



www.DirectieveTherapie.nl

Dank voor uw download

U kunt er natuurlijk uit citeren, graag zelfs, maar dan wel met bronvermelding. U mag dit artikel ook ruimhartig verspreiden mits het niet voor commerciële doeleinden is. In die gevallen pas na onze schriftelijke toestemming.

Opname in bloemlezingen en readers moedigen wij aan, maar wel graag eerst even overleggen.

Alle rechten van de artikelen liggen bij
de Stichting Cognitie en Psychose.

Voor alle vragen:
info@gedachtenuitpluizen.nl

LIESBETH BOEZEN-HILBERDINK, PAULINE JANSE, MAARTEN VAN DIJK
 & MARC VERBRAAK

De Outcome Rating Scale en de Session Rating Scale

De psychometrische eigenschappen van de Nederlandse versies

Er is inmiddels zoveel onderzoek verricht in de geestelijke gezondheidszorg dat we kunnen beschikken over zogenoemde *evidence-based* behandelprotocollen: specifieke interventies bij specifieke psychische aandoeningen, waarvan de effectiviteit in wetenschappelijk onderzoek is aangetoond (zie bijvoorbeeld Keijsers, van Minnen & Hoogduin, 2011). Toch wordt er in de praktijk nog heel vaak geen gebruik gemaakt van deze methoden (Becker, Zayfert & Anderson, 2003; Stobie, Taylor, Quigley, Ewing & Salkovskis, 2007; van Dijk, Verbraak, Oosterbaan & van Balkom, 2012). De inzet van dergelijke bewezen effectieve stoornisspecifieke behandelingen wordt namelijk als onvoldoende gezien voor het behalen van het optimale behandelresultaat in de ambulante praktijk van de reguliere ggz. Er zouden grote groepen patiënten zijn die niet van dergelijke bewezen effectieve protocollaire behandelingen profiteren. Onverminderd populair is sinds jaar en dag het argument dat de patiëntenpopulatie waarmee in de praktijk wordt gewerkt heterogener, complexer en zieker is dan de populatie waarover onderzoeksbevindingen worden gerapporteerd (zie bijvoorbeeld Havik & VandenBos, 1996), ondanks aanwijzingen voor het tegendeel (bijvoorbeeld Stirman, DeRubeis, Crits-Christoph & Brody, 2003; Stirman, DeRubeis, Crits-Christoph & Rothman, 2005). Het meten van de voortgang van de behandeling als ‘vinger aan de pols’ van de individuele patiënt (Routine Outcome Monitoring) zou hier wel eens een uitweg uit deze controverse kunnen bieden (Verbraak, 2012). Bij Routine Outcome Monitoring gaat het er immers om dat therapeuten gevalideerde meetinstrumenten routinematig gebruiken om het profijt van de ingezette behandeling van de individuele patiënt vast te stellen, op basis waarvan de behandeling eventueel (bij-) gestuurd kan worden.

Naast een onderzoekslijn naar *evidence-based* behandelmethoden, bestaat er een onderzoekslijn die zich nadrukkelijk heeft gericht op het ontwikkelen van *evidence-based psychotherapy relationships* (zie Lambert & Barley, 2002). Binnen deze lijn ligt de

nadruk op het meten van de kwaliteit van de therapeutische relatie en het ontwikkelen van manieren om deze in therapiecontacten te optimaliseren, met als doel het behandelresultaat te verbeteren. Gericht onderzoek laat zien dat een goede kwaliteit van de therapeutische relatie de uitkomst van een behandeling positief beïnvloedt (Horvath & Bedi, 2002; Horvath & Symonds, 1991; Martin, Garske & Davis, 2000). Het belang van deze therapeutische relatie staat veel minder ter discussie dan dat van de bewezen effectieve interventies.

Los van deze discussie laat onder andere onderzoek uitgevoerd door Hansen, Lambert en Forman (2002) onder zesduizend patiënten zien dat er meer dan voldoende reden is om te blijven kijken naar manieren om de kwaliteit en resultaten van behandeling in de dagelijkse praktijk te optimaliseren, hetzij door de inzet van bewezen effectieve vormen van behandeling, hetzij door het optimaliseren van de therapeutische relatie. Zij vonden dat (gemeten met gevalideerde uitkomstinstrumenten) slechts 35 procent van de patiënten uit deze steekproef echt opknapt na behandeling in de ggz; 57 procent van de patiënten liet geen tot nauwelijks verandering zien en 8 procent verslechterde gedurende de therapie.

Onderzoek laat ook zien dat therapeuten vaak op basis van hun eigen inschatting (klinische blik) niet goed kunnen inschatten of er sprake is van verslechtering bij een patiënt en welke patiënt risico loopt voortijdig te stoppen met therapie. Zo vonden Hannan et al. (2005) dat van een groep van 550 patiënten maar bij 3 van hen door de therapeuten een negatieve uitkomst voorspeld werd, terwijl het daadwerkelijke aantal patiënten met een negatieve uitkomst uiteindelijk 40 bedroeg. Een feedbackalgoritme daarentegen voorspelde in 77% van deze gevallen de negatieve uitkomst. Verder blijkt ook dat de kwaliteit van de therapeutische relatie binnen een behandeling nogal eens verschillend wordt ingeschat door de therapeut en de patiënt. Er is lang niet altijd sprake van overeenstemming (Hafkenscheid, Duncan & Miller, 2010; Horvath & Bedi, 2002).

Om de kwaliteit van individuele behandelingen te kunnen controleren en optimaliseren lijkt gestructureerd meten een belangrijke voorwaarde. Zo raadt de taakgroep van de American Psychological Association, gericht op evidence-based psychotherapy relationships, dan ook aan om patiëntfeedback standaard mee te nemen tijdens behandelingen om zo de uitkomsten van therapie te optimaliseren (Norcross & Wampold, 2011). Verschillende onderzoeken hebben inmiddels uitgewezen dat, waar in behandelingen gebruik wordt gemaakt van regelmatige patiëntfeedback, het resultaat van de therapie vergroot en de kans op drop-out verkleind wordt (Harmon et al., 2007).

