

LITERATUURLIJST

- 1 Cronbach, L. J.; Meehl, P. E. (1956) 'Construct validity in psychological tests'. In: *Minnesota Studies in the Philosophy of Science*. Vol. I, Minneapolis.
- 2 Beyaert, F. H. L. *Gedragsbeoordeling in het psychiatrisch ziekenhuis. De P.U.K.-schalen voor gedragsbeoordeling*. Stenfort Kroese, Leiden, 1966. (Diss. Utrecht).
- 3 Fischer, A. A. *Nosologisch denken in de psychiatrie; over theoretische en methodologische problemen*. Elinkwijk, Utrecht, 1963 (Diss. Utrecht).
- 4 Fischer, A. A. 'Enkele problemen bij het samenstellen van proefgroepen voor wetenschappelijk onderzoek in het psychiatrisch ziekenhuis'. Publikatie van de P.U.K.-groep. *Voordrachtenreeks Ned. Ver. Psychiat. Dienstverb.*, Jrg. 8, no. 29 (1966).
- 5 Kraepelin, E. (1918) 'Ziele und Wege der psychiatrischen Forschung'. *Z. ges. Neurol. Psychiat.* 42: 169.
- 6 Kraepelin, E. (1919) 'Die Erforschung psychischer Krankheitsformen'. *Z. ges. Neurol. Psychiat.* 51: 224.
- 7 Kraepelin, E. (1920) 'Die Erscheinungsformen des Irreseins'. *Z. ges. Neurol. Psychiat.* 62: 1—30.
- 8 Kreitmann, N. 'The reliability of psychiatric diagnosis'. *J. Ment. Sci.*, 107 (1961): 876—886.
- 9 v. d. Meer, J. 'Problemen bij de latere stadia van schizofrenie'. Publikatie van de P.U.K.-groep. *Voordrachtenreeks Ned. Ver. Psychiat. Dienstverb.*, 23 (1965) 69—90.
- 10 Rümke, H. C. 'Ueber alte Schizofrene'. *Schweiz. Arch. f. Neurol. Neurochir. u. Psychiat.*, 91 (1963) Heft 1: 201—210.

VOORBEELD VAN EEN MODERNE ONDERZOEKSMETHODE AAN DE HAND VAN EEN UITGEVOERD PROJECT

door Dr. F. H. L. BEYAERT *

Uit de voordracht van collega Fischer hebt u kunnen beluisteren voor welke problemen wij in de praktijk van onze research kwamen te staan en in vele opzichten nog staan. Wij wilden gedragsveranderingen vastleggen terwijl het gedrag van patiënten zich onder invloed van de ziekte, medicamenten en andere therapieën wijzigt. Naast de psychofarmacologische therapie, de creatieve therapie, de arbeidstherapie, de bewegingstherapie, ieder met een eigen discipline, geloven wij sterk in de zgn. socio-therapie d.w.z. in het resocialiseren

* Uit de Prof. Dr. H. C. Rümke Kliniek (hoofd Prof. Dr. J. H. Plokker), Tweede Psychiatrische Universiteits Kliniek deel uitmakend van het Willem Arntsz Huis te Utrecht (geneesheer-directeur: H. Loois).

Medewerkers PUK-groep o.l.v. Dr. A. A. Fischer: Dr. F. H. L. Beyaert, Prof. Dr. J. H. Dijkhuis, Mej. A. van Hilten, G. G. F. Pareau Dumont, Mej. E. Raven, W. G. Vogelaar, E. Werlemann, Mevr. A. C. Zwanikken-Delfos.

de klimaat van de afdeling van een psychiatrisch ziekenhuis. Een klimaat, dat gevormd wordt in de interactie tussen verplegenden en patiënten, waarbij de verplegenden, via de zich ontwikkelende discipline van de psychiatrische verpleegkunde, trachten er een resocialiserende koers aan te geven.

Bij het werkelijk vastleggen van gedragsveranderingen, zoals ons streven was, geldt als eerste eis, dat dit betrouwbaar, in de zin van 'reliable', gebeurt. Dat wil zeggen dat het onderzoek herhaalbaar en overdraagbaar is. In de tweede plaats moeten de gegevens, die met een dergelijk onderzoek verkregen worden, relevant zijn, wat dan weer inhoudt, dat de gegevens van betekenis zijn en in verband moeten staan met het beoogde metingsdoel.

