging uit naar instrumenten die (a) alle functioneringsaspecten bestreken, (b) op grond van observaties door elke discipline te scoren waren, (c) niet arbeidsintensief in het gebruik waren (wat in elk geval instrumenten uitsloot die alleen op basis van een uitgebreid interview te scoren waren), en (d) in zo concreet mogelijke, observeerbare termen uitgewerkt waren. Op grond van deze criteria werd een selectie gemaakt van de meest geschikte instrumenten. Geen enkel instrument bleek volledig aan alle eisen te voldoen. Uiteindelijk zijn stukken van een vijftal instrumenten gecombineerd, waarnodig aangevuld met items ontleend aan in totaal 18 andere instrumenten en items van eigen makelij. De vijf instrumenten die de basis vormden van de FOSS waren:

- de Beoordelingsschaal voor Oudere Patiënten (Van der Kam &

Wimmers, 1971);

- de Physical and Mental Impairment of Function Evaluation (Gurel, Linn & Linn 1972);
- de Nurses Observation Scale for Inpatient Evaluation (Honigfeld, Gilles & Klett 1966);
- de Geriatric Profile Scale (Evenson 1971);
- de Fourteen Symptoms Rating Scale (Wyatt & Kapfer 1968).

Na bewerking (vertaling, aanpassing tot vijf puntsschaal, uniformering van terminologie en lay-out, e.d.) kon beschikt worden over een experimentele versie van de observatieschalen. Deze bestond uit twee varianten, namelijk de FOSS-B (Basisvariant), bedoeld voor alle patiëntencategorieën, en de FOSS-U (Uitgebreide variant), bedoeld voor de diverse categorieën van niet-verblijfspatiënten. De FOSS-B bestond uit 134 items, de FOSS-U uit 241, dat wil zeggen: dezelfde 134 items van de FOSS-B plus een aantal specifieke items<sup>1</sup>.

In totaal 549 patiënten (218 mannen en 331 vrouwen) verdeeld over alle in het ziekenhuis opgenomen patiëntencategorieën zijn op een van deze twee FOSS-varianten gescoord. Daarvan zijn 303 patiënten op de FOSS-U variant gescoord. Dat betekent, dat met betrekking tot de 134 FOSS-B items beschikt kon worden over gegevens van 549 patiënten, en dat van 303 patiënten gegevens beschikbaar waren over de 107 specifieke FOSS-U-items. Deze gegevens zijn gebruikt om schalen te vormen, elk bestaande uit een reeks bij elkaar passende items.

## Schaalvorming

1. *De kwaliteit van de apriorischalen* – Omdat elk item was ondergebracht bij een van de 13 hiervoor genoemde functie-subcategorieën, is allereerst gekeken naar de mogelijkheid om schalen te vormen analoog aan deze indeling. Deze schalen, aangeduid als apriori-schalen, ontstonden door sommering van de betreffende itemscores per schaal (nadat sommige items zonodig waren omgescoord, zodat de richting van de items uniform was), gecorrigeerd voor het aantal items.

Nagegaan is, in hoeverre deze schalen aan een aantal kwaliteitseisen voldeden, waarbij met name is gelet op de betrouwbaarheid (interne consistentie) ervan en de onderlinge samenhang tussen de schalen. Daarnaast speelde ook het onderscheidend vermogen een rol, als voorlopige (en indirecte) indicatie voor de gevoeligheid van de schalen voor verschillen in het functioneren die in de loop van de behandeling kunnen optreden.

Wat het laatste betreft: de schalen bleken goed te differentiëren tussen patiënten van de diverse behandelprogramma's. De betrouwbaarheid van de schalen liet echter hier en daar nog te wensen over: enkele schalen waren duidelijk onbetrouwbaar en daarnaast bevatten op zich betrouwbare schalen items die de betrouwbaarheid ervan negatief beïnvloedden. Tot slot was er bij vrij veel schalen sprake van een niet onbe-

langrijke mate van samenhang gezien de aanzienlijke onderlinge correlaties. Zoiets is ongewenst, omdat het wijst op onduidelijke afgrenzing en overlap tussen schalen, waardoor de betekenis van de ene schaal niet scherp te onderscheiden is van die van een andere (daarmee samenhangende) schaal.

Al met al rechtvaardigden de bevindingen een grondiger analyse van het gegevensmateriaal, waarbij zoveel mogelijk recht wordt gedaan aan de feitelijke, empirisch vastgestelde, samenhangen tussen items, in plaats van op apriori-gronden aangenomen samenhangen.

2. *Schaalvorming op basis van empirische relaties* – Gezien de wijze waarop de criteria ontwikkeld en geoperationaliseerd zijn, is bij de analyses gericht op schaalvorming ervan uitgegaan, dat alle informatie die in het totaal van de items vervat was, in principe ook relevante informatie was, en dat derhalve *behoud van informatie* nagestreefd diende te worden<sup>3</sup>. De strategie die bij de analyses gevolgd is, was er op gericht om binnen dit kader een stelsel van schalen te vormen die zo weinig mogelijk overlap vertoonden en de betrouwbaarheid van de schalen te maximaliseren. Omwille van de eenvoud kon elk item slechts bij één schaal worden ondergebracht en werd de schaalscore gedefinieerd als de ongewogen som van de bijbehorende items, gecorrigeerd voor het aantal items.