Er zijn in de loop der jaren verschillende vragenlijsten ontwikkeld gericht op het meetbaar maken van de behandeluitkomst, zoals de Outcome Questionnaire 45 van Lambert et al. (1996). Ook zijn er vragenlijsten ontwikkeld gericht op het meetbaar maken van de alliantie (zie hierover het naschrift bij dit artikel), zoals de Helping Alliance Questionnaire (HAQ) van Luborsky et al. (1999). Het invullen van deze lijsten kost echter vaak enige tijd, waardoor ze in de klinische praktijk, ook al gaat het maar om 10 tot 15 minuten, minder praktisch routinematig en frequent bruikbaar zijn. Om dit probleem te verhelpen, ontwikkelden Miller en Duncan in 2003 twee zeer korte vragenlijsten om verschillende aspecten van *outcome* en alliantie te meten: de Outcome Rating Scale (ORS; Miller, Duncan, Brown, Sparks & Cloud, 2003) en de Session Rating Scale (SRS; Duncan et al., 2003).

Een aantal studies binnen specifieke behandelsettings en patiëntgroepen laat inmiddels zien dat het gebruik van deze lijsten de effectiviteit van de behandeling vergroot en deze mogelijk efficiënter maakt (Miller, Duncan, Brown, Sorell & Chalk, 2006; Reese, Norswothy & Rowlands, 2009).

In 2002 zijn de ORS en SRS vertaald naar het Nederlands (Miller, Duncan & Johnson, 2002). Sindsdien wordt er in toenemende mate in Nederland gewerkt met dit patiëntfeedbacksysteem (Baert, 2012; Beljouw & Verhaak, 2010). Er is echter nog maar weinig onderzoek verricht naar de Nederlandse versies van de ORS en SRS (met uitzondering van Hafkenscheid, 2010 en Hafkenscheid, Duncan & Miller, 2010). Dit maakt dat het onduidelijk is of de conclusies van Miller et al. (2006) over de toepasbaarheid en werkzaamheid zonder meer generaliseerbaar zijn naar andere behandelsettings en patiëntgroepen. Het onderzoek waarvan we in dit artikel verslag doen, heeft als doel om te bekijken of de psychometrische kwaliteiten van de Nederlandse versies van deze lijsten voldoende zijn. Daarnaast wordt onderzocht wat de representatieve *cut-off*-waarden en normgegevens voor een Nederlandse ambulante patiëntenpopulatie zijn.

De betrouwbaarheid en validiteit van de ORS zullen onderzocht worden door deze ultrakorte lijst te vergelijken met de langere SCL-90 en OQ-45. Voorts zullen de *cut-off*-score (grens tussen ziek en gezond) en mate van betrouwbare verandering worden berekend. De kwaliteiten van de SRS zullen worden onderzocht door deze lijst af te zetten tegen de Werk Alliantie Vragenlijst-12. Ook zal gekeken worden naar de voorspellende waarde van de therapeutische relatie zoals gemeten met de SRS op de uitkomst van de behandelingen.

METHODE

¶ Setting

Het onderzoek werd uitgevoerd binnen de HSK Groep, een landelijke organisatie die zich richt op het ambulant onderzoeken en behandelen van patiënten met psychische stoornissen voor wie psychologische behandeling meestal de, of één van de, eerste keuze behandelinterventies is. De gehanteerde behandelmethoden zijn evidence-based. Er wordt gewerkt met klachtspecifieke psychologische behandelprotocollen op basis van cognitieve gedragstherapie.

¶ Deelnemers

In totaal hebben 587 patiënten die zich tussen 1 september 2009 en 31 december 2010 bij zes locaties van de HSK Groep aanmeldden aan het onderzoek deelgenomen (zie tabel 1). Patiënten werden verwezen door huisarts of bedrijfsarts.

TABEL 1
Demografische kenmerken deelnemers

	N	%
Geslacht		
Man	281	47.9
Vrouw	306	52.1
Opleiding		
Lagere school	7	1.3
Beroepsonderwijs	360	69.6
Middelbaar onderwijs	69	13.3
Hbo/universitair	89	15.8
Burgerlijke staat		
Getrouwd/samenwonend	364	67.4
Ongetrouwd	137	25.4
Gescheiden	33	6.1
Weduwe/weduwnaar	6	1.1

Diagnose	N	%
Aanpassingsstoornis	164	28.0
Werkgerelateerde problematiek/burn-out	163	27.8
Stemmingsstoornissen	122	20.9
Angststoornis	102	17.3
Anders	13	5.6

¶ Instrumenten

De Outcome Rating Scale (Miller et al., 2003) is een korte zelfrapportage vragenlijst (visueel analoge schaal) die uit vier items bestaat. Er worden vier gebieden van functioneren gemeten, namelijk individueel functioneren, interpersoonlijke relaties, de mate waarin de patiënt zijn sociale rol kan vervullen en het algemeen welbevinden. Ieder item bestaat uit een lijn van 10 centimeter, waarbij een score uiterst links op de lijn aangeeft dat het zeer slecht gaat, en uiterst rechts dat het zeer goed gaat. De patiënt zet per schaal een streepje op de lijn. De lengte van het begin van de lijn tot waar het streepje is gezet, is de score op de desbetreffende schaal. De vier scores worden bij elkaar opgeteld en vormen de totaalscore (range 0-40). Hoe hoger de score, des te beter de outcome. De betrouwbaarheid (Cronbachs alfa) van de Engelstalige versie is goed (.87). De test-hertestbetrouwbaarheid is eveneens goed (.80). De overall validiteit van de Engelstalige versie van de ORS (wanneer afgezet tegen de OQ-45) is adequaat (.59). De validiteit (Pearsons product-momentcorrelatie) van de subschalen is voldoende gebleken in verschillende onderzoeken (tussen .56 tot .69; Miller et al., 2003). Er is een matig sterke samenhang tussen de subschalen van de ORS en OQ-45 (.57 tot .69), ondanks dat de eerste is afgeleid van de laatste (Bringhurst, Watson, Miller & Duncan, 2006; Miller et al., 2003), wat conform verwachting is als wordt gewerkt met ultrakorte vragenlijsten.