Betrouwbaarheid, reliability, herhaalbaarheid en overdraagbaarheid van een beoordelingsonderzoek is pas gewaarborgd als meerdere mensen het eens zijn over dat wat ze hebben waargenomen bij de ander, bij zichzelf, of in de interactie met die ander. Dit voert regelrecht naar de problematiek van de beoordelaar en het beoordelen.

Bekend is, dat de psychiatrische beoordelingen, voor zover het diagnosticeren betreft, maar ook voor zover het symptomen betreft, naamgeving van ziekelijke gedragingen dus, in ruime mate onbetrouwbaar zijn. Het wordt meer en meer duidelijk, dat psychiaters misschien wel met elkaar kunnen praten maar dat zij zodra er gerubriceerd moet worden, tot grote onenigheid geraken. Getallen als 30 procent onderlinge overeenstemming behoren in de desbetreffende research-literatuur niet tot een uitzondering. Als u zich realiseert, dat dit betekent dat 7 van de 10 diagnoses door 2 psychiaters over dezelfde patiënten gesteld, verschillend luiden, terwijl deze psychiaters het maar in 3 van de 10 keer met elkaar eens zijn, beseft u dat er iets mis moet zijn met het benoemen van de dingen. Sommige onderzoekers menen, dat met deze gegevens het laatste woord gezegd is. Het bewijs is geleverd: de psychiatrie is als een soort charlatannerie ontmaskerd en in ieder geval als een nauwelijks wetenschappelijke zaak terzijde geschoven.

Met wéér anderen ben ik van mening, dat ook dit een voorbarige conclusie is en dat bij nadere beschouwing de beoordelingen, die aan de psychiater in dergelijke onderzoeken worden gevraagd, een beetje lijken op het volgende: 'Is dit een olifant?', indien dit zo is: welke soort, óf komt hij uit Afrika, óf heeft hij geen behoorlijke stamboom, óf is het allemaal gedeeltelijk waar. Indien u het niet weet zegt u het dan'.

Dat neemt ondertussen niet weg, dat de psychiater, practicus als hij is en vaak moet zijn, en misschien geïmponeerd door de vraagsteller die aan research doet, vaak trouwhartig een antwoord tracht te geven zonder te vermelden dat de gestelde vraag eigenlijk niet te beantwoorden en onzinnig is.

Gedeeltelijk lijkt dit zijn oorzaak te vinden in het feit, dat we nog

steeds doen alsof we dezelfde taal spreken. Alsof woorden als: schizofrenie — autisme — slapeloosheid — depressie en noem maar op, niet voor elk psychiatrisch centrum een heel eigen betekenis hebben. Een betekenis die alleen maar door veelvuldige onderlinge gesprekken van een staf, voor die afzonderlijke stafleden dezelfde is; terwijl bij elke communicatiestoornis babylonische spraakverwarringen ontstaan die in de eerste plaats deze technische termen treffen. Misschien is dit in elke wetenschap zo, maar in ieder geval heeft de psychiatrie er een moeilijkheid bij als het om technische termen gaat. Technische termen verwijzen in ons vak niet alleen naar dat wat we ermee willen benoemen maar heel duidelijk ook naar waardeeringen daarachter, die bijv. ofwel moreel ofwel prognostisch zijn. Bij termen als hysterisch en autistisch is dat heel duidelijk. Het eerste is vaak tevens een scheldwoord, het tweede vaak tevens een sombere toekomst. Het hangt dus niet alleen van de opvatting van de beoordelaar over de technische term af wat hij zal gaan zeggen bij een beoordeling, maar ook nog van zijn persoonlijke voorkeur hierin en van de mate waarin hij gemakkelijk of moeilijk kan kiezen tussen goed en kwaad of tussen somber en opgewekt.

