

Over de legitimiteit van Routine Outcome Monitoring (ROM) lopen de gemoederen geregeld hoog op. Vanuit hun kennis en expertise over de grondslagen van psychologische metingen hadden psychologen, al dan niet via ROM, veel eerder een hoofdrol mogen opeisen in de discussie over de zin en onzin van ROM, stellen psycholoog Anton Hafkenscheid en psychiater Jim van Os. ‘Het is merkwaardig hoe zelfrapportagelijsten door voorstanders van verplichte ROM als valide indicatoren voor ‘behandeluitkomst’ worden gezien.

ROUTINE OUTCOME MONITORING 'BEWEZEN EFFECTIEF'? TWEE MISVATTINGEN OVER ROM

De toch al verhitte gemoederen over de legitimiteit van *Routine Outcome Monitoring* (ROM) liepen de eerste helft van 2017 verder op, gevoed door de kritische conclusies van de Algemene Rekenkamer en het kort geding tegen de Stichting Benchmark GGZ (SBG), aangespannen door de beweging 'Stop Benchmark ROM. Te midden van de steeds chaotischer en tumultueuzer verlopende discussies over ROM riep de voorzitter van het Nederlands Instituut van Psychologen (NIP), Linde Gonggrijp, vakgenoten op om het debat over ROM met inhoudelijke (in plaats van gevoels-) argumenten aan te gaan (haar blog van 22 maart 2017: <https://tinyurl.com/y79c7n9k>).

In het zomernummer van *De Psycholoog* werd de inhoudelijke discussie over rom in de ggz –toegesplitst op benchmarking – binnen deze beroepsvereniging voor het eerst publiekelijk opgepakt, met De Beurs (2017a) als verklaard voorstander, Cnubben & Delespaul (2017) als critici

en met Cramer (2017) die een tussenpositie inneemt. We zouden bijna vergeten dat de eigen visie van het nip op rom al jaren is terug te vinden op de nip-website (<https://tinyurl.com/yc7b9v3e>): 'Het nip ziet veel kansen met rom. rom opent de black box van de ggz. rom biedt een mogelijkheid om een objectieve maatstaf in te zetten om de voortgang van de behandeling te meten. Of dat nu de effectiviteit is van de behandeling, de kwaliteit van leven of het dagelijks functioneren. Inzicht in de voortgang kan helpen bij de besluitvorming over het voortzetten van de behandeling in een bepaalde richting of in een ander domein.'

Tegelijkertijd stelt het nip zich op het standpunt dat de rom-data (begin- en eindmetingen) die momenteel moeten worden aangeleverd 'veelal nog niet voldoende zeggingskracht' hebben om 'als sturingsinstrument gebruikt te worden'. Daarom meent het nip dat nader onderzoek naar de psychometrische waarde van de gebruikte meetinstrumenten een voorwaarde is om deugdelijk wetenschappelijk

onderzoek te kunnen doen naar de kwaliteit en effectiviteit van bepaalde behandelingen en zorgprogramma's (zie eerder genoemde link).

Het nip lijkt met haar stellingname op twee benen te hinken: je zou denken dat conclusies over de vele kansen die rom zou bieden (de nip-website noemt er vier: ondersteuning van passende zorg, vergelijken van uitkomsten op afdelings- en instellingsniveau, het genereren van 'big data' die uitnodigen voor wetenschappelijk onderzoek en verantwoording in benchmark) worden uitgesteld totdat (1) gebleken is hoe rom het er empirisch van afbrengt en (2) als goed is doordacht in hoeverre de theoretische grondslagen van rom deugdelijk zijn. Aan die twee condities is tot nu toe voorbij gegaan.

De ontwikkeling en psychometrische toetsing van testen, vragenlijsten en beoordelingsschalen in de ggz is bij uitstek de expertise en het werkterrein van psychologen. Als geen andere beroepsgroep kent de beroepsgroep van psychologen de waarde en de voordelen van gestandaardiseerde psychologische meetinstrumenten. Hoe meer professionals weten over een bepaald onderzoeks- of toepassingsgebied, des te beter zij ook de grenzen en beperkingen ervan in het vizier hebben. rom is grotendeels gebaseerd op subjectieve metingen. Vanuit hun kennis en expertise over de grondslagen van psychologische metingen hadden psychologen, al dan niet via het rom, veel eerder een hoofdrol mogen opeisen in de discussie over de zin en onzin van rom. De argumenten die ter verdediging van het vigerende rom-systeem in Nederland naar voren worden geschoven zijn in de regel niet erg inhoudelijk. Ze getuigen van weinig kennis of bewustzijn van de beperkingen van psychologische metingen in het algemeen, en van het 'objectiveren van subjectiviteit' in het bijzonder. Zoals wij verderop zullen laten zien, hebben deze argumenten vaak een hoog retorisch gehalte.

Het nip stelt terecht dat rom in de ggz staat of valt met de empirisch gebleken kwaliteiten van het gebruikte psychologisch meetinstrumentarium. Dat betekent dus in de eerste plaats meer (en beter) psychometrisch onderzoek naar rom-instrumenten zelf. Pas wanneer het monitorinstrumentarium

psychometrisch voldoende deugdelijk blijkt wordt de volgende stap zinvol: empirisch onderzoek waarmee wordt aangetoond dat 'uitkomsten' van zorg inderdaad beter worden door toevoeging van rom.

In dit artikel richten wij ons op twee misvattingen over rom, die in de discussies tot nu toe nauwelijks aandacht hebben gekregen: (1) dat rom-feedback onomstotelijk tot betere 'uitkomsten' van behandeling leidt en (2) dat rom-instrumenten als vanzelfsprekend 'uitkomsten' meten.