Met opzet is gekozen voor een strategie met een conservatief karakter. Er is met andere woorden gekozen voor een benadering waarbij het elimineren van items zo lang mogelijk kon worden uitgesteld, om daarmee te voorkomen, dat er geen informatie verloren zou gaan als gevolg van de gebruikte analysetechnieken.

Bij de vorming van de schalen is gebruik gemaakt van zowel factoranalyse als itemanalyse. Begonnen is met een factoranalyse (volgens de principes van Jöreskog (1963)) op een deel van de FOSS-B-items, namelijk die welke aan een aantal formele vereisten<sup>4</sup> voldeden. Op deze manier zijn een aantal dimensies (factoren) in het materiaal opgespoord, die als startpunt hebben gefungeerd bij de schaalvorming. De betreffende items werden op basis van factorladingen tot diverse factorschalen gegroepeerd. Items die op deze manier niet bij zo'n schaal ondergebracht konden worden en de items die niet bij de factoranalyse waren meegenomen, werden vervolgens met deze schalen gecorreleerd. Deze items werden daarna bij die schaal ingedeeld waarmee ze de hoogste correlatie hadden. Met behulp van itemanalyse is daaropvolgend nagegaan welke items een negatieve bijdrage hadden ten aanzien van de betrouwbaarheid van de schaal waarbij ze ingedeeld waren. Items met een negatieve bijdrage werden verwijderd uit de schaal en zo mogelijk ondergebracht bij een andere schaal. Deze procedure is een aantal keren herhaald totdat geen verbetering meer optrad. Aldus zijn 16 schalen binnen de FOSS-B gevormd.

Deze 16 schalen vormden het startpunt bij de schaalvorming binnen

de FOSS-U-items. Uitgangspunt was dat de 107 items die specifiek zijn voor deze variant zoveel mogelijk ondergebracht moesten worden bij de schalen van de FOSS-B. Dit uitgangspunt is gestalte gegeven door als eerste stap in de analyse factoranalyses uit te voeren op de 16 reeds gevormde schalen tezamen met (selecties uit) de 107 genoemde items. Voor het overige is dezelfde strategie gevolgd als die bij de vorming van de schalen binnen de FOSS-B. Dit resulteerde in een aantal van 24 schalen.

### Degevormde schalen

1. *De schalen van de FOSS-B* – In tabel 1 wordt een overzicht gegeven van de schalen binnen de FOSS-B. Conceptueel kan daarbij een onderscheid gemaakt worden tussen schalen die algemene aspecten van het functioneren representeren (verder aangeduid als ADL-schalen), schalen die naar specifieke aspecten van psychopathologie verwijzen (verder aangeduid als P-schalen) en de schaal voor somatisch functioneren (de C-schaal). In tabel 1 zijn de schalen opgesomd met daarbij vermeld het aantal items waaruit de schaal bestaat en de hoogte van coëfficiënt alpha (= Cronbach's alpha, een maat voor de interne consistentie van een schaal). De laatste maat is met name van belang, omdat de analyses erop gericht waren deze te optimaliseren binnen de randvoorwaarden zoals omschreven onder 'schaalvorming'.

Tabel 1: Overzicht van gevormde schalen binnen de FOSS-B.

| Schaal |                             | Aant.<br>items | Alpha  |
|--------|-----------------------------|----------------|--------|
| ADL1   | Zelfredzaamheid             | 24             | .974   |
| ADL2   | Contactgerichtheid          | 19             | .944   |
| ADL3   | Gemak in de omgang          | 4              | .879   |
| ADL4   | Vitaliteit                  | 5              | .836   |
| ADL5   | Zelfreflectie               | 2              | (.680) |
| ADL6   | Interesse voor buitenwereld | 3              | .844   |
| P1     | Agressie                    | 15             | .918   |
| P2     | Depressiviteit              | 5              | .870   |
| P3     | Verwardheid                 | 7              | .787   |
| P4     | Achterdocht                 | 4              | .798   |
| P5     | Paranoia                    | 2              | (.737) |
| P6     | Lastigheid                  | 5              | .733   |
| P7     | Grootheidswaan              | 4              | .751   |
| P8     | Hallucinaties/wanen         | 7              | .819   |
| P9     | Zelfdestructie              | 2              | (.680) |
| C      | Somatisch functioneren      | 12             | .913   |

Vanwege het nog experimentele karakter van de schalen zijn in tabel 1 alle schalen opgenomen, die voldeden aan de minimale eis dat ze uit twee of meer items dienden te bestaan. Bijgevolg is een aantal schalen in feite niet geschikt voor praktische toepassing. Dat geldt zeker voor die 'schalen' die uit slechts twee items bestaan. Bij deze schalen is het zelfs niet mogelijk om coëfficiënt alpha te berekenen. (De in tabel 1 opgenomen coëfficiënten betreffen correlaties.) Het overzicht moet evenwel ook gelezen worden met een andere optiek. De bedoeling ervan is om een overzicht te bieden van de verschillende aspecten van het functioneren die door de FOSS-B worden gerepresenteerd, zij het deels op een gebrekkige, voor verbetering in aanmerking komende, manier.