Dit alles pleit dus voor een psychiatrisch beoordelingssysteem waarin opgewekt opgewekt is en kwaad kwaad wordt genoemd en waarin veelvuldige technische termen zoveel mogelijk worden vermeden. Het pleit echter ook voor een beoordelingssysteem waarin duidelijk wordt gemaakt met welk doel beoordeeld werd. En dat is trouwens van belang voor de hele geneeskunde. Als iemand zgn. 'genoegzaam hersteld' is en het ziekenhuis kan verlaten moet men zich afvragen wat er allemaal toe heeft bijgedragen, dat de medicus deze beoordeling uitsprak. Was dat omdat de ziekte was verdwenen of teruggedrongen, of omdat er te weinig bedden waren, of omdat de omstandigheden in het eigen milieu van de patiënt zodanig waren dat thuiskomst gewenst was, of omdat de patiënt in het ziekenhuis zo lastig was dat men hem graag kwijt wilde, of omdat men meende dat de patiënt weliswaar niet genezen, maar toch gerevalideerd zich in de maatschappij wel weer zou kunnen handhaven, of omdat er bij röntgenologisch en laboratoriumonderzoek niets afwijkends meer aantoonbaar was, of omdat de patiënt geen klachten meer had, en zo kan ik wel doorgaan. Een beoordeling als 'genoegzaam hersteld' is zo ongehoorlijk veelomvattend en veelvuldig, dat hierdoor alleen al de ontslagdatum in bijv. de psychofarmacologische research als criterium onbruikbaar lijkt, tenzij het om zeer grote aantallen patiënten gaat. Het feit dat 20 van de 30 patiënten, die een bepaald geneesmiddel gebruikten, na verloop van tijd het ziekenhuis 'genoegzaam hersteld' kunnen verlaten, zegt hoegenaamd niets over dat medicament maar veel meer over de dokter, die zich intensief met die patiënten bezig hield en ook andere voorwaarden schiep waardoor het ontslag uit het psychiatrisch ziekenhuis mogelijk was.

Bij al deze problematiek rond de beoordeling, met name in de psychiatrie, voegen zich nog de volgende problemen.

In de eerste plaats is er de deskundigheid of misschien moet ik zeggen de ondeskundigheid van de psychiater als het niet om ziekte gaat, maar bijv. om produktiviteit in een bepaald arbeidsmilieu. Een ander punt is bijv. in dit verband de zwijgzaamheid. Het is onverdraaglijk en sociaal onaangepast indien iemand nooit een woord spreekt, en daar zelfs geen behoefte toe voelt. De psychiater zal uit hoofde van zijn beroep eerder geneigd zijn dit wél te 'accepteren' dan de niet-psychiater en hij zal dus minder geneigd zijn om het zwijgen onaangepast te noemen. Dat wil zeggen: te beoordelen als een onoverkomelijke belemmering voor een terugkeer naar de maatschappij. Door het 'begrijpen' of het 'willen begrijpen' van de zwijgzaamheid verliest de psychiater zijn 'deskundigheid' op het gebied van de beoordeling van sociale aangepastheid.

In de tweede plaats is er de beperktheid van de psychiatrische beoordeling, die in het algemeen wordt gegeven naar aanleiding van het psychiatrisch interview. Het is bekend dat we een gesprek met een patiënt kunnen hebben waaruit we zouden kunnen besluiten, dat de patiënt in kwestie zich naar onze mening goed in de maatschappij zal kunnen handhaven. In het algemeen zijn we echter zo voorzichtig geworden om eerst even aan de hoofdzuster te vragen wat die ervan denkt. Soms merken we dan ineens, dat de hoofdzuster kan vertellen, dat de patiënt elke nacht incontinent is of dat hij 's morgens eerst alle tegels van de badkamer moet tellen voor hij zich kan wassen. Wat ik hiermee wil zeggen is dat de psychiatrische beoordeling, zoals elke beoordeling, situatie-gebonden is en dat niet vanzelfsprekend de beoordeling in situatie A van toepassing is in de situatie B. Dat men dus ook ten onrechte de psychiatrische beoordeling, verkregen uit een gesprek met een patiënt, als maatgevend beschouwt voor het algemene wel en wee van de patiënt. Dat men eveneens ten onrechte meent, dat de psychiater het algemeen functioneren van een mens kan beoordelen, terwijl we toch tegelijkertijd weten, dat iemand door zijn werkgever beschouwd kan worden als een nietsnut, door zijn echtgenote als een held en door de clubgenoten als een bekwaam politicus of een geweldige grappenmaker. Als we dan bovendien nog weten, dat dit sociaal functioneren van iemand juist in evenwicht kan worden gehouden door het feit, dat de beoordelingen in de verschillende situaties verschillend zijn, dan maant dit tot grote bescheidenheid. Met deze bescheidenheid bedoel ik dan, dat de psychiater zich realiseren moet, dat zijn beoordeling aan de situatie gebonden is en in de meeste gevallen niet meer kan zijn dan een beoordeling van het met de patiënt gevoerde gesprek. Wat dus een beoordeling is van de manier hoe een patiënt zich tijdens een gesprek met de dokter gedraagt. Dat wil dus ook weer niet zeggen, dat de practiserend psychiater in het psychiatrisch zie-