MISVATTING 1: ROM-FEEDBACK LEIDT ONOMSTOTELIJK TOT BETERE UITKOMSTEN

Auteurs die de huidige rom als kwaliteitsinstrument willen redden (bijv. De Jong & Van 't Spijker (2013), De Jong, Tiemens & Verbraak (2017), De Beurs (2017b)) grijpen steevast terug op de veelbelovende resultaten van het innovatieve onderzoeksprogramma van Lambert en collega's (Lambert, 2007) in de Verenigde Staten. Dit onderzoeksprogramma omvat een serie gerandomiseerde vergelijkende effectstudies. Alle deelnemende patiënten vullen voorafgaand aan elk behandelcontact de *Outcome Questionnaire* (OQ-45) als rom-instrument in. Binnen de onderzoekconditie (Feedbackconditie: Fb) krijgt de therapeut (en in sommige studies ook de patiënt) de scores die de patiënt toekent op de OQ-45 elke keer te zien. Het individuele scoreverloop van de gevolgde patiënt wordt in een grafiek weergegeven. In die grafiek wordt dat individuele scoreverloop gerelateerd aan het collectieve scoreverloop van grote normgroepen die eerder (succesvol of niet succesvol) in behandeling zijn geweest. Op grond van de vergelijking van het individuele met het (genormeerde) collectieve scoreverloop krijgt de behandelaar terugkoppeling over de mate waarin er progressie (*On Track*) of stagnatie (*Not On Track*: NOT) is in de behandeling. Deze methode staat bekend als *outcome tracking*. In de controleconditie (TAU: *Treatment As Usual*) vult de patiënt de OQ-45 eveneens voorafgaand aan elke behandelingsessie in, maar de behandelaar (of de patiënt) krijgt nimmer feedback over het scoreverloop. De Lambert-groep toetst in elk van de afzonderlijke rom-studies of de patiënten

Van een 'sturend' gebruik van routine monitoren is sprake wanneer de monitorscores bedoeld zijn om de behandeling rechtstreeks te beïnvloeden

in de Fb-conditie meer en/of sneller vooruitgang boeken dan de patiënten in de TAU-conditie, waarbij het behandelverloop dus operationeel wordt gedefinieerd als het scoreverloop op de OQ-45.

De Lambert-groep voerde zelf een meta-analyse uit (Lambert & Shimokawa, 2011), die zes studies met in totaal meer dan zesduizend patiënten omvat. Uit die meta-analyse blijkt dat continue systematische feedback aan de therapeut over het behandelverloop inderdaad tot een hogere effectiviteit van ggz-behandelingen leidt, gecombineerd met een lagere voortijdige uitval ('drop out'). Met name patiënten die een verhoogd risico lopen op een niet geslaagde behandeling en/of op voortijdige uitval (NOT), zouden van ROM-feedback profiteren. In genoemde meta-analyse worden ook drie kleinere studies met een ander ROM-instrumentarium geïncorporeerd. Die zijn qua opzet vergelijkbaar met de onderzoeken van Lambert en collega's. In plaats van de meer omvattende OQ-45 gebruiken deze drie studies de *Outcome Rating Scale* (ORS): een ultrakorte zelfrapportageschaal. Anders dan in de studies van Lambert en collega's, waar de kwaliteit van de therapeutische alliantie niet routinematig wordt gemeten, wordt in deze drie kleinere studies een ultrakorte alliantievragenlijst, de *Session Rating Scale* (SRS), standaard afgenomen. Met enkele methodologische slagen om de arm komt ook routine monitoren op basis van de ORS uit de meta-analyse als werkzaam naar voren.

Een recente, bredere, aanzienlijk grotere en onafhankelijker meta-analyse is die van Gondek, Edbrooke-Childs, Fink, Deighton & Wolpert (2016). De helft van de 32 studies (waarin vrijwel alle studies uit de meta-analyse van Lambert & Shimokawa (2011) zijn opgenomen) uit hun meta-analyse gebruikte de OQ-45, nog eens zes andere studies gebruikte de SRS of een combinatie van de ORS en SRS. Andere gebruikte monitorinstrumenten waren bijvoorbeeld de *Symptom Checklist* (SCL-90), en in sommige studies werden meerdere monitorinstrumenten (effectmaten) tegelijk gebruikt. In iets meer dan de helft van de studies (56%) bleek de Fb-conditie op tenminste één effectmaat statistisch significante betere behandeluitkomsten op te leveren dan de TAU-conditie. ROM-feedback werkte het best als die zowel aan patiënt als behandelaar werd verschaft. Op grond van hun meta-analyse achten de auteurs van deze meta-analyse de potentiële waarde van ROM afdoende bewezen. Ze bepleiten vooral meer onderzoek naar de condities waaronder en de patiëntgroepen waarvoor ROM meerwaarde heeft.

en van Gondek et al. (2016) lijken op het eerste gezicht een aardige steun in de rug voor het over een brede linie handhaven van verplichte ROM in Nederland. Daarbij passen echter twee kanttekeningen.

Allereerst steekt de Nederlandse ROM, met laagfrequente metingen en klinici die meestentijds ongetraind zijn in de verantwoorde interpretatie van ROM-feedback, schraal af tegen de veel verfijndere studies (met continue metingen en directe grafische feedback) uit genoemde meta-analyses. De ROM-studies door de Lambert-groep verschillen inhoudelijk dermate drastisch van het armzalige Nederlandse ROM-systeem dat de Amerikaanse resultaten geen rechtvaardiging kunnen zijn voor ggz-brede implementatie in Nederland.

In de tweede plaats: ondanks de vernuftige opzet van het Amerikaanse onderzoek naar *outcome tracking* is op de validiteit van die opzet wel iets af te dingen. Lambert en collega's gebruiken hun OQ-45 als meetinstrument om 'sturend' te monitoren (Hafkenscheid, 2009). Van een 'sturend' gebruik van routine monitoren is sprake wanneer de monitorscores bedoeld zijn om de behandeling rechtstreeks te beïnvloeden. Therapeuten en/of patiënten ontvangen gedurende de behandeling regelmatig feedback over het scoreverloop op het monitorinstrument. Hetzelfde ROM-instrument dat wordt gebruikt om de therapeut (en/of de patiënt) systematisch feedback te verschaffen, wordt tegelijkertijd gebruikt om vast te stellen wat de toegevoegde waarde is van die systematische feedback. De OQ-45 is dus tegelijkertijd zowel monitorinstrument (predictor) als effectmaat (criterium). Deze keuze zet de deur wijd open voor 'contaminatie' van de onafhankelijke variabele (feedback op basis van de OQ-45) met de afhankelijke variabele (de OQ-45 als criterium voor de effectiviteit van feedback). Anders gezegd: door dit design wordt principieel onbeslisbaar of de scores veranderen (a) omdat feedback over de scores de behandeling daadwerkelijk sneller en effectiever maken, of (b) omdat de patiënt anders gaat scoren in de wetenschap dat de therapeut over de scores gaat praten, nadat deze ze onder ogen heeft gehad.