De betrouwbaarheid van de schalen is over het algemeen redelijk tot goed te noemen: van de 16 schalen hebben 9 een betrouwbaarheid van .80 of meer, terwijl slechts twee (zeer korte) schalen een betrouwbaarheid van iets minder dan .70 hebben.

Tabel 2: Overzicht van gevormde schalen binnen de FOSS-U.

| Schaal |                                | Aant.<br>items | Alpha  |
|--------|--------------------------------|----------------|--------|
| ADL1X  | Zelfredzaamheid                | 27             | .946   |
| ADL2X  | Sociale zelfredzaamheid        | 33             | .957   |
| ADL3X  | Gemak in de omgang             | 6              | .905   |
| ADL4X  | Extraversie                    | 17             | .910   |
| ADL5X  | Inzicht/zelfkennis             | 12             | .902   |
| ADL6X  | Interesse in de buitenwereld   | 3              | .768   |
| ADL7X  | Sociabiliteit                  | 6              | .825   |
| P1X    | Agressie                       | 16             | .940   |
| P2X    | Depressiviteit                 | 8              | .880   |
| P3X    | Verwardheid                    | 7              | .808   |
| P4X    | Achterdocht                    | 10             | .827   |
| P5X    | Paranoia                       | 4              | .782   |
| P6X    | Klaaggedrag                    | 8              | .811   |
| P7X    | Grootheidswaan                 | 8              | .794   |
| P8X    | Hallucinaties/wanen            | 8              | .797   |
| P9X    | Zelfdestructie                 | 5              | .791   |
| P10X   | Angstigheid/gespannenheid      | 3              | .718   |
| P11X   | Objectrelaties                 | 3              | .868   |
| P12X   | Beïnvloedbaarheid              | 3              | .645   |
| P13x   | Onzekerheid m.b.t. identiteit  | 6              | .729   |
| P14X   | Inzichtelijke communicatie     | 2              | (.608) |
| P15X   | Reactiviteit op contactverlies | 4              | .752   |
| P16X   | Affectarmoede                  | 3              | .648   |
| CX     | Somatisch functioneren         | 13             | .766   |

De onderlinge correlaties tussen de schalen zijn meestal gering (minder dan .40). Met name bij de ADL-schalen zijn daarop een aantal uitzonderingen te constateren<sup>5</sup>.

2. *De schalen van de FOSS-U* – Op een wijze analoog aan die welke bij de FOSS-B is gevolgd wordt in tabel 2 een overzicht gegeven van de gevormde FOSS-U-schalen. Van de in totaal 24 gevormde schalen hebben er 13 een betrouwbaarheid van .80 of meer, terwijl van de overige schalen een aantal van 8 een betrouwbaarheid van meer dan .70 heeft. Deze resultaten zijn alleszins bevredigend te noemen.

Wat de onderlinge samenhang tussen de schalen betreft doet zich in grote lijnen weer hetzelfde patroon voor dat ook reeds bij de FOSS-B-schalen geconstateerd kon worden. Binnen de ADL-schalen treden een aantal behoorlijke sterke samenhangen (correlaties tussen .60 en .80) op. Met name de schalen ADL1X en ADL2X hangen zowel onderling als met een aantal andere schalen behoorlijk sterk samen. Bij de P-schalen vindt men slechts bij uitzondering dergelijke samenhangen.

In de bijlage zijn enkele voorbeelden van items opgenomen, waaruit de schalen zijn samengesteld.

## **Onderscheidend vermogen van de schalen**

Als eerste indicatie van de validiteit van de schalen voor het meten van veranderingen kan het onderscheidend vermogen ervan worden gehanteerd. Indien de schalen niet gevoelig zijn voor verschillen tussen onderscheiden groepen patiënten, dan mag weliswaar niet worden verondersteld dat ze eveneens ongevoelig zijn om veranderingen te meten (omdat de groepen niet op alle schalen hoeven te verschillen), maar indien geen enkel verschil zou kunnen worden vastgesteld, dan zou desalniettemin de geldigheid van de schalen op zijn minst twijfelachtig worden.