De Session Rating Scale (Duncan et al., 2003) is eveneens een korte zelfrapportage vragenlijst. De schaal bestaat uit vier elementen, namelijk kwaliteit van de therapeutische relatie, mate van overeenstemming in doelen tussen patiënt en therapeut, mate van overeenstemming in aanpak en werkwijze en een algehele beoordeling van de sessie. De patiënt zet per item een streepje op een lijn van 10 centimeter, waarbij uiterst links betekent dat de patiënt niet tevreden was over dat onderdeel in de sessie en uiterst rechts betekent dat de patiënt zeer tevreden was. De lengte van het begin van de lijn tot waar het streepje is gezet, is de score op de desbetreffende

schaal. Ook hier worden de vier scores bij elkaar opgeteld en vormen zo de totaalscore (range 0-40), waarbij geldt dat hoe hoger de score is, des te beter de kwaliteit van de alliantie is. De betrouwbaarheid van de Engelstalige SRS is berekend en vergeleken met die van de Helping Alliance Questionnaire (HAQ-II). De interne consistentie van de SRS bleek vergelijkbaar goed met die van de HAQ-II (respectievelijk .88 en .90; Duncan et al., 2003). De test-hertestbetrouwbaarheid voor de SRS is voldoende (.64). Concurrente validiteit werd berekend door de product-momentcoëfficiënt te berekenen tussen de SRS totaalscore en de HAQ-II totaalscore. Deze was voor deze korte vragenlijst voldoende (.48; Duncan et al., 2003).

¶ Instrumenten voor het onderzoeken van de psychometrische kwaliteiten van de ORS

De Outcome Questionnaire 45 (OQ-45; Lambert et al., 1996), bestaat uit 45 items van het Likert-type en kost ongeveer 15 minuten om in te vullen. De respondent wordt verzocht aan te geven hoe vaak beschrijvingen op hem van toepassing zijn in de afgelopen week, inclusief de dag van invullen. Er zijn vijf antwoordmogelijkheden (variërend van 0 = nooit tot 4 = bijna altijd). Naast de totaalscore omvat het instrument drie subschalen, namelijk ernst van de symptomen, interpersoonlijk functioneren en maatschappelijk functioneren. De interne consistentie van de OQ-45 is goed (.93) evenals de test-hertestbetrouwbaarheid (.84). De concurrente validiteit (de OQ-45 afgezet tegen de SCL-90) was significant op het 0.01 level (range 0.50-0.85 uit de verschillende onderzoeken; Lambert et al., 1996; Mueller, Lambert & Burlingame, 1998; Umphress, Lambert, Smart & Barlow, 1997). De interne consistentie van de Nederlandstalige versie van de OQ-45 blijkt goed te zijn voor de totaalscore (tussen .91 en de .93), maar onvoldoende voor de subschaal die vraagt naar het maatschappelijk functioneren (alfa tussen .53 tot .69; de Jong et al., 2007). De test-hertestbetrouwbaarheid is goed (variërend van .76 tot .79) en de constructvaliditeit van de OQ-45 in de studie van de Jong et al. (2007) was goed in een klinische groep ($r = .72-.80$) en voldoende in de groep van studenten ($r = .42-.78$).

De Symptom Check List-90 (SCL-90; Derogatis, 1977; Nederlandse vertaling Arrindell & Ettema, 1986) is een multidimensionele klachtenlijst die vaak wordt gebruikt als meetinstrument voor de ervaren mate van psychopathologie (als toestandsbeeld) en voor de evaluatie van behandelingen (Arrindell & Ettema, 1986, 2003). De SCL-90 meet de mate waarin iemand gedurende de afgelopen week last heeft gehad van bepaalde psychische en lichamelijke klachten. De klachtenlijst is gebaseerd op de zelfbeoordeling van de patiënt. De SCL-90 bestaat uit negentig items die worden beantwoord op een vijfpuntsschaal (1 = helemaal niet, 2 = een beetje, 3 = nogal, 4 = tame-

lijk veel, 5 = heel erg). De totaalscore geeft de algemene mate voor het welbevinden aan. Uit Nederlands onderzoek van Arrindell en Ettema (2003) blijkt dat de interne consistentie van de subschalen grotendeels goed is ($> .80$). Ook de betrouwbaarheid is goed (.82).

De Assessment of DSM-IV Personality Disorders (ADP-IV; De Doncker, Schotte, Vertommen & Vankerckhoven, 1997) zal gebruikt worden om de divergente validiteit te bepalen. De verwachting is dat, omdat deze vragenlijst een geheel ander construct meet dan de ORS, namelijk persoonlijkheidsproblematiek, er geen samenhang tussen de scores op de ADP-IV en de ORS zullen zijn. De ADP-IV is een zelfrapportage vragenlijst en bestaat uit 94 items. Elk item wordt gescoord op een zevenpuntsschaal (range 1 = helemaal niet tot 7 = helemaal wel) en is gericht op het onderzoeken van persoonlijkheidspathologie volgens de DSM-IV. Bij een score 5, 6 of 7 dient de patiënt tevens aan te geven in welke mate hij last heeft van een bepaald kenmerk (bijvoorbeeld angst om in de steek gelaten te worden of de angst dat de ander misbruik van je zal maken), door de ernst hiervan aan te geven met een 1, 2 of 3. De interne consistentie van de ADP-IV is goed (.79), de test-hertestbetrouwbaarheid is eveneens goed (.79).

¶ Instrumenten voor het onderzoeken van de psychometrische kwaliteiten van de SRS

De Werk Alliantie Vragenlijst 12 is afgeleid van de WAV-36, de Nederlandstalige vertaling van de Working Alliance Inventory (WAI; Horvath & Greenberg, 1989; Nederlandstalige versie WAV: Vervaecke & Vertommen, 1996). De WAV meet de band tussen patiënt en therapeut, overeenstemming over de doelen van de behandeling en overeenstemming over de taken in de therapie. Uit onderzoek blijkt dat de interne consistentiecoëfficiënten van de drie schalen goed is (tussen .82 en .85).