Lambert & Shimokawa (2011) bespreken deze zwakte gek genoeg uitsluitend in relatie tot de drie studies naar *outcome tracking* met de ORS (en de SRS). Binnen het ROM-systeem met de ORS (en SRS) is het elke sessie bespreken van de scores integraal onderdeel van de aanpak. Dat is slechts een gradueel verschil met hun eigen ROM-systeem, waarin het bespreken van de scores op de OQ-45 facultatief (ter keuze aan de behandelaar) is. Hun kritiek op het *outcome tracking*

TABEL 1. VERGELIJKENDE STUDIES TOEGEVOEGDE WAARDE ROM-FEEDBACK BINNEN NEDERLAND EN VLAANDEREN

Studie	N=	Onderzoekslokatie/ Diagnostische classificaties	Instrumenten	Conclusies
Bovendeerd & De Groot (2013)	Fb: 9 TAU: 8	<i>Dimence/vestiging Steenwijk</i> Angststoornis of Stemmingsstoornis	Fb: SRS, ORS Effect: OQ-45	Toevoeging van monitoring met feedback leidt niet tot betere behandelresultaten bij kortdurende behandeling.
Janse et al. (2016)	Fb: 461 TAU: 545	<i>HSK/</i> Angst-, Stemnings-, Aanpassings-, Somatoform, soms (< 10%) gecombineerd met PS	Fb: SRS, ORS Effect: SCL-90	Monitoring met feedback leidt niet tot effectievere behandeling, behalve bij patiënten met een stemmingsstoornis.
De Jong (2012)	Fb therapeut: 159 Fb therapeut én patiënt: 172 TAU: 144	<i>Psychiatrische poliklinieken, vrijgevestigde praktijken/</i> Angst-, Stemnings-, Aanpassingsstoornis, relatieproblemen, PS (37%), Comorbiditeit As I en As II: 36%.	Fb: OQ-45 Effect: OQ-45 Focus op <i>Not On Track (NOT)-</i> patiënten	Monitoren met feedback is effectiever wanneer zowel therapeut als patiënt feedback ontvangen. Feedback aan therapeut voegt alleen iets toe bij NOT-patiënten die kortdurend in behandeling zijn.
Van Oenen (2017)	Fb: 72 TAU: 57	<i>Arkin Crisiscentrum/</i> Psychose, Stemmings-, Aanpassings-, PS	Fb: SRS, ORS Effect: BSI, OQ-45	Monitoren met feedback verbetert de behandel-effectiviteit niet en kan zelfs contraproductief zijn.
Smits (2016)	Fb: 25 TAU: 27	<i>Vlaamse Centra Geestelijke Gezondheidszorg of vrijgevestigde praktijken/</i> DSM-IV-classificaties niet beschikbaar	Fb: diverse instrumenten QIT- Online, waaronder OQ-45, WAI, SRL Effect: WAI, SRL	Therapeuten van feedback voorzien over het proces en de tussentijdse effecten van de behandeling heeft geen positieve invloed op de productiviteit van de behandeling op proces- of uitkomstniveau.
Winkelhorst e.a. (2013)	Fb: 25 TAU: 25	<i>Dimence Afdeling Persoonlijkheidsstoornissen.</i> DSM-IV-TR As I en As II	Fb: SRS, ORS Effect: OQ-45	Routine monitoren met feedback voegt weinig toe aan de behandel-effectiviteit.

Toelichting: Fb: monitoren met feedback; TAU: monitoren zonder feedback. NOT (*Not On Track*): herstelcurve voor achtereenvolgende OQ-45-scores slechter dan statistisch voorspeld.

BSI: Brief Symptom Inventory; OQ-45: Outcome Questionnaire; ORS: Outcome Rating Scale; SCL-90: Symptom Checklist; SRL: Session Rating List; SRS: Session Rating Scale; WAI: Working Alliance Inventory; QIT-Online: Vlaams digitaal ROM-systeem; PS: persoonlijkheidsstoornis. Dik gedrukt in kolom Conclusies: ondersteunende evidentie.

met de ORS geldt in beginsel dus net zozeer voor hun eigen ROM-systeem. Als de resultaten van de Amerikaanse ROM-studies worden aangehaald als legitimering van het nationale ROM-systeem in Nederland, moet een belangrijk voorbehoud worden gemaakt: het is onbekend in hoeverre de relatief gunstige Amerikaanse resultaten geflatteerd zijn door 'sturend monitoren' als artefact.

STUDIES NAAR DE TOEGEVOEGDE WAARDE VAN

ROM-FEEDBACK DICHTER BIJ HUIS Er zijn ons zes recente vergelijkende effectstudies uit Nederland en Vlaanderen bekend, waarin is onderzocht of en in hoeverre routine monitoren met feedback aan de therapeut de effectiviteit van behandeling verhoogt. In totaal omvatten de onderzoeksgroepen in de Fb-conditie 923 patiënten; de TAU-conditie bestaat, geaggregeerd over de zes studies, uit 806 patiënten. In tabel 1 staan de zes studies samengevat weergegeven:

De studies van Janse et al. (2016) en van De Jong (2012) zijn veruit de grootste in steekproefaantallen. De studies van Bovendeerd & De Groot (2013), van Winkelhorst et al. (2013) en van Smits (2016) zijn in steekproefomvang nogal klein, waarmee ze *underpowered* zijn. De studie van Van Oenen (2017) zit er wat betreft aantal onderzochte patiënten tussenin. De N's in de tabel zijn steeds gebaseerd op de patiënten die uiteindelijk in de statistische analyses konden worden betrokken, dus voortijdige uit- of afvallers maken de steekproeven waarmee de onderzoekers begonnen groter. De vergelijkende effectstudies omvatten ambulante patiënten en alle gebruikte effectmaten zijn gebaseerd op zelfbeoordelingen van de patiënt. Het zwaartepunt in de zes studies (met uitzondering van Van Oenen, 2017) ligt op mildere psychopathologie. Vier van de zes studies gebruiken als effectmaat een ander meetinstrument dan het ROM-instrument zelf, in de twee andere studies (De Jong, 2012; Smits, 2016) is het ROM-instrument, net als in de Amerikaanse studies, de effectmaat.