1. *Onderscheidend vermogen van de FOSS-B-schalen* – Bij de FOSS-B-schalen is nagegaan in hoeverre ze differentieerden tussen te volgen patiëntengroepen:

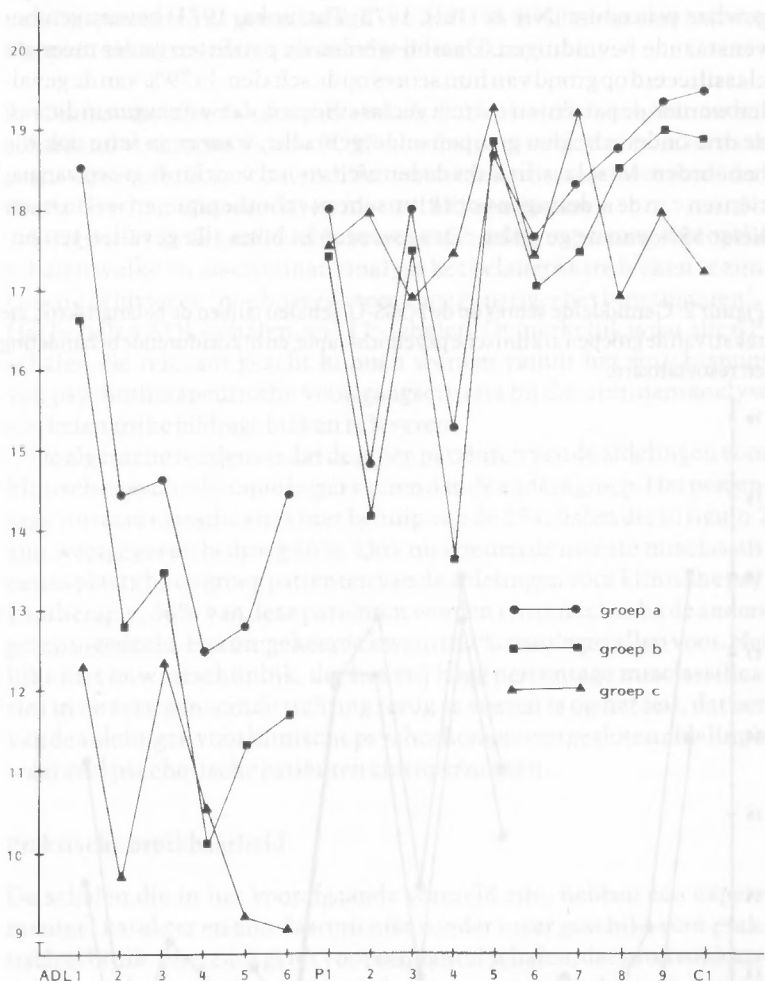
- a. verblijfspatiënten;
- b. patiënten van de afdelingen voor kortdurende behandeling en socialisatie;
- c. patiënten van de afdelingen voor klinische psychotherapie.

De gemiddelde scores van deze drie groepen zijn in figuur 1 weergegeven. Om de interpretatie te vergemakkelijken zijn de schalen, indien nodig, zo getransformeerd, dat voor alle geldt 'hoe hoger de score hoe gunstiger het functioneren'. Dit dient men met name te bedenken bij de P-schalen. Een hoge score op bij voorbeeld P1 (agressie) betekent juist weinig agressie, enzovoorts.

Over het algemeen tekent zich een duidelijke driedeling af, met name bij de ADL-schalen en de C-schaal. Daarbij scoren de verblijfspatiënten



Figuur 1: Gemiddelde scores op de FOSS-B-schalen voor de afdelingen voor: a) klinische psychotherapie, b) kortdurende behandeling en resocialisatie, en c) verblijfspatiënten.