kenhuis, geruggesteund door zijn ervaring en na informaties te hebben gekregen van de hoofdzuster, de arbeidstherapeuten, de maatschappelijk werker en wie al meer, niet competent zou zijn om een oordeel te vormen over de mate waarin een patiënt 'genoegzaam hersteld' zou zijn. Het wil alleen maar zeggen, dat men zich, als er een research begonnen wordt, goed realiseert hoe de zaken liggen. Voor de research betekent dit, dat men zich realiseert dat er zeer bepaalde aspecten van het psychiatrisch werk geïsoleerd moeten worden uit de complexe veelzijdige bezigheden van ons alledaagse werk en dat we moeten bedenken waaruit we de dingen, die we onderzoeken, isoleren. Het is om deze reden van groot belang, dat juist de psychiater zich blijft bezig houden met de research in zijn eigen vakgebied en dat dit werk niet overgelaten wordt aan andere wetenschappelijke onderzoekers wanneer deze tenminste niet, evenals de psychiater, in het dagelijks werk zijn ingevoerd. Omdat de psychiater dan weer niet in staat is om het researchgebied te overzien, de methodieken die mogelijk zijn e.d., is het vanuit dit gezichtspunt noodzakelijk om research te doen in teamverband zoals dat in de Rümkekliniek gebeurt.

Er is nog één punt dat ik aan de orde zou willen stellen voor ik u ga beschrijven hoe we een deel van de problematiek in onze kliniek hebben opgelost. Het is het feit, dat een bepaalde situatie, een beoordelingssituatie, zich zal wijzigen door de beoordeling zelf. In experimenteel onderzoek wordt de beoordelingssituatie tijdens deze beoordeling door allerlei maatregelen zoveel mogelijk constant gehouden. U kent dat: om te kijken of iemand iets kan, wordt hem niet voor een tweede of derde keer dezelfde opdracht gegeven, anders heeft hij dat misschien intussen geleerd. Het is wel mogelijk om het zo te doen, maar dan zou de vraag moeten luiden niet zozeer: wat kan iemand in deze opdrachtsituatie, maar veel meer: wat kan iemand, terwijl hij bij herhaling in dezelfde opdrachtsituatie is, hiervan leren?

Een dergelijk probleem doet zich voor bij andere intermenselijke beoordelingsprocedures. Om de herhaalbaarheid en de betrouwbaarheid van een bepaalde beoordeling te kunnen vaststellen is het noodzakelijk om een gestandaardiseerd instrument te vervaardigen. Op dezelfde manier als de koortsthermometer en temperatuurlijst nodig waren om de temperatuurverschillen op betrouwbare manier vast te leggen. Maar ook de koortsthermometer verandert natuurlijk bij het aanleggen de temperatuur in de omgeving, maar we zijn gewoon deze wijziging in de situatie door het toepassen van dit beoordelingsinstrument, te verwaarlozen. Natuurkundigen kunnen ons ook nog verzekeren, dat dat verantwoord is. Maar het wordt moeilijker wanneer de wijziging door de toepassing van het beoordelingsinstrument aanzienlijk is, of van onbepaalbare grootte en hoedanigheid. Een

eenvoudig voorbeeld uit het psychiatrisch ziekenhuis kan dit mis-schien verduidelijken. Vraagt men aan verplegenden om, met behulp van tevoren opgestelde vragenlijsten (zoals schaal 18 van de PUK-schalen), te beoordelen hoe en op welke manier patiënten in een psychiatrisch ziekenhuis met hun geld omgaan, dan zal deze vraag aan de verplegenden de interactie tussen patiënt en verplegenden wij-zigen. We hebben dat in concreto meegemaakt toen de hoofdzuster zich opeens ging afvragen waarom de patiënten op een gegeven mo-ment zoveel met haar spraken over hun geldlijsten. Dat was nadat de andere verplegenden voor het eerst hadden beoordeeld hoe de patiënten met hun geld omgingen. Het vragen, al dan niet op ge-standaardiseerde wijze, naar een bepaald gedrag in een bepaalde si-tuatie kan dus de situatie waarin dat gedrag werd vastgelegd aan-zienlijk veranderen. In dit geval nu toevallig in de goede, althans in de door ons gewenste richting.