De rechter kolom van tabel 1 vat de conclusies van elke studie samen. Al met al wordt in de zes Nederlandse studies weinig empirische ondersteuning gevonden voor verhoging

van behandel-effectiviteit door ROM-feedback. De tegenval-lende Nederlandse en Vlaamse resultaten kunnen deels worden gerelativeerd: sommige studies hadden een lage statistische *power* en geen van de zes studies had de verfijndheid die het onderzoeksprogramma van de Lambert-groep kenmerkt. Daar staat tegenover dat in het merendeel van de studies ROM-instrument en effectmaat – terecht – verschillen, waardoor de effecten minder geflatteerd worden door 'sturend monitoren'.

Op grond van deze nogal povere resultaten in het Nederlandse taalgebied kunnen degenen die pleiten voor continuering van een verplichte ggz-brede ROM zich in elk geval niet beroepen op het argument dat ROM inmiddels ondubbelzinnig 'bewezen effectief' is. Er is echt meer en beter onderzoek op dit gebied nodig. Een lopend (dissertatie) onderzoek (Bovendeerd et al., ingediend) probeert de zwaktes van eerdere studies binnen het Nederlandse taalgebied te ondervangen.

MISVATTING 2: ROM-INSTRUMENTEN METEN ALS VANZELFSPREKEND 'UITKOMST'

De Beurs (2017a; 2017b), directeur van de SBG, en Westert (2017), onafhankelijk lid van de Wetenschappelijke Raad van de SBG, menen dat critici van ROM 'een karikatuur maken'. Deze critici voeren volgens hen 'een achterhoedegevecht' met hun 'fabeltje dat de ggz zo ingewikkeld is dat de behandelresultaten niet zijn vast te stellen of niet zijn te meten'. Deze vertegenwoordigers van de SBG schuwen de retoriek niet. Westert (2017): 'Een patiënt met psychische problemen komt niet in zorg voor een goed gesprek, of om tevreden te zijn over de bejegening of de fijne relatie met de behandelaar. In de curatieve ggz wil de gemiddelde patiënt van zijn psychische klachten af en zo snel mogelijk weer functioneren zoals voorheen, of op zijn minst weer zo goed als mogelijk met een chronische psychiatrische aandoening.' Westert (2017) impliceert dat critici van de opgelegde nationale ROM niet 'uitkomstgericht' zouden zijn en zich liever onderdompelen in een plezierig en conflictvermijdend therapeutisch proces, waarbij patiënt en therapeut zich allebei comfortabel

Degenen die pleiten voor continuering van een verplichte ggz-brede rom kunnen zich niet beroepen op het argument dat rom ondubbelzinnig 'bewezen effectief' is

In de zes Nederlandse studies is weinig empirische ondersteuning gevonden voor verhoging van behandel-effectiviteit door ROM-feedback

voelen, zonder zich er al te druk om te maken of de patiënt baat heeft bij de behandeling.

In onze kritiek op de door de Lambert-groep gevolgde onderzoeksopzet lag al besloten dat ROM meer *proces*beïnvloeding behelst dan dat er sprake is van onafhankelijke effectmeting. Natuurlijk hebben pleitbezorgers van het huidige ROM-systeem een punt als zij betogen dat het proces (dus ook de procesmeting) in de ggz per saldo ondergeschikt moet zijn aan de 'uitkomst' of het 'product'. Wat zij echter lijken te vergeten, is dat 'proces' en 'product' in de ggz bepaald niet altijd eenvoudig uit elkaar zijn te trekken. Als bijvoorbeeld een extreem sociaal fobische patiënt in de loop van de behandeling de therapeut (als autoriteitsfiguur) voor het eerst durft tegen te spreken, kan dat met recht een klinisch relevante 'tussentijdse uitkomst' worden genoemd, zonder dat die 'uitkomst' zich noodzakelijk vertaalt in een scoredaling op een zelfrapportagelijst voor sociale angst.

Er bestaat een complexe verhouding tussen de 'uitkomsten' die met een gestandaardiseerde en psychometrische zelfrapportagelijst worden gevonden en de 'uitkomsten' die op een informele wijze tijdens het therapeutisch proces worden vastgesteld. Het is merkwaardig hoe zelfrapportagelijsten door voorstanders van verplichte ROM zonder blikken of blozen met de grootste vanzelfsprekendheid als valide indicatoren voor 'behandeluitkomst' (anders gezegd: als valide veranderingsmaten) worden gezien. Over de veel gebruikte ROM-instrumenten als de OQ-45 en de HONOS schrijft De Beurs (2017b) bijvoorbeeld: 'Dat generieke meetinstrumenten ongeschikt zijn voor ROM is overigens niet juist: met instrumenten als de OQ-45 of de HONOS is bij de meeste patiënten goed te meten of zij baat hebben bij de behandeling (De Jong et al., 2007; Mulder et al., 2004).' Bij raadpleging van de door De Beurs (2017b) aangehaalde literatuurbronnen (het artikel van De Jong et al. (2007) is in

het Nederlands vertaald en in 2008 gepubliceerd in *Psychologie & Gezondheid*) blijft raadselachtig hoe hij tot zijn conclusies komt. Genoemde psychometrische evaluaties van de OQ-45 en de HONOS hielden zich voornamelijk bezig met klassieke betrouwbaarheidskenmerken, dimensionale structuur en concurrente validiteit. Op deze kenmerken komen beide meetinstrumenten er inderdaad behoorlijk goed van af.

Wat betreft de verandergevoeligheid van de HONOS op het niveau van de individuele patiënt ligt dat anders. Nugter et al. (2012) concluderen uit hun empirisch onderzoek: 'De 'reliable change index' liet weinig individuele verandering zien, ook bij groepen waar die verandering werd verwacht.' De mate waarin de OQ-45 gevoelig is voor verandering werd slechts getoetst in een selecte subgroep van zestig patiënten die aan een behandeling van maximaal vijf sessies voldoende hadden. De stabiliteit (test-hertestbetrouwbaarheid) bij herhaalde meting werd in een groep van niet meer dan 42 patiënten bepaald. Het *berekenen* van de stabiliteit van een test of vragenlijst is simpel als er voldoende herhaalde metingen beschikbaar zijn. Het *verzamelen* van herhaalde metingen is vaak echter een bewerkelijke en dure aangelegenheid. Om deze reden wordt doorgaans voor een praktische, maar ongewenste gelegenhedsoplossing gekozen: de indexen voor statistisch betrouwbare verandering worden simpelweg berekend met Cronbach's alpha (interne consistentie) als betrouwbaarheidscoëfficiënt. Cronbach's alpha schat betrouwbaarheid op basis van één enkel meetmoment en negeert daarmee de onbetrouwbaarheidsfluctuaties in scores die zijn toe te schrijven aan meerdere afnames van hetzelfde instrument op verschillende tijdstippen. Deze onbetrouwbaarheidsbron moet worden verdisconteerd om de onbetrouwbaarheidsmarges te bepalen, die in acht moeten worden genomen om vast te stellen of intra-individuele scoreveranderingen statistisch betrouwbaar zijn (Crutzen, 2014). Gebeurt dat niet, dan kunnen conclusies over de effectiviteit van een behandeling (met name positief) worden vertekend (Hafkenscheid, 1994).