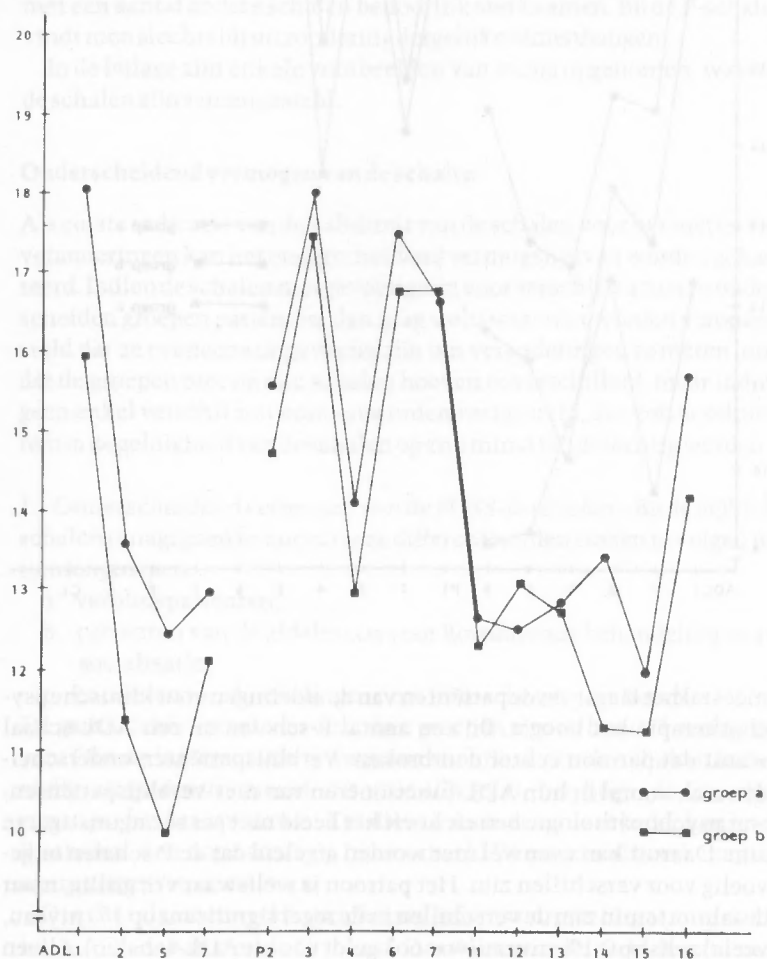


meestal het laagst, en de patiënten van de afdelingen voor klinische psychotherapie het hoogst. Bij een aantal P-schalen en een ADL-schaal wordt dat patroon echter doorbroken. Verblijfspatiënten onderscheiden zich vooral in hun ADL-functioneren van niet-verblijfspatiënten; wat psychopathologie betreft hoeft het beeld niet per se ongunstiger te zijn. Daaruit kan evenwel niet worden afgeleid dat de P-schalen ongevoelig voor verschillen zijn. Het patroon is weliswaar vrij grillig, maar desalniettemin zijn de verschillen in de regel significant op 1% niveau, veelal zelfs op 0,1% niveau (wat ook geldt voor de ADL-schalen). Alleen

de schalen P1 (agressie), P5 (paranoia) en P6 (lastigheid/klaaggedrag) laten geen significante verschillen zien.

Resultaten van discriminantanalyse, gebruik makend van de stapsgewijze procedure (Nie & Hull, 1975, Tatsuoka, 1971) bevestigen bovenstaande bevindingen. Daarbij werden de patiënten onder meer geclassificeerd op grond van hun scores op de schalen. In 79% van de gevallen werden de patiënten correct geclassificeerd, dat wil zeggen in die van de drie onderscheiden groepen ondergebracht, waar ze in feite ook toe behoorden. Misclassificaties deden zich vooral voor bij de groep van patiënten van de afdelingen voor klinische psychotherapie, en wel in maar liefst 55% van de gevallen. Deze werden in bijna alle gevallen ten on-

Figuur 2: Gemiddelde scores op de FOSS-U-schalen (alleen de belangrijkste; zie tekst) van de groepen a) klinische psychotherapie, en b) kortdurende behandeling en resocialisatie.



rechte ingedeeld bij de groep patiënten van de afdelingen voor kortdurende behandeling en resocialisatie. Voorlopig lijkt de conclusie gewettigd, dat voor het onderscheiden van de betreffende groepen van niet-verblijfspatiënten de gevoeligheid van de FOSS-B-schalen tekortschiet. Hetgeen overigens niet verrassend genoemd kan worden.

2. *Onderscheidend vermogen van de FOSS-U-schalen* – Het onderscheidend vermogen van de FOSS-U-schalen is onderzocht door na te gaan in hoeverre deze schalen differentieerden tussen de twee reeds eerder onderscheiden groepen van niet-verblijfspatiënten.

In figuur 2 zijn de gemiddelde scores van beide groepen uitgezet op die schalen welke bij discriminantanalyse het belangrijkste bleken te zijn. Ook nugeldt weer: 'hoe hoger de score hoe gunstiger het functioneren'. Het betreft 4 ADL-schalen, en 11 P-schalen. Opmerkelijk is dat alle 6 P-schalen die relevant geacht kunnen worden vanuit het gezichtspunt van psychotherapeutische voortgangscriteria bij discriminantanalyse een belangrijke bijdrage bleken te leveren.

De algemene tendens is dat de groep patiënten van de afdelingen voor klinische psychotherapie hoger scoren dan de andere groep. Het percentage correcte classificaties met behulp van de 15 schalen die in figuur 2 zijn weergegeven bedroeg 86%. Ook nu vonden de meeste misclassificaties plaats bij de groep patiënten van de afdelingen voor klinische psychotherapie: 36% van deze patiënten werden ten onrechte bij de andere groep ingedeeld. Het omgekeerde kwam in 7% van de gevallen voor. Het lijkt niet onwaarschijnlijk, dat het vrij hoge percentage misclassificaties in de eerstgenoemde richting terug te voeren is op het feit, dat een van de afdelingen voor klinische psychotherapie een gesloten afdeling is waar veel psychotische patiënten zijn opgenomen.