PROCEDURE

Alle volwassen patiënten die zich aanmeldden voor behandeling in de periode van september 2009 tot en met december 2010 werd bij intake gevraagd deel te nemen aan het hier beschreven onderzoek. Zij werden schriftelijk en mondeling geïnformeerd en vulden een toestemmingsformulier (*informed consent*) in. Om naast de onderzoekspopulatie de uitkomsten te kunnen vergelijken met een normale populatie, werden de partners van de deelnemers eveneens benaderd om deel te nemen als niet-klinische controlegroep.

Deze partners vulden ter toelating aan het onderzoek de SCL-90 in, zodat bepaald kon worden of de controlegroep vergelijkbaar was

met de normale populatie zoals bekend uit de normen van de SCL-90. In het onderzoek vulden zij alleen de OQ-45 en ORS in. Er waren 166 deelnemende partners. Van deze steekproef zijn uiteindelijk de scores van vijftig partners buiten beschouwing gelaten, omdat zij zelf onder psychologische behandeling bleken te staan of geen complete scores gaven.

Bij intake vulden patiënten de SCL-90, ADP-IV, OQ-45 en ORS in. Bij aanvang van de behandeling vulden ze voorafgaand aan de sessie de ORS in en na afloop van de eerste sessie de WAV-12 en de SRS. Bij alle sessies die daarop volgden vulden patiënten de ORS van tevoren en de SRS na afloop in. Om de vijf sessies en bij ontslag vulden patiënten tevens de WAV-12 vooraf en de OQ-45 nadien in.

¶ Data-analyse

De interne consistentie van de ORS en SRS werd onderzocht door berekening van Cronbachs alfa. De test-hertestbetrouwbaarheid van de ORS werd berekend met Pearsons product-momentcorrelatie. Die van de SRS werd, in verband met een scheve verdeling van de scores, gemeten met Spearman's correlatiecoëfficiënt. Omdat het herhaald meten met deze instrumenten de mogelijkheid bood, werden de interne consistentie en de test-hertestbetrouwbaarheid meerdere malen berekend. Om de concurrente validiteit te onderzoeken van de ORS werd naar de correlatie tussen de ORS en de OQ-45 gekeken door middel van het berekenen van Pearsons product-momentcorrelatie. De ORS-scores werden eveneens afgezet tegen de SCL-90-scores bij intake. Om te bepalen wanneer er sprake is van klinisch significante verbetering, moet aan twee criteria worden voldaan (Jacobson, Follette & Revenstorf, 1986): statistisch betrouwbare verandering (*reliable change*) en het bereiken van een score boven de cut-off-waarde. Als alleen aan het eerste criterium wordt voldaan, dan is er wel sprake van verbetering, maar nog niet van herstel of terugval. Het bereiken van een score boven de grenswaarde markeert het verschil tussen ziek en gezond. De cut-off-score werd bepaald middels de formule C van Jacobson en Truax (zie het naschrift bij dit artikel; 1991), de RCI door het verschil tussen voor- en nameting te delen door de variantie van de standaardmeetfout.

De concurrente validiteit van de SRS werd berekend door correlaties met de WAV-12 te berekenen. De predictieve validiteit van de SRS werd berekend middels lineaire regressieanalyse, waarbij het verschil tussen de voor- en nameting van de SCL-90 als uitkomstmaat werden gebruikt.

De divergente validiteit werd berekend door de samenhang tussen totaalscores op de ADP-IV te vergelijken met de totaalscores op de ORS en SRS en de mate van samenhang te bepalen.

Alle analyses werden uitgevoerd met SPSS versie 17 (SPSS, Chicago, Verenigde Staten).

RESULTATEN OUTCOME RATING SCALE (ORS)

¶ Normatieve data

In tabel 2 staan de gemiddelde scores en standaarddeviaties voor de klinische groep bij intake en de controlegroep op de ORS. De totaalscore van de klinische groep op de ORS is lager dan gevonden door Miller et al. ($M = 19.6, SD = 8.7; 2003$). De scores van de controlegroep zijn respectievelijk 29.6 voor de Nederlandse controlegroep en 26 voor de Amerikaanse controlegroep. Er zijn geen significante verschillen tussen mannen en vrouwen gevonden op de scores bij intake ($t(524) = 0.58, p > .05$), OQ-45 ($t(483) = -4.49, p > .05$) en totaalscore op de SCL-90-R ($t(555) = -1.10, p > .05$). De gemiddelde totaalscore op de OQ-45 was 70.5 ($SD = 22.2$), wat indicatief is voor een hoog klachtenniveau. De gemiddelde totaalscore op de SCL-90 bedraagt 180.7 ($SD = 47.3$) en is indicatief voor een hoog klachtenniveau.

De controlegroep heeft een gemiddelde score van III op de SCL-90 ($SD = 21.8$), wat aangeeft dat de deelnemers van de controlegroep een beneden gemiddeld klachtenniveau rapporteren in vergelijking met de normale populatie (normgroep algemeen; Arrindell & Ettema, 2003).

TABEL 2
Gemiddelden en standaarddeviaties van de ORS totaalscores van de klinische en controlegroep

	Controlegroep N = 116		Klinische groep N = 524	
	M	SD	M	SD
ORS Individueel	7.3	1.8	3.6	2.1
ORS Relationeel	7.4	1.7	5.5	2.4
ORS Sociale rol	7.5	1.6	3.9	2.4
ORS Algemeen	7.5	1.6	4.0	2.0
ORS Totaal	29.6	6.0	17.0	7.2
Cut-off-score RCI			24 9	

¶ Betrouwbaarheid

De interne consistentie van de ORS in de klinische groep en de controlegroep is berekend middels Cronbachs alfa (zie tabel 3). Voor de controlegroep is de alfa .94. Dit is vergelijkbaar met de studies van Miller et al. (2003) en Hafkensheid (2010). De interne consistentie (r) van de ORS voor de klinische groep ligt tussen .82 en .96.

TABEL 3
Interne consistentie van de ORS van de klinische groep

	N	in- take	N	Sessie 1	N	Sessie 3	N	Sessie 5
ORS	524	.82	349	.89	361	.94	351	.96

De test-hertestbetrouwbaarheid van de ORS is berekend met Pearsons product-momentcorrelatie. De correlatie tussen de ORS totaalscores op verschillende meetmomenten is vergelijkbaar met maar enigszins hoger dan wat Miller vond (2003; r tussen .49 en .66) en Hafkensheid (2010; r tussen .16 en .63), namelijk een r tussen .57 en .69. (zie tabel 4).