Los van deze psychometrische kanttekeningen kan er een inhoudelijk vraagteken worden geplaatst bij de aanname dat de OQ-45 een onbevooroordeelde 'uitkomstmaat' is. Eerder werd als kritiek op de iteminhoud van de OQ-45 geuit dat maar liefst acht van de OQ-45 items veronderstellen dat de patiënt werkt en/of een schoolopleiding volgt, waarmee bijna 18% van de items uit dit ROM-instrument irrelevant (en onbedoeld stigmatiserend) is voor patiënten die door diverse omstandigheden niet (meer) tot werken of tot het volgen van

een opleiding in staat zijn (Hafkenscheid, 2008). Daarmee kon de OQ-45 al niet echt een generiek ROM-instrument worden genoemd. Die kritiek is inmiddels met een stopmiddel ondervangen: aan de oorspronkelijke instructie is toegevoegd dat 'werk' wordt gedefinieerd als '... baan, school, huishoudelijk werk, vrijwilligerswerk, enz.'

Twee andere vragen uit de OQ-45 zijn evenmin generiek: de vraag naar hoe ongelukkig de patiënt is over haar of zijn huwelijk/relatie (item 7) en de vraag naar hoe onbevredigend de seksleven van de patiënt is (item 17). De patiënt wordt geïnstrueerd om bij item 7 'Nooit' in te vullen wanneer zij of hij alleenstaand is. Het *ontbreken* (of *ontberen*) van een liefdesrelatie – voor heel wat patiënten een ongelukkig makend probleem – wordt in de scoretelling dus hetzelfde gewaardeerd als *het gelukkig zijn* binnen een liefdesrelatie. Een soortgelijke merkwaardigheid doet zich voor bij de beantwoording van de vraag naar seksuele satisfactie (item 17). Uit klinische ervaring weten wij hoe onprettig en disrespectvol patiënten de ambiguïteit ervaren van dergelijke vragen die niet op hen van toepassing zijn en toch beantwoord dienen te worden. Ten overvloede: het is gezinszins onze bedoeling om onze kritiek specifiek op de OQ-45 (HONOS) te richten. In het verleden zij wij niet minder kritisch geweest op onderzoekers die over andere ROM-instrumenten zoals de SCI-90 (Hafkenscheid, Maassen & Veeninga, 2007) of de *Brief Symptom Inventory* (BSI; Hafkenscheid, 2007) wat al te kritiekloos rapporteerden.

BLACK BOX GGZ Collega's en organisaties die het heersende ROM-systeem hebben omarmd, zien ROM als middel bij uitstek om de 'black box van de ggz te openen. Benchmarken op basis van ROM-scores zou ertoe leiden dat inzicht wordt verkregen in 'hoe het beter kan door bij anderen af te kijken hoe ze goede resultaten boeken', waarbij die opgedane kennis kan worden gebruikt om eigen resultaten te verbeteren (De Beurs, 2017a).

Deze pleitbezorgers van het opgelegde ROM-systeem zetten hun argumenten graag kracht bij door nogal klakkeloze vergelijkingen te trekken met de somatische geneeskunde en met het onderwijs. Onder andere in de discussie die in *Medisch Contact* volgde op het eerder genoemde stuk van Westert (2017) worden deze sectoren de ggz ten voorbeeld gesteld: daar zou over uitkomstindicatoren tenminste niet zo moeilijk worden gedaan. In genoemde discussie haalt Koekkoek (2017) instemmend de ranglijsten aan, waarmee dagblad *Trouw* jaarlijks de verschillen in slagingspercentages en gemiddelde schoolcijfers tussen scholen in het voortgezet onderwijs per provincie presenteerde. Deze overzichten gaven stakeholders – aankomende leerlingen, hun ouders, de onderwijsinspectie en de vergeleken onderwijsinstellingen zelf – de gelegenheid om hun keuzes te bepalen (ouders en aankomende leerlingen) en beleidsmaatregelen (inspectie, besturen: de bekende 'verbetertrajecten') te treffen. Koekkoek vermeldt niet dat *Trouw* in 2012 stopte met deze 'Schoolprestaties', zoals de jaarlijkse overzichten van rapportcijfers in het voortgezet onderwijs werden genoemd. De ranglijsten waren namelijk uiterst controversieel. Tegenstanders benadrukten dat de vergelijking van schoolprestaties tussen verschillende onderwijsinstellingen zinloos is zonder de context in acht te nemen. Een middelbare school met veel allochtone leerlingen uit achterstandsgezinnen met laag opgeleide ouders staat in schoolprestaties a priori op achterstand ten opzichte van een middelbare school met veel autochtone kinderen uit welgestelde gezinnen met hoog opgeleide ouders. Vergelijkbaar met Koekkoek in zijn verdediging van de huidige nationale ROM bleven de voorstanders van de ranglijsten binnen de onderwijssector volhouden dat critici die de eerlijkheid van de vergelijking tussen onderwijsinstellingen in twijfel trokken zich wilden onttrekken aan hun maatschappelijke verantwoordingsplicht.

Hoe het ook zij: de vergelijking die Koekkoek trekt tussen

Er bestaat een complexe verhouding tussen de 'uitkomsten' die met een gestandaardiseerde en psychometrische zelfrapportagelijst worden gevonden en de 'uitkomsten' die op een informele wijze tijdens het therapeutisch proces worden vastgesteld

het huidige nationale ROM-systeem met 'Trouw Schoolprestaties' gaat inhoudelijk mank. Rapportcijfers en slagingspercentages zijn onmiskenbaar duidelijke indicatoren waarmee geleverde prestaties in elk geval objectief kunnen worden vergeleken, nog los van de vraag hoe eerlijk of zinvol deze prestatievergelijkingen zijn. Datzelfde geldt voor andere psychologische variabelen zoals IQ-scores. De Jong, Tiemens & Verbraak (2017) halen eveneens ongelukkige voorbeelden aan ter ondersteuning van benchmarken in de ggz: percentages patiënten met overmatig middelengebruik of percentages sterfte na maagkankeroperaties zijn ondubbelzinnig telbaar en daarmee onvergelijkbaar met percentages verbeterde of verslechterde patiënten, vastgesteld aan de hand van scoreveranderingen op ROM-instrumenten.