### Praktische bruikbaarheid

De schalen die in het voorafgaande vermeld zijn, hebben een experimenteel karakter en zijn daarom niet zonder meer geschikt voor praktisch gebruik. Enerzijds geldt voor een aantal schalen, dat ze onvoldoende betrouwbaar zijn, anderzijds geldt voor een aantal andere schalen, dat ze weliswaar betrouwbaar genoeg zijn, maar dat het wellicht mogelijk is om te verkorten zonder dat de betrouwbaarheid ervan wezenlijk wordt aangetast.

Om met het laatste te beginnen: zowel bij de FOSS-B als bij de FOSS-U zijn er schalen gevormd die bestaan uit een groot aantal items. Onderzocht is of het mogelijk was deze te verkorten met zo weinig mogelijk vermindering van de betrouwbaarheid door van elk van deze schalen de 10 'beste' items te nemen. Onder de 'beste' items worden verstaan die items welke de hoogste correlatie hebben met de betreffende schaal (in feite de hoogste item-restcorrelatie).

Bij de FOSS/B is dit nagegaan bij de schalen ADL1, ADL2, P1 en C. Bij

de FOSS-U kwamen hiervoor de schalen ADL1X, ADL2X, ADL4X, ADL5X, P1X en CX in aanmerking. Omdat het nogal onpraktisch zou zijn om met twee verschillende verkorte versies te werken van met elkaar corresponderende schalen (zoals ADL1 en ADL1X) is bij deze schalen gezocht naar een compromis tussen de optimale verkorte versies van respectievelijk FOSS-B en FOSS-U.

De betrouwbaarheden (interne consistenties) van deze verkorte versies zijn te vinden in tabel 3.

Tabel 3: Overzicht van betrouwbaarheden (coëfficiënt alpha) van de verkorte versies van schalen

| Schaal  |                          | Alpha<br>FOSS-B | Alpha<br>FOSS-U |
|---------|--------------------------|-----------------|-----------------|
| ADL1(X) | Zelfredzaamheid          | .951            | .906            |
| ADL2(X) | Contactgerichtheid       | .934            | .897            |
| ADL4X   | Extraversie              | n.v.t.          | .895            |
| ADL5X   | Inzicht/zelfkennis       | n.v.t.          | .900            |
| P1(X)   | Agressie                 | .915            | .926            |
| C(X)    | Lichamelijk functioneren | .908            | .762            |

Door in de praktijk gebruik te maken van de verkorte versies kan het aantal items van beide FOSS-versies behoorlijk worden verminderd. De FOSS-B kan dan worden ingekort met 30 items en de FOSS-U met 58 items.

Daar staat echter tegenover dat vrij veel schalen verlengd zouden moeten worden willen ze een betrouwbaarheid van minimaal .80 bereiken. Te berekenen is met welke factor de betrouwbaarheid van een schaal toeneemt als deze verlengd wordt met een gegeven aantal parallelle items (Lord & Novick 1968). Daaruit kan op eenvoudige wijze een formule worden afgeleid die het aantal extra toe te voegen items oplevert.

Als we de balans opmaken van het aantal items dat nodig is om *alle* in de vorige paragraaf vermelde schalen op betrouwbare en efficiënte wijze te operationaliseren, dan zal de FOSS-B zijn samengesteld uit 102 items en de FOSS-U uit in totaal 185 items.

## Discussie

De resultaten van de analyses hebben een tentatief karakter. De gevormde schalen moeten dan ook, zoals al eerder vermeld, experimenteel genoemd worden. Dat leidt als vanzelf tot de vraag wat er moet gebeuren om tot definitieve schalen te komen.

In zijn algemeenheid kan gesteld worden, dat nader onderzoek naar de

validiteit en de interbeoordelaarsbetrouwbaarheid van de schalen noodzakelijk is. Daarnaast zullen de schalen geïkt en genormeerd moeten worden.

In de vorige paragraaf is reeds gesteld dat een aantal schalen voor verbetering (i.c. verlenging) in aanmerking komen gezien hun onvoldoende betrouwbaarheid.

Wat de validiteit betreft het volgende. Over validiteit (geldigheid) kan pas gesproken worden als men aangeeft waarvoor – voor welk doel – de schalen valide dienen te zijn. Daarom ook zal door toepassing van de schalen voor diverse doeleinden in de loop der tijd een meermovattend beeld ontstaan van de validiteit ervan. Het ligt voor de hand in elk geval de validiteit van de schalen te onderzoeken als instrument voor het meten van veranderingen in het functioneren. Daarnaast zijn diverse andere toepassingsmogelijkheden denkbaar. Men mag verwachten, dat voor de ene toepassing een andere selectie van schalen geëigend is dan voor een andere toepassing. Voor differentiatie tussen patiënten voor klinische psychotherapie enerzijds en patiënten voor kortdurende of resocialiserende behandeling anderzijds blijkt dat bij voorbeeld het geval te zijn (zie onderscheidend vermogen FOSS-U). Inmiddels is reeds onderzocht in hoeverre de schalen valide zijn als voorspellers voor plaatsingskeuzes ten aanzien van de programma's voor kortdurende behandeling en resocialisatie, en voor overplaatsingsbeslissingen binnen het resocialisatietraject. Daarbij bleken de schalen voor eerstgenoemde toepassing van slechts marginale betekenis te zijn (kenmerken van biografische aard waren veel belangrijker). Voor de laatstgenoemde toepassing was een groot aantal schalen wel nodig (Koopmans 1984). Meer van dergelijk onderzoek zal moeten volgen. Op deze manier zal het duidelijk worden welke schalen voor welke doeleinden bruikbaar zijn.