TABEL 4
Test-herstebetrouwbaarheid van de ORS tussen vijf afnames

	N	Sessie 1-2 r	N	Sessie 2-3 r	N	Sessie 3-4 r	N	Sessie 4-5 r
ORS	323	.64	341	.57	339	.69	334	.63

Alle correlaties zijn significant bij $p < 0.01$

¶ Concurrente validiteit

De concurrente validiteit (de samenhang tussen ORS en OQ-45) is middels de Pearson correlatiecoëfficiënt berekend. De samenhang tussen de ORS en OQ-45 subschalen is matig sterk (r tussen .19 en .58; Cohen, 1988). De gevonden correlatie is vergelijkbaar met de bevindingen van Miller (2003). De samenhang voor de totaalscores is vergelijkbaar met hun bevindingen (r van .62 versus een r tussen .53 en .69; zie tabel 5).

TABEL 5
Samenhang tussen ORS en OQ-45 subschalen en totaalschalen

	OQ-45 SD* (N = 493)	OQ-45 IR** (N = 482)	OQ-45 SR*** (N = 492)	OQ-45 To- taal (N = 455)
ORS Individueel	-.53	-.40	-.30	-.52
ORS Relatieeel	-.36	-.54	-.19	-.45
ORS Sociale Rol	-.46	-.36	-.46	-.50
ORS Algemeen	-.55	-.45	-.34	-.56
ORS Totaal	-.58	-.54	-.40	-.62

Alle correlaties zijn significant bij $p < 0.01$

* OQ-45 SD = Symptomatische Distress

** OQ-45 IR = Interpersoonlijke Relaties

*** OQ-45 SR = Sociale Rol

De concurrente validiteit werd ook berekend aan de hand van de correlaties tussen de ORS en SCL-90 totaal- en subschaalscores bij intake. In de klinische groep varieerden de correlaties van $r = .09$ tot $.56$ ($N = 481$). De sterkste relaties werden gevonden tussen de ORS Algemeen subschaal en de ORS totaalscore en de SCL-90 Depressie subschaal ($r = .54$ en $.56$, respectievelijk), en tussen de ORS totaal-score en de SCL-90 totaalscore ($.50$). In de controlegroep ($N = 111$) waren de correlaties sterker (r variërend tussen $.19$ en $.70$). Ook hier werden de sterkste relaties gevonden tussen de ORS Algemeen subschaal en de SCL-90 Depressie subschaal ($r = .70$) en tussen de ORS en de SCL-90 totaalscores ($.66$).

¶ Divergente validiteit

Om de divergente validiteit te bepalen is de ORS vergeleken met de Nederlandse versie van de Assessment of DSM-IV Personality Disorders (De Doncker et al., 1997). De correlaties liggen tussen de $-.04$ en $.26$. Hoewel enkele correlaties significant zijn ($p < .01$) is de sterkte van de samenhang laag, zoals verwacht.

¶ Gevoeligheid voor verandering

De ORS wordt gebruikt als uitkomstmaat en als instrument om vooruitgang in een therapie te meten. Daarom is het belangrijk dat de ORS gevoelig is voor verandering van klachteniveau. Van de groep respondenten hebben 172 patiënten de lijsten zowel bij intake als ontslag ingevuld. De gemiddelde score bij intake is 16.9. De gemiddelde score bij ontslag is 29.2. Er is een duidelijk verschil tussen voormeting en nameting ($t(171) = -12.3, p < .05, r = .81$).

¶ Cut-off-score en Reliable Change Index

Het punt van de ORS dat aangeeft of iemand in het bereik van de gezonde populatie scoort (cut-off-score), is 24. De gevonden cut-off-score voor de Amerikaanse populatie is 25 (Miller et al., 2003).

De Reliable Change Index (RCI) is de mate waarin de verandering in uitkomst toe te schrijven is aan betrouwbare verandering. Een verandering is betrouwbaar als het verschil tussen de metingen significant is en niet toe te schrijven aan meetfouten van het instrument. De RCI uit onderzoek van Miller et al. (2003) bedraagt 5. De RCI van het huidige onderzoek bedraagt 9, wat aanzienlijk hoger is, maar vergelijkbaar met de gevonden waarde (8) door Hafkenscheid (2010).

RESULTATEN SESSION RATING SCALE (SRS)

¶ Normatieve data

Tabel 6 laat de gemiddelde scores en standaarddeviaties zien voor de SRS totaalscores om de vijf sessies.

TABEL 6
Gemiddelde en standaarddeviaties op de SRS na eerste, vijfde, tiende en vijftiende sessie

	Sessie 1			Sessie 5			Sessie 10			Sessie 15		
	N	M	SD	N	M	SD	N	M	SD	N	M	SD
SRS	349	30.1	6.1	321	32.0	4.7	208	32.6	4.7	121	33.6	4.4

¶ Betrouwbaarheid

De interne consistentie is berekend middels Cronbachs alfa voor de eerste vijf sessies. De interne consistentie blijkt tussen .85 en .95 uit te komen (zie tabel 7).

TABEL 7
Interne consistentie van de SRS

	N	Ses- sie 1	N	Ses- sie 2	N	Ses- sie 3	N	Ses- sie 4	N	Ses- sie 5
SRS	349	.93	356	.85	342	.94	339	.91	321	.95

Omdat de scores op de SRS scheef verdeeld zijn, is een non-parametrische test, Spearman's correlatiecoëfficiënt, gebruikt om betrouwbaarheid te berekenen. De test-hertestbetrouwbaarheid (tabel 8)

voor de gevonden SRS-scores zijn vergelijkbaar met wat Duncan et al. (2003) vonden (een overall r van .64). De correlaties (r) liggen tussen .48 en .72.