Methodologisch heeft zoiets natuurlijk wel konsekventies. Met name: dat het gebruiken van een beoordelingsinstrument op zichzelf een variabele is, die in het oog moet worden gehouden bij de inter-pretatie van de *oorzaken* van de met het beoordelingsinstrument vastgelegde gedragswijzigingen. Dit hoeft echter geen bezwaar te zijn bij een longitudinaal opgezet onderzoek, want gedragswijzigingen wer-den betrouwbaar vastgelegd en hypothesen over de oorzaken hiervan kunnen steeds opnieuw worden gesteld en met behulp van het bij herhaling gebruikte instrument worden geverifieerd.

Op welke wijze nu hebben wij een beoordelingsinstrument (PUK-schalen) gemaakt, dat, voor zover dit tot nu toe is te overzien, aan deze voorwaarden voldoet?

Het uitgangspunt van de PUK-schalen was een gedragsbeoordeling, gebonden aan de situatie van de afdeling van een psychiatrisch zie-kenhuis, waarbij de beoordeling dan ook juist gericht moest zijn op de interactie tussen de beoordeelde en de beoordelaar. We moes-ten daarvoor opnieuw het terrein verkennen, de handboeken vooral thuis laten en de technische termen, ook het vakjargon van de ver-plegenden, kwijt zien te raken. Om aan de basisgegevens te ko-men, dat wat ik later 'observabele eenheden' ben gaan noemen, heb-ben we werkstudenten ingeschakeld. De werkstudenten werden aan-gesteld als verpleeghulp en stonden in die hoedanigheid dus on-der het gezag van de verplegenden en de hoofdzuster. Daarnaast hadden zij een observerende taak waaraan een rapportage vastzat. Na allerlei eerste pogingen, die misliepen, ontstond de volgende observatie-opdracht: aan de studenten werd gevraagd om onafhan-kelijk hun ervaringen met een bepaalde patiënt, met wie zij in relatie stonden als verplegende-patiënt, zo minitueus mogelijk te be-schrijven en om deze observatieperiode te beperken tot 10 minuten. De observatieperioden werden op verschillende momenten van de dag

gekozen. De studenten, die eerst hun steun zochten in handboeken en met termen als 'mutistisch' e.d. gingen schermen, werden in onze wekelijkse besprekingen in dit opzicht sterk ontmoedigd. Ook de interpretatieve behoefte kon worden tegengegaan, d.w.z. beschrijvingen als: 'een patiënt zat agressief met zijn benen te slingeren' werden geweerd. Dat kon ook gemakkelijk omdat de meeste interpretaties over eenzelfde patiënt elkaar tegenspraken. Zo had een ander in het bovengenoemde geval beschreven, dat het meer een lusteloos slingeren met het been was. Aan de andere kant werden de studenten aangemoedigd om hun eigen belevingen te beschrijven. Zo konden zij het bijv. wel ééns worden over het feit, dat zij graag met een bepaalde patiënt een praatje gingen maken. Ik zal u nog een voorbeeld geven van een beschrijving van een van de observatoren waarin duidelijk naar voren komt hoe de studenten, die minder 'gewend' waren aan de gebeurtenissen en gedragingen in een psychiatrisch ziekenhuis, een frisse kijk konden geven op wat zij meemaakten.

'Waarom ik dit zo uitvoerig beschrijf is eigenlijk omdat ik vermoed, dat de heer X zich slechts op een bepaald niveau kan handhaven en dat dit in de 2e PUK zó vanzelfsprekend is, dat het mij nauwelijks meer opvalt. Wanneer er bijv. een grapje gemaakt wordt aan zijn adres verliest hij de grond onder de voeten en kan hij de situatie niet aan. Het valt bijna niet meer op en hij redt zich meestal door over iets anders te beginnen zoals zojuist werd beschreven of door schouderophalend weg te lopen. Daarbij is onze houding in de 2e PUK niet die welke je t.o.v. vrienden hebt tegenover wie je zeker geen genoeg zou nemen met een dergelijke oplossing. Hier doe je dat wel en het valt je nauwelijks op dat je het doet'.