Daarnaast kan men de vraag stellen, of de nu gemaakte keuze van een nogal detaillistisch aggregatieniveau wel noodzakelijk is. Wellicht is voor bepaalde doeleinden een zeker verlies aan informatie niet bezwaarlijk. Te denken valt aan bij voorbeeld nogal grofmazige toewijzingsbeslissingen, waarbij wel informatie over een breed spectrum van het functioneren van belang is, maar op een minder gedetailleerd niveau. Uitgaande van hetzelfde itembestand zouden zo diverse varianten ontwikkeld kunnen worden, elk geschikt voor omschreven toepassingen. Het itembestand zelf lijkt, gezien de ervaringen tot nog toe en de wijze waarop het tot stand is gekomen, daarvoor een voldoende brede basis te vormen.

Dit artikel is gebaseerd op een onderzoek waaraan velen een bijdrage hebben geleverd. Voor het gedeelte van het onderzoek dat in dit artikel verwerkt is, zij met name genoemd: mevr. P.D. Bouma, R.K. Buiter, W.R.A. Duurkoop, L.A.J.M. van Eck en P.A.B.J. de Natris. Hen, alsmede J. Mollen J.W.W. Coebergh, dank ik voor hun commentaar op eerdere versies.

## Noten

1. Voor het opsporen van bruikbare instrumenten is gebruik gemaakt van op dat terrein beschikbare overzichten en naslagwerken (*Documentatie van tests en testsresearch*, 1974; Van Eekelen 1978; *Psychological tests and teaching aids*, 1981; Hargreaves et al. 1977; Buros 1978). Daarnaast zijn middels trefwoorden de bestanden doorzocht van Medlars *Psychological abstracts* en *social science search*. Daarbij zijn als criteria gehanteerd: (a) de toepasbaarheid bij de evaluatie van psychiatrische behandelvormen, (b) de toepasbaarheid bij (deel)populaties binnen een psychiatrisch ziekenhuis.

2. Een meer uitgebreide beschrijving van de totstandkoming van de schalen is te vinden in Koopmans et al. (1982) en Koopmans & Bouma (1983).

3. Een meer uitgebreide beschrijving van de analyses en de daaraan voorafgaande overwegingen is te vinden in Koopmans & Buitter (1984).

4. Deze vereisten hadden betrekking op de zgn. normaliteitscriteria, inhoudende dat een item voldeed aan de volgende kenmerken: een scheefheid kleiner dan 1,5, niet tweetoppig, een standaardafwijking groter dan 0,9 en een gemiddelde tussen 1,5 en 4,5.

5. Deels is dit het gevolg van een doelbewuste ingreep. Op conceptuele gronden is namelijk een onderscheid gemaakt tussen de schalen ADL1 en ADL2, hoewel daar statistisch gezien geen argumenten voor aanwezig waren.

## Literatuur

- Kuros, O.K. (ed.) (1978), *The eight mental measurements yearbook, Vol. 1*. Gryphon Press, Highland Park, New Jersey.
- Dekkers, J.M., W.R.A. Duurkoop, G.T. Koopmans en J. van der Zijde (1980), *Doelenonderzoek Hoofdafdeling I; SES-rapport 1*. Provinciaal Ziekenhuis Santpoort, Santpoort-Zuid.
- Dekkers, J.M., en G.T. Koopmans (red.) (1980), *Doelen en middelenonderzoek Hoofdafdeling II; SES-rapport 3*. Provinciaal Ziekenhuis Santpoort, Santpoort-Zuid.
- Documentatie van tests en tests-research* (1974), Nederlands Instituut van Psychologen Amsterdam.
- Eekelen, C.W.J.M. (1978), *Beleving en gedrag in de geriatrie*. Intern rapport, Intervakgroep Sociale Gerontologie i.o., KU Nijmegen.
- Evenson, R. (1971), *Geriatric Profile Scale*. Missouri Institute of Psychiatry, St. Louis, Miss.
- Gurel, D., M.W. Linn en F.S. Linn (1972), Physical and mental impairment-of-function evaluation in the aged: the PAMIE scale. *J. of Gerontology*, 27, 83-90.
- Hargreaves, W.A., C.C. Attkisson en J.E. Sorensen (1977), *Resource materials for community mental health program evaluation*. National Institute of Mental Health, Rockville, Maryland.
- Honigfeld, G., R.D. Gilles en J.C. Klett (1966), A treatment-sensitive ward behavior scale. *Psychological Reports*, 19, 180-182.
- Jöreskog, K.G. (1963), *Statistical estimation in factor analysis: a new technique and its foundation*. Almqvist & Wiksell, Stockholm.