Benchmarken op basis van opinie of gevoelens is uitermate riskant

ZELFRAPPORTAGE ROM-scores op (meestal) zelfrapportagelijsten verwijzen vrijwel nooit naar objectieve prestaties, maar representeren niet meer en niet minder dan eigen belevingen van die prestaties. Benchmarken op basis van opinie of gevoelens is uitermate riskant. In de onderwijssector zouden maar weinigen bijvoorbeeld genoeg willen nemen met een benchmark die tot stand is gekomen met een vragenlijst waarin leerlingen zelf mogen aangeven hoe goed ze zijn in schoolvakken zoals wiskunde, in plaats van te benchmarken op 'harde' rapportcijfers voor dat vak. En in de somatische geneeskunde verlaten diabetologen zich toch liever op gemiddelde bloedsuikerspiegels (HbA1c) die met laboratoriumonderzoek zijn bepaald dan door patiënten hun bloedsuikerspiegels te laten schatten op basis van hoe moe of fit zij zich voelen. Daarmee is geenszins gezegd dat de beleving van de patiënt er niet toe zou doen.

Zelfrapportagelijsten zijn bij uitstek geschikt om belevingen (subjectiviteit) te objectiveren. Met 'objectiveren' wordt slechts bedoeld het 'gestandaardiseerd vastleggen in scores'. Deze beperkte betekenis mag niet worden verward met het blootleggen van 'een objectieve (dus niet vertekende) werkelijkheid' (Hafkenscheid & Van Os, 2016)

Zelfrapportagelijsten leggen de werkelijkheid niet bloot, maar zijn constructies van die werkelijkheid: constructies van

degenen die de items en antwoordalternatieven hebben ontworpen, en van de degenen die de vragenlijst invullen. Respondenten geven hun eigen interpretaties en kleur aan formuleringen aan vragen naar bijvoorbeeld de mate waarin hun huwelijksleven vervuld is. Niet de beleving van de patiënt staat bij de vragenlijstmethode ter discussie, maar wel de wijze waarop voorstanders van de nationaal verplichte ROM-vragenlijstscores interpreteren alsof het laboratoriumwaarden zijn. Degenen die beweren dat ROM-metingen 'de black box van de ggz openen' vallen wat ons betreft ten prooi aan een bedenkelijke vorm van naïef empirisme: de aanname dat de dingen zijn zoals men ze in eerste instantie waarneemt.

Het huidige verplichte ROM-systeem past typisch binnen een klassiek medisch model, waarbij patiënten binnen afzonderlijke, goed afgebakende diagnosegroepen een min of meer vast (eventueel protocollair) behandeltraject doorlopen. Iedere wat meer ervaren behandelaar weet dat zelfs iets ingewikkelder klachtgerichte behandelingen meestal een onvoorspelbaar en soms ronduit grillig verloop kennen, waarbij patiënten zich geenszins puur op basis van rationele argumenten (voor hun eigen bestwil) 'willoos' aan een bewezen effectieve behandeling onderwerpen. Dat onvoorspelbare en op het oog rommelige karakter van veel behandelingen zal naar verwachting toenemen met de opkomst van de gedeelde besluitvorming in de ggz. Zeker bij persoonlijkheidsproblematiek en bij ernstige psychiatrische aandoeningen bestaat een groot deel van het therapeutisch werk uit het motiveren van patiënten voor veranderingstrajecten waar zij ambivalent of in eerste instantie afkerig tegenover staan. Dat misschien wel belangrijkste deel van de therapeutische arbeid laat zich niet gemakkelijk in het keurslijf van de verplichte ROM wringen.

De gewenste 'uitkomsten' van ggz-behandelingen staan zelden gelijk aan simpele, lineaire klacht- of symptoomreductie. Bij heel veel ggz-problematiek is er sprake van door de patiënt (en diens omgeving) als onoverbrugbaar ervaren intra-persoonlijke of interpersoonlijke spanningen. Intra-persoonlijke spanningen ontstaan onder meer wanneer mensen hun zelfbeeld en hun zelfideaal niet in overeenstemming weten te krijgen, interpersoonlijke spanningen onder andere wanneer mensen zich te vaak en te zeer anders naar anderen gedragen dan die anderen willen of acceptabel vinden. Het criterium voor behandel succes is dan de mate waarin discrepantie zich naar convergentie beweegt – bijvoorbeeld het opschuiven van het zelfbeeld in de richting van het zelfideaal (of andersom), of het verkleinen van de kloof tussen de gedragspatronen die patiënt en

naastbetrokkenen van elkaar verwachten en als acceptabel ervaren. Wanneer patiënten over de loop van de behandeling meer in harmonie met zichzelf en/of hun omgeving gaan leven, volgt niet zelden klachtenreductie, en staan de klachten en symptomen op zijn minst minder centraal.

CONCLUSIES

De afgelopen jaren hebben wij, in diverse auteurssamenstellingen, onze bedenkingen tegen de nationaal verplichte ROM, en de ideologie erachter, in een reeks publicaties naar voren gebracht (Hafkenscheid, 2008, 2009, 2010, 2012, 2013, 2017; Hafkenscheid & Van Os, 2013, 2016; Van Os et al., 2012; Van Os, Berkelaar & Hafkenscheid, 2017). In elk van die publicaties hebben wij verschillende kritiekpunten geaccentueerd. We hebben ons ook ingespannen om niet met kritiek te volstaan, maar om ook met alternatieven te komen (Hafkenscheid, 2016; Hafkenscheid & Van Os, 2014a, 2014b; Van Os & Delespaul, 2017).