TABEL 8
Test-hertestbetrouwbaarheid van de SRS tussen vijf meetmomenten

	N	Sessie 1-2 r	N	Sessie 2-3 r	N	Sessie 3-4 r	N	Sessie 4-5 r
SRS	317	.48	313	.72	315	.61	296	.59

¶ Concurrente validiteit

Om te corrigeren voor de scheve verdeling van de SRS-scores werd gebruikgemaakt van Spearman's rho om de concurrente validiteit te kunnen berekenen. De concurrente validiteit van de SRS is berekend door de samenhang tussen SRS en WAV-12 te bepalen bij aanvang van de behandeling (zie tabel 9). De correlaties zijn matig sterk maar significant ($p < .01$). Een tweede vergelijking tussen SRS en WAV-12-scores is rond de vijfde sessie gemaakt, de correlaties tussen SRS en WAV-12 zijn dan tussen .26 en .34, wat duidt op matig sterke verbanden, die wel significant zijn ($p < .01$). Dit houdt in dat de samenhang tussen WAV-12 en SRS gering lijkt.

TABEL 9
Correlaties tussen de SRS en de WAV-12 subschalen en totaalscores bij begin van de behandeling

	WAV-12 Band (N = 235)	WAV-12 Doel (N = 252)	WAV-12 Taak (N = 248)	WAV-12 Totaal (N = 234)
SRS Relatie	.32	.36	.37	.37
SRS Doel	.38	.41	.40	.43
SRS Aanpak	.31	.41	.46	.43
SRS Algemeen	.37	.40	.45	.44
SRS Totaal	.39	.43	.45	.46

Alle correlaties zijn significant bij $p < .01$ (2-tailed)

¶ Predictieve validiteit

Om de voorspellende waarde van de therapeutische relatie (zoals gemeten met de SRS) op de uitkomst van de behandeling (gedefini-

eerd door het verschil tussen SCL-90-score bij intake en ontslag) te bepalen, werd lineaire regressie toegepast.

De SRS-totaalscore bij sessie 2 en 3 is enigszins voorspellend voor de uiteindelijke uitkomst van de therapie ($p < .05$), waarbij vooral de SRS-score bij sessie 2 voorspellende waarde heeft ($\beta_1 = -.14, p < .05$). De invloed van de SRS blijkt echter zeer beperkt ($R^2 = .02$). Om te onderzoeken of deze resultaten toe te schrijven zijn aan de instrumenten of dat de invloed van de therapeutische relatie gering is, werd dezelfde analyse toegepast, gebruikmakend van de WAV-12. Noch de eerste meting aan het begin van de behandeling, noch de meting bij de vijfde sessie voorspellen de therapie-uitkomst ($p > .05$).

¶ Divergente validiteit

De overlap tussen de SRS en de ADP-IV werd bepaald om de divergente validiteit te bepalen. De correlaties blijken gering en niet significant ($p < .01$). Er is wel een uitzondering: er blijkt een significante (negatieve) samenhang tussen de schaal voor antisociale persoonlijkheidstrekken en enkele schalen van de SRS (Aanpak, Algemeen en totaalscore).

DISCUSSIE

Uitgebreid (psychometrisch) onderzoek naar de Nederlandse versies van de ORS en SRS is nog niet voorhanden. Doelstelling van dit onderzoek was om enerzijds de psychometrische kwaliteiten van deze versies te onderzoeken en anderzijds om normen te verschaffen voor therapeuten, zodat duidelijker wordt hoe de scores op de lijsten geïnterpreteerd kunnen worden.

De vragenlijsten lijken *an sich* bruikbare instrumenten te zijn om de voortgang van de therapie en kwaliteit van de therapeutische relatie te meten. Als uitkomstmaat (om bijvoorbeeld te benchmarken) zijn ze, zoals blijkt uit de hier gepresenteerde validiteitsgegevens, echter psychometrisch te zwak. De betrouwbaarheid van beide lijsten blijkt voldoende tot goed, de validiteit is echter matig voldoende. De correlaties tussen de ORS en SRS en uitgebreidere uitkomstmaten (respectievelijk OQ-45 en SCL-90 en WAV-12) zijn niet zo sterk als we zouden verwachten en liggen lager dan wat Miller et al. (2003) en Duncan et al. (2003) vonden. Voor de ORS geldt dat de totaalscore de sterkste correlatie laat zien met de totaalscore op de OQ-45, terwijl de samenhang met de verschillende subschalen duidelijk minder sterk is. De samenhang met de SCL-90 is zwakker, ook omdat de lijsten deels verschillende constructen meten. Deze scores komen overigens overeen met de bevindingen uit het onder-

zoek van Hafkenscheid (2010). De predictieve validiteit van de SRS is beperkt. De enige voorspeller van de (goede) uitkomst van de therapie blijkt de SRS-score van de tweede en derde sessie, maar ook dit verband blijkt zwak. De divergente validiteit is voor beide lijsten goed.

Dit onderzoek heeft verder tot doel normen voor de ORS en SRS vast te stellen voor de Nederlandse populatie. Miller vindt in zijn onderzoek in 2003 een cut-off-score van 25 punten voor de ORS en een RCI van 5 punten. Uit dit onderzoek komt een cut-off-score van 24 punten, als grens tussen 'ziek' en 'gezond' en een RCI van 9 punten, wat beduidend hoger is. Ook Hafkenscheid vond in een Nederlandse populatie een hogere RCI, namelijk 8 (Hafkenscheid, 2010). Dit laat zien dat de Amerikaanse normen dus niet zonder meer bruikbaar zijn voor Nederlandse patiënten. Op basis van de Amerikaanse normen zou een Nederlandse patiënt bij een positieve verandering in de scores van 5 punten een duidelijke verbetering laten zien. Hierin zou geen aanleiding worden gezien tot het aanpassen van de behandeling. Echter, op basis van de Nederlandse normering zou er wel reden zijn om te kijken naar een aanpassing van de behandeling of het eventueel overgaan tot een therapeutenwissel.

Ook zijn er verschillen in de SRS-scores tussen beide groepen. Duncan et al. (2003) stellen de cut-off van de SRS-score op 36 (als minimumscore voor een goede kwaliteit van de therapeutische alliantie), terwijl in onze onderzoekspopulatie de gemiddelde SRS-score ligt op 34.

Het niet toepassen van de feedback uit zowel ORS als SRS kan gezien worden als een beperking van dit onderzoek, omdat de instrumenten juist zijn ontwikkeld om expliciet mee te werken. Er is echter voor gekozen om de scores juist niet inzichtelijk te maken voor therapeuten, om zo een basismeting te verkrijgen. Uit (lopend) vervolgonderzoek zal blijken wat het effect is van het wel toepassen van de feedback van de patiënten.