Uit dergelijke beschrijvingen uit een periode van ruim een half jaar heb ik tenslotte 300 'observabele eenheden' gevonden, die de basis vormden van de schalen. Het was dus niet alleen een 'observabele eenheid' wanneer alle observatoren het erover eens waren, dat zij het gezien hadden, maar ook wanneer zij het erover eens waren, dat zij dit voelden t.o.v. een bepaalde patiënt. Bijv. wanneer zij allen meenden, dat zij de afdeling best even konden verlaten indien een bepaalde patiënt aanwezig was die de verantwoordelijkheid in dat geval dan even overnam, werd dit een 'observabele eenheid' geacht. De 'observabele eenheden' zijn door mij geordend en gerangschikt. Bij de groepering heb ik me laten leiden door het overigens arbitraire gevoel, dat bepaalde eenheden bij elkaar hoorden. Zo ontstond een groep van observabele eenheden die betrekking hadden op het communicatieve gedrag, een andere groep die betrekking had op het handelende gedrag e.d.

Bij de ordening van observabele eenheden in een groep of subgroep, werd getracht een rangorde aan te brengen van meer- naar minder sociaal aangepast. Het is evident dat deze rangorde subjectief

is, en opgesteld werd volgens een impliciet gebleven gedachtengang over sociale aanpassing. Dat wil zeggen: de ideeën over wat sociaal aangepast is of zou moeten zijn, worden slechts expliciet gemaakt in de schalen en in de aangebrachte rangorde. Daarna werden de schalen door de verplegenden geprobeerd. De verplegenden werden verzocht en aangemoedigd om hun kritiek te uiten en om aan te geven wanneer gedragsomschrijvingen niet duidelijk waren, niet toepasselijk of op een andere wijze tot misverstanden aanleiding gaven. Tot mijn vreugde hebben de verplegenden hieraan intensief meegewerkt. Talrijke wijzigingen werden aangebracht. Een enkele schaal werd, omdat deze onbruikbaar bleek, bijv. in tweeën gesplitst. Een voorlopig betrouwbaarheidsonderzoek waarbij de onderlinge overeenkomst tussen meerdere beoordelaars statistisch werd berekend, was zeer bemoedigend. Sommige schalen, die laag betrouwbaar bleken, werden op grond hiervan veranderd. Tenslotte had ik 21 schalen, die alle in de eigen kliniek hoog betrouwbaar bleken en zeer goed bruikbaar voor het doel waarvoor ze werden ontworpen. (Voor statistici onder u: de correlatiecoëfficiënten bevonden zich steeds rond de .90) De vraag bleef echter bestaan of nu niet schalen waren gemaakt, die weliswaar goed in de eigen kliniek gebruikt konden worden, maar die bij nader inzien misschien zó waren gesteld in de taal van de eigen kliniek, dat ze in een ander psychiatrisch ziekenhuis onbruikbaar waren of althans onbetrouwbaar zouden zijn. Er werd dus een derde onderzoek ingesteld naar de statistische betrouwbaarheid in een ander psychiatrisch ziekenhuis met meerdere en andere, in dit geval niet 'geschoolde' beoordelaars. Met name 1e, 2e en 3e-jaars leerlingen verplegenden. Ook hier waren de resultaten bijzonder gunstig. Na het verschijnen van mijn proefschrift in maart 1966 heeft de psycholoog A. W. Burger in 'Groot Graffel' het onderzoek nog eens herhaald, waarover hij rapporteerde in het Ned. Tijdschrift voor de Psychologie en haar Grensgebieden. Wederom waren er ten aanzien van de betrouwbaarheid soortgelijke gunstige resultaten. Men kan dus zeggen, dat wij erin zijn geslaagd een betrouwbaar beoordelingsinstrument te vormen, een instrument dat betrouwbaarder is dan de meeste tot dusverre in het buitenland gevormde en gebruikte zgn. 'rating-scales'.

Er is een ander punt waar ik even op in wil gaan. Het is de vraag naar de validiteit van de schalen. Uit het feit dat de PUK-schalen een in hoge mate betrouwbaar instrument zijn, mag in ieder geval worden afgeleid dat met dit instrument 'iets' betrouwbaar wordt gemeten. De vraag naar 'wat' er nu wordt vastgelegd kan langs twee wegen beantwoord worden.

In de eerste plaats is er de praktische weg, die meer handelt over de bruikbaarheid van de schalen voor de psychiater. De psychiater kan de uitslag van de beoordeling op de antwoord-formulieren door

een administratieve kracht laten noteren op de jaar kaarten (die zich in de status van de patiënt bevinden*).