In dit artikel hebben wij ons gericht op twee misvattingen over het huidige ROM-systeem, die in de discussies tot nu toe onderbelicht bleven: (1) dat ROM-feedback onomstotelijk tot betere 'uitkomsten' van behandeling leidt en (2) dat ROM-instrumenten als vanzelfsprekend 'uitkomsten' meten. Wij willen niet onder stoelen of banken steken dat het huidige, opgelegde ROM-systeem ons een ware gruwel is. Als fervente voorstanders van kwaliteitsmonitoring, van toetsbaarheid en van reflectief diagnostisch en therapeutisch handelen, hebben wij met lede ogen moeten toezien hoe een log, opgelegd bureaucratisch systeem, zonder enige afstemming, ervoor heeft gezorgd dat lokale initiatieven om de kwaliteit van zorg *lean and mean* (dus afgestemd op de specifieke problematiek van de patiënt) te monitoren, vakkundig om zeep zijn geholpen. Ongetwijfeld troostrijk en verbindend bedoelde aanbevelingen om het schrale en nutteloze 'minimale' ROM-systeem gewoon aan te vullen door vaker tussentijds te meten en er monitorinstrumenten aan toe te voegen die wel zinvol zijn (De Beurs, 2017a, 2017b), zetten de wereld op zijn kop. We beschouwen ze als obligate doekjes voor het bloeden. Degenen die pleiten om ROM 'van binnenuit' te veranderen, voeren nogal weleens ter verdediging aan dat er nu eenmaal niets beters voorhanden is, dat ROM gewoon 'werk in uitvoering' is, en dat 'niets perfect' kan zijn. Ook wordt doorgaan op de ingeslagen weg verdedigd met het argument dat het veld een slechte beurt maakt door er voortijdig de brui aan te geven: de reputatie van de GGZ is namelijk toch al niet zo best (Schoevers & Beekman, 2017). Wij vinden deze argumenten echt onvoldoende om een

Liever ten halve gekeerd dan ten hele gedwaald

inherent krakkemikkig systeem overeind te houden. Een systeem dat alleen ideologisch en retorisch (en nauwelijks met inhoudelijke argumenten) kan worden gerechtvaardigd.

Om misverstanden te voorkomen willen wij benadrukken dat onze bedenkingen tegen de nationaal verplicht gestelde ROM niets te maken hebben met animositeit tegen zorgverzekeraars, al hebben wij persoonlijk sterke twijfels of de stelselwijziging in de zorg ooit de ambitie kan waar maken van betere zorg voor een lagere (althans minder hoge) prijs. Het zou unfair zijn om de zorgverzekeraars als niet ter zake deskundigen aan te rekenen dat zij de voor de hand liggende, maar naïeve, gedachte koesterden dat met ROM twee vliegen in één klap geslagen konden worden: optimalisering van het individuele behandelproces (inclusief gedeelde besluitvorming) en benchmarken op afdelings- of instellingsniveau. Niet ter discussie staat dat professionals van buitenaf op elk moment moeten kunnen worden aangesproken op de kwaliteit en de doelmatigheid van hun professionele handelen. Professionals dreigen echter het vertrouwen van hun patiënten te verliezen zodra ze hun oren laten hangen naar een opgelegd systeem, dat haaks staat op hun eigen professionele handelen, zelfs als zij zich conformeren met het mes op de keel.

Liever ten halve gekeerd dan ten hele gedwaald: om kwaliteitstoetsing niet verder verloren te laten gaan zullen de beroepsgroepen in de ggz bereid moeten zijn de nationaal opgelegde ROM in de huidige vorm vaarwel te zeggen en zelf weer voortvarend te gaan bouwen aan voldoende gedifferentieerde routine monitorsystemen, die recht doen aan de complexiteit van de klinische realiteit, en die zich lenen voor evaluaties van beoogde ten opzichte van geëffectueerde zorg. Om met Schoevers & Beekman (2017) te spreken: 'Van 'stop ROM naar 'hier met die ROM!'

OVER DE AUTEURS

Anton Hafkenscheid is klinisch psycholoog-psychotherapeut en werkzaam bij Arkin/Sinai Centrum. Jim van Os is hoogleraar psychiatrische epidemiologie en publieke GGZ, UMC Utrecht Hersencentrum, Utrecht. Voor beide auteurs geldt dat dit artikel op persoonlijke titel is geschreven. Correspondentie aangaande dit artikel: a.hafkenscheid@sinaicentrum.nl.