Een volgende beperking in dit onderzoek is de relatief kleine controlegroep. Grootschaliger onderzoek met zowel 'normale' controlegroepen als grotere patiëntgroepen is nodig om tot normtabellen te kunnen komen.

Een derde beperking is het hoge aantal uitgevallen metingen. Van de 587 patiënten die aanvankelijk toezegden deel te willen nemen aan het onderzoek, bleven 323 patiënten over die bij de eerste en tweede sessie nog steeds deelnamen aan het onderzoek. Van deze patiënten is er in 121 gevallen ook een eindmeting. Er zijn meerdere factoren die een rol hebben gespeeld hierin. Deze studie is de eerste in een reeks onderzoeken naar het toepassen van formele feedbacksystemen en had als doel een basismeting te zijn. In vervolgonderzoek wordt de feedback wel toegepast. Doordat thera-

peuten geen inzage hadden in of hun patiënten goed deelnamen, kon moeilijk gecontroleerd worden of er voldoende metingen per patiënt waren. Daarnaast vergde het implementeren van het onderzoek het nodige van de therapeuten (het op tijd verstrekken van OQ-45 en WAV-12) en bleek op individueel niveau verschil te bestaan tussen de nauwgezetheid waarmee therapeuten meewerkten aan het onderzoek. Ten slotte was er eveneens sprake van enige dropout, dat wil zeggen, patiënten die voortijdig de behandeling beëindigden zonder bericht.

In (lopend) vervolgonderzoek wordt gekeken naar het effect van het gebruik van de lijsten op de behandelresultaten. Interessant zal zijn of het gebruik ervan de therapie effectiever en korter maakt en of de kwaliteit van de werkrelatie in vervolgonderzoek in het voorstellen van het resultaat een grotere rol inneemt.

Dit artikel is een bewerking van: Janse, P., Boezen-Hilberdink, L., van Dijk, M.K., Verbraak M.J.P.M. & Hutschemaekers, G.J.M. (2013). *Measuring feedback from clients: The psychometric properties of the Dutch Outcome Rating Scale and Session rating Scale*. *European Journal of Psychological Assessment*. Advance online publication. doi: 10.1027/1015-5759/a000172

Een definitie van een 'therapeutische relatie', ook wel 'alliantie' genoemd, werd gegeven door Bordin (1979). 'Alliantie' definieert hij als de relatie tussen patiënt en therapeut en de consensus over doelen en onderwerpen. Het actief bewaken van deze alliantie stelt therapeuten in staat in te grijpen als de alliantie zwak is of dreigt te verslechteren, waardoor voorkomen kan worden dat de patiënt voortijdig stopt met therapie. Bordin stelt dat verstoring in de therapeutische relatie op drie gebieden kan ontstaan: onenigheid over de doelen van de therapie, onenigheid over de taken binnen de therapie en spanning op relationeel niveau.

Formule C van Jacobson en Truax (1991):

M_1 = gemiddelde van de klinische groep bij intake, M_0 = gemiddelde van de controlegroep, S_0 , S_1 zijn de standaarddeviaties van de controle- en de klinische groep.

$$c = \frac{S_0 M_1 + S_1 M_0}{S_0 + S_1}$$

DRS. LIESBETH BOEZEN-HILBERDINK is klinisch psycholoog/psychotherapeut/cognitief gedragstherapeut bij Diaconessenhuis Meppel.

DRS. PAULINE JANSE is onderzoeker en gz-psycholoog/cognitief gedragstherapeut VGCT bij de HSK Groep in Utrecht.

DRS. MAARTEN VAN DIJK is hoofd wetenschappelijk bureau van de HSK Groep in Arnhem en gz-psycholoog in opleiding tot specialist bij de HSK Groep in Arnhem.

PROF.DR. MARC VERBRAAK is bijzonder hoogleraar Gezondheidszorgpsychologie aan de Radboud Universiteit Nijmegen, Behavioural Science Institute; hoofd-

opleider gz-psychologen van SPON/ACSW Nijmegen; en inhoudelijk directeur van de HSK Groep in Arnhem.

Referenties

- Arrindell, W.A. & Ettema, J.H.M. (1986). *SCL-90: Handleiding bij een multidimensionele psychopathologie indicator*. Lisse: Swets & Zeitlinger.
- Arrindell, W.A. & Ettema, J.H.M. (2003). *SCL-90. Symptom Checklist. Handleiding bij een multidimensionele psychopathologie-indicator*. Lisse: Swets Test Publishers.
- Baert, S. (2012). Het gebruik van patiëntenfeedback in de klinische praktijk: een nieuw ROM-instrument. *Psychopraktijk*, 4, 27-31.
- Becker, C.B., Zayfert, C. & Anderson, E. (2003). A survey of psychologists' attitude towards and utilization of exposure therapy for PTSD. *Behaviour Research and Therapy*, 42, 277-292.
- Bordin, E.S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research and Practise*, 16, 252-260.
- Bringhurst, D.L., Watson, C.W., Miller, S.D. & Duncan, B.L. (2006). The reliability and validity of the outcome rating scale: A replication study of a brief clinical measure. *Journal of Brief Therapy*, 5, 23-30.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (second edition). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- De Doncker, D., Schotte, C., Vertommen, H. & Vankerckhoven, C. (1997). Assessment van de DSM-IV persoonlijkheidsstoornissen: Ontwikkeling en voorlopige resultaten van de ADP-IV-zelfbeoordelingsvragenlijst. *Tijdschrift Klinische Psychologie*, 27, 171-186.
- de Jong, K., Nugter, M.A., Polak, M.G., Wagenborg, J.E.A., Spinhoven, Ph. & Heiser, W.J. (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology & Psychotherapy* 14, 288-301.
- Derogatis, L.R. (1977). *SCL-90-R, administration, scoring & procedures manual-I for the R(evised) version*. Baltimore: Johns Hopkins University School of Medicine.
- Duncan, B.L., Miller, S.D., Sparks, J.A., Claud, D.A., Reynolds, L.R., Brown, J. & Johnson, L.D. (2003). The session rating scale: Preliminary psychometric properties of a 'working' alliance measure. *Journal of Brief Therapy*, 3, 3-12.
- Hafkenscheid, A. (2010). De Outcome rating scale (ORS) en de Session rating scale (SRS): Enkele psychometrische kenmerken van de Nederlandse versies. *Tijdschrift voor Psychotherapie*, 36, 394-403.
- Hafkenscheid, A., Duncan, B.L. & Miller, S.D. (2010). The Outcome and Session Rating Scales: A Cross-Cultural Examination of the Psychometric Properties of the Dutch translation. *Journal of Brief Therapy*, 7, 1-12.
- Hannan, C., Lambert, M.J., Harmon, C., Nielsen, S.L., Smart, D.W., Shimokawa, K. & Sutton, S.W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology*, 61, 155-163.
- Hansen, N.B., Lambert, M.J. & Forman, E.M. (2002). The psychotherapy dose-response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice*, 9, 329-343.
- Harmon, S.C., Lambert, M.J., Smart, D.W., Hawkins, E.J., Nielson, S.L., Slade, K. & Lutz, W. (2007). Enhancing outcome for potential treatment failures: Therapist-client feedback and clinical support tools. *Psychotherapy Research*, 17, 379-392.
- Havik, O.E. & VandenBosch, G.R. (1996). Limitations of manualized psychotherapy for everyday practice. *Clinical Psychology: Science and Practice*, 3, 264-267.