Hij kan zonodig steeds op eenvoudige wijze terugvinden wat de gegeven beoordeling betekent; welke omschrijving van toepassing is of was op de beoordeelde patiënt. Dit is mogelijk omdat de inhoud, de betekenis, die door hem gegeven zal worden aan de gedragsomschrijving als gevolg van de hoge betrouwbaarheid, dezelfde zal zijn als die de beoordelaar eraan gaf. In de praktijk zal de behandelend arts dus een getrouwe beschrijving kunnen vinden van een aantal gedragsaspecten van de patiënt die hij behandelt. Hij zal de voor- en achteruitgang kunnen volgen en hij zal kunnen merken welke gebieden, bijv. communicatie, initiatief, groepsinteracties, of zelfverzorging, bij deze voor- of achteruitgang het meest betrokken zijn.

Verder geeft de totaalscore, dus de summatie van de getallen waarmee de gedragsomschrijvingen worden aangeduid, enig beeld van de voor- of achteruitgang in het algemeen. Ik zeg met nadruk enig beeld omdat nog niet bewezen werd en waarschijnlijk ook niet bewezen zal worden, dat de getallen in hun numerieke waarde mogen worden opgevat. Het verschil tussen 3 en 5 is niet echt 2; 2×2 is niet echt 4; en het mag niet zonder meer worden aangenomen, dat het verschil tussen 4 en 5 hetzelfde is als dat tussen 5 en 6 of tussen 90 en 91. Toch geeft de totaalscore een 'aanwijzing' voor het beloop. Om een paar voorbeelden te noemen: een patiënte kreeg tijdens het verblijf bij ons een terugval. De patiënte, die reeds vóór deze terugval bij de maandelijkse beoordelingen met de PUK-schalen een lagere totaalscore behaalde, zakte op het dieptepunt van de 'Schub' met niet minder dan 40 punten. (U moet dan nog weten dat de maximum totaalscore 110 bedraagt en de minimum 21).

Bij andere patiënten zagen wij, dat de diverse enthousiaste 'para'-therapeuten een lichte verbetering constateerden, terwijl de PUK-schalen geen enkele vooruitgang noteerden. Na verloop van tijd, en ik spreek hier van een of meer jaren, gaven dezelfde maar nu teleurgestelde therapeuten een terugval aan terwijl de PUK-schalen nog steeds dezelfde score aangaven. Omgekeerd zagen wij op de PUK-schalen bij sommige (schizofrene) patiënten duidelijk periodische, cyclische voor- en achteruitgang in de totaalscore, terwijl daarvan uit de andere informatiebronnen niets was gebleken en terwijl ook de behandelend psychiater dit nooit duidelijk was gebleken.

Soms ook kon een voor- of achteruitgang op de PUK-schalen gezien worden, die iedereen was ontgaan, maar die achteraf verklaard kon worden uit een geslaagde vakantie, de dood van een familielid of de wijziging in de psychofarmacologische medicatie. We hebben zelfs één maal een wijziging in de psychofarmacologische me-

* Een met de PUK-schalen vertrouwde verplegende (of ook leerling-verplegende zoals bleek) blijkt ongeveer 15 minuten nodig te hebben voor het invullen van het antwoordformulier betreffende één patiënt.

dicatie, die door een misverstand was ontstaan, langs deze weg kunnen terugvinden.

Intussen is met al deze gegevens uit de praktijk strikt genomen nog niets *bewezen*. Bijvoorbeeld de vraag of de PUK-schalen nu juist het beloop van de ziekte of het verloop van de resocialisatie aangeven blijft onopgelost, zoals nog tal van andere vragen. Er is een exploratief correlatie-onderzoek gaande ten aanzien van een groot aantal biochemische variabelen, dat misschien enig inzicht zal verschaffen. Daarnaast loopt er een exploratief correlatie-onderzoek met een beoordelingsinstrument, dat werd ontwikkeld voor de longitudinale beoordeling in groepstherapieën.

Zo zijn er meer onderzoeken waarin gepoogd wordt om systematische samenhangen te vinden tussen de scores op de PUK-schalen en andere parameters. Uit een aantal van deze onderzoeken zullen gegevens geput worden, die de basis kunnen vormen voor hypothesen ter verificatie. Op deze wijze zal de begripsvaliditeit van de PUK-schalen nauwer omschreven kunnen worden en zullen de PUK-schalen kunnen dienen om de begripsvaliditeit van andere meetmethoden nauwer te omschrijven.