Literatuur

- De Beurs, E. (2017a). ROM en benchmark: goed of slecht huwelijk. *De Psycholoog*, 52, 46-49.
- De Beurs, E. (2017b). Over ROM en benchmarken; reactie van SBG op rapport van de Algemene Rekenkamer en de actie 'Stop ROM'. *Tijdschrift voor Psychiatrie*, 59, 238-241.
- Bovendeerd, A.M. & de Groot, E.I.W. de (2013). Het effect van systematische feedback op de uitkomst van een kortdurende wachtlijstbehandeling in de ggz. *Tijdschrift voor Psychotherapie*, 39, 157-171.
- Bovendeerd, B, de Jong, K., Colijn, S., de Groot, E.I.W., Hafkenscheid, A., Duncan, B.L., de Keijser, J. (ingediend). Systematic client feedback to brief therapy in basic mental health care: study protocol for a multicenter cluster randomized controlled trial.
- Cnubben, W. & Delespaul, Ph. (2017). Benchmark ROM: de verkeerde administratie. *De Psycholoog*, 52, 51-53.
- Cramer, A. (2017). Hoe verder met benchmarking in de ggz: pleidooi voor betere methodologie en meer slow science. *De Psycholoog*, 52, 54-55.
- Crutzen, R. (2014). Time is a jailer: what do alpha and its alternatives tell us about reliability? *The European Health Psychologist*, 16, 70-74.
- Gondek, D., Edbrooke-Childs, J., Fink, E., Deighton, J., & Wolpert, M. (2016). Feedback from outcome measures and treatment effectiveness, treatment efficiency, and collaborative practice: a systematic review. *Administration and Policy in Mental Health and Mental Health Services Research*, 43, 325-343.
- Hafkenscheid, A. (1994). *Rating scales in treatment efficacy studies: individualized and normative use*. Academisch proefschrift, Rijksuniversiteit Groningen.
- Hafkenscheid, A. (2007). De Brief Symptom Inventory: reactie op het gelijknamige artikel van E. de Beurs & F. Zitman (MGv 06-2). *Maandblad Geestelijke Volksgezondheid*, 62, 819-821.
- Hafkenscheid, A. (2008). *Routine Process Monitoring: ervaringen uit de praktijk*. *Tijdschrift Cliëntgerichte Psychotherapie*, 46, 327-345.
- Hafkenscheid, A. (2009). *Routine Process Monitoring (RPM) in partnerrelatiebehandelingen*. *Directie Therapie*, 29, 5-25.
- Hafkenscheid, A. (2010). Rammelende ROM in de ggz: geen ROM zonder Routine Process Monitoring. *GZ-Psychologie*, 2, 12-17.
- Hafkenscheid, A. (2012). Subjectiviteit bij de interpretatie van het grafische scoreverloop op monitorinstrumenten. *Tijdschrift voor Psychiatrie*, 54, 29-134.
- Hafkenscheid, A. (2013). Geen rad voor de ogen: reactie op De Jong & Van 't Spijker. *Tijdschrift voor Psychotherapie*, 39, 203-207.
- Hafkenscheid, A. (2016). Waarom stagneert deze therapie? Een zoekschema voor het opsporen van mogelijke obstakels. *Tijdschrift voor Psychotherapie*, 42, 291-307.
- Hafkenscheid, A. (2017). Behandelplannen voor patiënten met een psychose: In hoeverre schieten deze plannen tekort? *Nederlands Tijdschrift voor de Geneeskunde*, 161, 10-11.
- Hafkenscheid, A.J., Maassen, G.H. & Veeninga, A.T. (2007). The dimensions of the Dutch SCL-90: more than one, but how many? *Netherlands Journal of Psychology*, 63, 29-35.
- Hafkenscheid, A. & Van Os, J. (2013). Huidige ROM doet afbreuk aan valide kwaliteitsmeting. *Tijdschrift voor Psychiatrie*, 55, 179-181.
- Hafkenscheid, A. & Van Os, J. (2014a). Naar een deugdelijke ROM. *MGv*, 69, 20-28.
- Hafkenscheid, A. & Van Os (2014b). ROM van geïndividualiseerde behandeldoelen. *PsychoPraktijk*, 6, 29-32.
- Hafkenscheid, A. & Van Os, J. (2016). Wat ieder die betrokken is bij ROM zich over de metingen moet realiseren. *Tijdschrift voor Psychiatrie*, 58, 388-396.
- Janse, P.D., De Jong, K., Van Dijk, M.K., Hutschemaekers G.J.M., & Verbraak, M. J.P.M. (2016). Improving the efficiency of cognitive-behavioural therapy by using formal client feedback. *Psychotherapy Research*, DOI: 10.1080/10503307.2016.1152408.
- De Jong, K. (2012). *A chance for change: building an outcome monitoring feedback system for outpatient mental health care*. Academisch proefschrift Universiteit Leiden.
- 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 and Psychotherapy*, 14, 288-301.
- De Jong, K., Nugter, M.A., Polak, M.G., Wagenborg, J.E.A., Spinhoven, Ph., & Heiser, W.J. (2008). De Nederlandse versie van de Outcome Questionnaire (OQ-45): een crossculturele validatie. *Psychologie & Gezondheid*, 36, 35-45.
- De Jong, K., Tiemens, B., & Verbraak, M. J.P.M. (2017). Red ROM als kwaliteitsinstrument. *Tijdschrift voor Psychiatrie*, 59, 242-245.
- Koekkoek, B. (2017). Reacties, 3 april 2017, <https://www.medischcontact.nl/nieuws/laatste-nieuws/artikel/protesterende-psychiaters-maken-karikatuur-van-rom.htm>
- Lambert, M. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, 17, 1-14.
- Lambert, M.J. & Shimokawa (2011). Collecting client feedback. In: J.C. Norcross (ed.). *Psychotherapy relationships that work: evidence-based responsiveness*. New York: Oxford University Press, pp. 203-223.
- Mulder, C., Staring, A., Loos, J., Buwalda, V., Kuijpers, D., Sytema, S., & Wierdsma, A.I. (2004). De Health of the Nation Outcome Scales (HONOS) als instrument voor 'routine outcome assessment'. *Tijdschrift voor Psychiatrie*, 46, 273-284.
- Nugter, M.A., Buwalda, V.J.A., Dhondt, A.D.F., Draisma, S. (2012). Gebruik van de HONOS bij de behandeling van patiënten. *Tijdschrift voor Psychiatrie*, 54, 153-159.
- Van Oenen, F.J. (2017). *Feedback Informed Treatment in emergency psychiatry: a poor FIT?* Academisch proefschrift Vrije Universiteit Amsterdam.
- Van Os, J., Berkelaar, J., Hafkenscheid, A. (2017). Benchmarken: doodlopende weg onder het mom van 'ROM'. *Tijdschrift voor Psychiatrie*, 59, 247-250.
- Van Os, J. & Delespaul, (in druk). Een valide kwaliteitskader voor de GGZ: van Benchmark-ROM aan de achterkant naar regionale regie en creatie aan de voorkant. *Tijdschrift voor Psychiatrie*, 59.
- Van Os, J., Kahn, R., Denys, D., Schoevers, RA, Beekman, A.T.F., Hoogendijk, W.J.G., van Hemert, A.M., Hodiament, P.P.G., Scheepers, F., Delespaul, A.E.G., Leentjens, A.F.G. (2012). ROM: Gedragsnorm of Dwangmaatregel? Commentaar op het Themanummer ROM. *Tijdschrift voor Psychiatrie*, 54, 245-253.
- Schoevers, R.A. & Beekman, A.T.F. (2017). Van 'stop ROM naar 'hier met die ROM!' *Tijdschrift voor Psychiatrie*, 59, 245-246.
- Smits, D. (2016). *The effect of monitoring on the attunement process in psychotherapy: a mixed-method approach*. Academisch proefschrift Katholieke Universiteit Leuven.
- Westert, G. (2017). Protesterende psychiaters maken een karikatuur van ROM: behandeluitkomst is juist uitstekende indicator van kwaliteit. *Medisch Contact*, <https://www.medischcontact.nl/nieuws/laatste-nieuws/artikel/protesterende-psychiaters-maken-karikatuur-van-rom.htm>
- Winkelhorst, Y., Hafkenscheid, A., & De Groot, E. (2013). Verhoogt Routine Process Monitoring (RPM) de effectiviteit van behandeling? Een vergelijkende pilot study. *Tijdschrift voor Psychotherapie*, 39, 146-156.