- Horvath, A.O. & Bedi, R.P. (2002). The Alliance. In: J. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 37-70; Reviewed). New York: Oxford University Press.
- Horvath, A.O. & Greenberg, L.S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, 64, 223-233.
- Horvath, A.O. & Symonds, B.D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology*, 38, 139-149.
- Jacobson, N.S., Follette, W.C. & Revenstorf, D. (1986). Toward a standard definition of clinically significant change. *Behavior Therapy*, 17, 308-311.
- Jacobson, N.S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19.
- Keijsers, G.P.J., van Minnen, A. & Hoogduin, C.A.L. (2011). Toepassing van protocollaire behandelingen bij psychische stoornissen. In: G. Keijsers, A. van Minnen & K. Hoogduin (Eds.), *Protocollaire behandelingen voor volwassenen met psychische klachten 2* (pp. 13-34). Amsterdam: Uitgeverij Boom.
- Lambert, M.J. & Barley, D.E. (2002). Research summary on the therapeutic relationship and psychotherapy outcome. In: J.C. Norcross (Ed.), *Psychotherapy relationships that work* (pp. 17-32). New York: Oxford University Press.
- Lambert, M.J., Burlingame, G.M., Umphress, V.J., Hansen, N.B., Vermeersch, D., Clouse, G. & Yanchar, S. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, 3, 106-116.
- Luborsky, L., Diger, L., Seligman, D.A., Rosenthal, R., Krause, E.D., Johnson, S., ... Schweizer, E. (1999). The researcher's own therapy alliances: a 'wild card' in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice*, 6, 95-106.
- Martin, D.J., Garske, J.P. & Davis, M.K. (2000). Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *Journal of Consulting and Clinical Psychology*, 68, 438-450.
- Miller, S.D., Duncan, B.L., Brown, J., Sorrell, R. & Chalk, M.B. (2006). Using formal client feedback to improve retention and outcome: Making ongoing, real time assessment feasible. *Journal of Brief Therapy*, 5, 5-22.
- Miller, S.D., Duncan, B.L., Brown, J., Sparks, J. & Claud, D. (2003). The Outcome Rating Scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analogue measure. *Journal of Brief Therapy*, 2, 91-100.
- Miller, S.D., Duncan, B.L. & Johnson, L. (2002). *SRS en ORS* (Nederlandse vertaling A. Hafkenscheid, i.s.m. D. Been, S. de Boer, A. Boon, P. Breukers, M. Crouzen & P. Teune).
- Mueller, R.M., Lambert, M.J. & Burlingame, G.M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment*, 70, 248-262.
- Norcross, J.C. & Wampold, B.E. (2011). What works for whom: Tailoring psychotherapy to the person. *Journal of Clinical Psychology*, 67, 127-132.
- Reese, R.J., Norsworthy, L.A. & Rowlands, S.R. (2009). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy Theory, Research, Practice, Training*, 46, 418-431.
- Stirman, S.W., DeRubeis, R.J., Crits-Christoph, P. & Brody, P.E. (2003). Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *Journal of Consulting and Clinical Psychology*, 71, 963-972.
- Stirman, S.W., DeRubeis, R.J., Crits-Christoph, P. & Rothman, A. (2005). Can the randomized controlled trial literature generalize to non-randomized patients? *Journal of Consulting and Clinical Psychology*, 73, 127-135.

- Stobie, B., Taylor, T., Quigley, A., Ewing, S. & Salkovskis, P.M. (2007). 'Contents may vary': A pilot study of treatment histories of OCD patients. *Behavioural and Cognitive Psychotherapy*, 35, 273-282.
- Umpress, V.J., Lambert, M.J., Smart, D.W. & Barlow, S.H. (1997). Concurrent and construct validity of the outcome questionnaire. *Journal of Psychoeducational Assessment*, 15, 40-55.
- van Beljouw, I.M.J. & Verhaak, P.F.M. (2010) *Geschikte uitkomstmaten voor routinematige registratie door eerste-lijnspsychologen*. Utrecht: Nivel.
- van Dijk, M.K., Verbraak, M.J.P.M., Oosterbaan, D.B. & van Balkom, A.J.L.M. (2012). Implementing practice guidelines for anxiety disorders in secondary mental health care: A case study. *International Journal of Mental Health Systems*, 6, 20. doi:10.1186/1752-4458-6-20
- Verbraak, M. (2012). Illusies van de psycholoog: Gezondheidszorgpsychologie in tijden van doelmatigheid. *Directieve Therapie*, 32, 235-255.
- Vervaecke, C.A.G. & Vertommen H. (1996). De Werk Alliantie Vragenlijst (wAV). *Gedragstherapie*, 2, 139-144.