Voorlopig is er nog maar één gegeven met betrekking tot de validiteit van de PUK-schalen. Ik heb u gesproken over het onderzoek in het andere psychiatrische ziekenhuis, dat vooral werd gedaan met het oog op de bepaling van de betrouwbaarheid van de schalen, de onderlinge overeenstemming tussen meerdere beoordelaars dus, ditmaal buiten de eigen kliniek. Dat onderzoek vond plaats op een afdeling waar, door allerlei oorzaken, de therapeutische mogelijkheden zeer beperkt waren. Een van de belangrijkste oorzaken was in ieder geval gebrek aan geld. Te verwachten was dus dat, gezien over de hele populatie, het verschil tussen deze patiënten en de patiënten van de Rümke Kliniek groot zou zijn en wel in die zin, dat de patiënten van de Rümke Kliniek in het algemeen hogere scores zouden behalen op de PUK-schalen dan de patiënten uit het andere ziekenhuis. Deze verwachting kon worden bevestigd met behulp van de mediaantoets. De verschillen waren significant. ($P < 0.001$)

Tenslotte, ik hoop dat ik u een indruk kon geven van hoe wij werken in de praktijk. Hoe in een team, samengesteld uit deskundigen van diverse pluimage, oplossingen kunnen worden gezocht en voor zover het deelproblemen betreft ook gevonden, voor de vragen die steeds bij ons opkomen tijdens het dagelijks werk.

SUMMARY

The aim of the present study was the construction of a sensitive objective instrument for the evaluation of changes in social adjustment of chronic psychotic patients.

The patients were subjected to an intensive multidimensional therapeutic programme.

In assessing social adjustment emphasis is laid on interpersonal behavioural patterns, rather than on the various external, and often arbitrarily chosen 'qualities' that are considered in the usual symptom rating scales.

The instrument had to be sensitive enough to register small differences in the interaction between judge and subject with a reasonably high degree of reliability (interjudge agreement). The (un)reliability of the psychiatric judgements is discussed. Four aspects of the psychiatric judgement receive special attention:

- 1 The expertness of the judge with regard to the judgement he is required to give. (E.g. it is generally known that in certain respects the expertness of the nurses in judging social adjustment is greater than that of the psychiatrists).
- 2 The fact that the judgement on adjustment is of value if formed in the particular social situation on which it bears.
- 3 The fact that in the course of forming a judgement the situation conditioning the judgement changes.
- 4 The distinctive power of a given rating instrument. The descriptions used should be unambiguous and so chosen as to ensure the greatest possible interjudge agreement, while retaining maximum significance for the purpose they serve (i.e. judging social adjustment).

The validity of judgements in psychiatry is discussed.

The existence of any measurable and valid external criterion is denied, considering that the decisions of the psychiatrist in the mental hospital are inevitably influenced and conditioned by a host of still intangible factors.

It is one of the primary functions of a rating instrument, to help make the factors determining the psychiatrists' decisions explicit.

Apart from this practical value of an instrument, reliable in terms of interjudge agreement, it is important, from a theoretical point of view, to assess its construct validity.

The construction of the PUK rating scales is discussed. An exposition is given of the specially devised method by which the building-stones of the rating scales were obtained. Six undergraduate students, working as parttime nurse aids in the above mentioned hospital were instructed to write down in detail their remarks (as participants) on the interaction patterns immediately following ten minutes' observation periods, which each individual student had with every patient over a period of approximately one year. They were told that neither symptoms nor any other characteristic qualities of the patients were of significance of that they had to confine their attention to the interaction patterns on the ward, while keeping in view social adjustment and social acceptability.

Those observations which recurred in a majority of the individual reports on one and the same patient, were assumed to constitute reliable judgements. It were those judgements which formed the basis for the construction of our rating scales.

Twenty-one scales consisting of 5, 6 or 7 ordinally arranged descriptions were subsequently tested for their interrater reliability.

Both in the Second Psychiatric Hospital of Utrecht University as well as in another psychiatric hospital (Zon en Schild, Amersfoort), the results proved to be highly satisfactory (each of the 21 scales used by 4—8 raters, in three different trials, showed a reliability of about .90). As a preliminary indication of the validity of the PUK rating scales it may be mentioned that the median test showed a highly significant differentiation of two populations in each scale.

($P < 0.001$). Investigations on the construct validity are continued.