- Kam, P. van der, en M.F.H.G. Wimmers (1971), *Beoordelingsschaal voor oudere patiënten*. Van Loghum Slaterus, Deventer.
- Koopmans, G.T. (red.) (1981), *Doelen en middelenonderzoek Hoofdafdeling IV; SES-rapport 4*. Provinciaal Ziekenhuis Santpoort, Santpoort-Zuid.
- Koopmans, G.T. (1984), *Een eerste validering van de Functie Observatie Schalen Santpoort op de Hoofdafdelingen I en III; SES-rapport 14*. Provinciaal Ziekenhuis Santpoort, Santpoort-Zuid.
- Koopmans, G.T., en P.D. Bouma (1983), *Ontwikkeling van de Functie Observatie Schalen Santpoort voor de Hoofdafdelingen II en III; SES-rapport 9*. Provinciaal Ziekenhuis Santpoort, Santpoort-Zuid.
- Koopmans, G.T., en R.K. Buijter (1984), *Schaalvorming binnen de Functie Observatie Schalen Santpoort; SES-rapport 11*. Provinciaal Ziekenhuis Santpoort, Santpoort-Zuid.
- Koopmans, G.T., W.R.A. Duurkoop en P.D. Bouma (1982), *Ontwikkeling van de Functie Observatie Schalen Hoofdafdeling IV; SES-rapport 7*. Provinciaal Ziekenhuis Santpoort, Santpoort-Zuid.
- Lord, F.M., en M.R. Novick (1968), *Statistical theories of mental test scores*. Addison-Wesley Publ. Comp., Reading, Mass.
- Nie, N.H., C.H. Hull, J.G. Jenkins en K. Steinbrenner (1975), *Statistical package for social sciences* (2nd ed.). McGraw-Hill, New York.
- Psychological tests and teaching aids (1981), Swets & Zeitlinger, Lisse.
- Swalen, A.P.M., en P.J. Tack (red.) (1981), *Doelen en middelenonderzoek Hoofdafdeling III; SES-rapport 5*. Provinciaal Ziekenhuis Santpoort, Santpoort-Zuid.
- Tatsuoka, M.M. (1971), *Multivariate analysis; techniques for educational and psychological research*. Wiley, New York.
- Wyatt, R.J., en D.J. Kapfer (1968), A fourteen-symptom behavioral and mood rating scale for longitudinal patient evaluation by nurses. *Psychological Reports*, 23, 1331-1334.

(Voor bijlage zie p. 554)

---

Schrijver is psycholoog, wetenschappelijk medewerker bij de Studierichting Algemene Gezondheidszorg van de Erasmus Universiteit Rotterdam; voorheen werkzaam bij het Provinciaal Ziekenhuis te Santpoort. Werkadres: EUR-SAG, postbus 1738, 3000 DR Rotterdam.

## Bijlage: enkele voorbeelden van items

De hieronder vermelde items zijn voorbeelden van items die bij een bepaalde schaal horen. Een schaal bestaat uit een reeks van items.

*Voorbeelditems bij schaal ADL2(X): Contactgerichtheid/Sociale Zelfredzaamheid*

- Gebruik van telefoon:
  1. Gebruikt telefoon(ce)l op eigen initiatief (zoekt nummers op, draait ze zelf, enz.)
  2. Kan enkele bekende nummers draaien
  3. Kan alleen van telefoon op afdeling gebruik maken, draait zelf nummers
  4. Neemt zelf de telefoon op, maar belt zelf niet
  5. Gebruikt de telefoon helemaal niet

In FOSS-U toegevoegd aan deze schaal (voorbeeld):

- Werk/studie:
  1. Werkt of studeert in het geheel niet
  2. Werkt resp. studeert binnen het ziekenhuis, bijv. arbeidstherapie op spec. georganiseerde arbeidstherapieplaats
  3. Werkt of studeert buiten het ziekenhuis in relatief beschutte omstandigheden (geen normaal arbeidsproces)
  4. Werkt of studeert buiten het ziekenhuis in normale omstandigheden, maar maximaal 3 middagen per week
  5. Idem, maar dan meer dan 3 middagen p.w.

*Voorbeelditems bij schaal P4(X): Achterdocht*

- Is wantrouwend en achterdochtig
  1. Nooit of niet duidelijk te merken
  2. Bij uitzondering
  3. Soms
  4. Dikwijls
  5. Altijd

In FOSS-U toegevoegd aan deze schaal (voorbeeld):

- Voelt zich snel afgewezen, gekwetst of gekrenkt
  1. Nee, duidelijk niet
  2. Nee, waarschijnlijk niet
  3. Onzeker
  4. Ja, waarschijnlijk wel
  5. Ja, duidelijk wel