

.....

Psychometrische evaluatie van de Mental Health Continuum – Short Form (MHC-SF)

Constructvaliditeit, responsiviteit voor
verandering, normen en T-scores

EDWIN DE BEURS, SJOERD KOSTERMAN, SUZANNE ANTEN,
ERNST BOHLMEIJER & GERBEN WESTERHOF

Samenvatting

.....

De Mental Health Continuum – Short Form (MHC-SF) is ontwikkeld om positieve geestelijke gezondheid te meten. Hieronder wordt verstaan emotioneel, sociaal en psychisch welbevinden. Doel van deze studie was om de betrouwbaarheid en validiteit van de MHC-SF te onderzoeken. De constructvaliditeit en concurrente validiteit kwamen hierbij aan bod. Ook onderzochten we de responsiviteit van de MHC-SF door verandering op de MHC-SF te vergelijken met verandering op de schaal voor Symptomatic Distress van de Outcome Questionnaire (OQ-SD). Ten slotte leveren we normgegevens en T-scores voor de MHC-SF. De psychometrische kenmerken van de MHC-SF werden onderzocht met gegevens van een bevolkingssteekproef ($n = 1802$) en een steekproef van cliënten die werden behandeld in de specialistische ggz ($n = 971$). De resultaten ondersteunden de eerder gevonden factorstructuur met een algemene factor voor positieve geestelijke gezondheid, en drie subschaalscores voor emotioneel, sociaal en psychologisch welbevinden. De responsiviteit van de MHC-SF was vergelijkbaar met de OQ-SD, alleen de verandering in score

op sociaal welbevinden bleef significant achter bij de OQ-SD. Hiermee vormt de MHC-SF een zinnige toevoeging aan een ROM-set, maar we mogen niet geheel dezelfde veranderingen op de MHC-SF verwachten als met de OQ-SD-schaal gemeten worden. Met gestandaardiseerde T-scores kunnen meetresultaten eenvoudiger geïnterpreteerd worden: de T-score geeft direct weer hoe ver een testuitkomst verwijderd is van het gemiddelde en daarmee hoe uitzonderlijk de testuitkomst is. In vergelijking met traditionele normtabellen kan met een T-score op een verfijndere manier gerapporteerd worden over de uitkomst van een test. T-scores vergemakkelijken zo de communicatie over de meetresultaten met cliënten of collega's.

Trefwoorden: positieve psychologie, welzijn, psychometrie, responsiviteit, T-scores

Kernboodschappen voor de klinische praktijk

- ▶ Het is de moeite waard om naast symptomen of klachten ook positieve geestelijke gezondheid te meten om goed zicht te krijgen op de voortgang in de behandeling.
- ▶ De MHC-SF en de OQ-SD zijn vergelijkbaar responsief, met uitzondering van de schaal voor sociaal welbevinden, die minder gevoelig voor verandering blijkt.
- ▶ De testuitslag weergegeven in genormaliseerde T scores maakt het mogelijk om schaalcores onderling direct te vergelijken en vergemakkelijkt de interpretatie en de communicatie met de client.

INLEIDING

.....

Van oudsher is gezondheid gedefinieerd als het ontbreken van ziekte. Door de Wereldgezondheidsorganisatie werd echter al in 1948 gepleit voor een bredere visie op gezondheid. Meer recent onderscheidt Machteld Huber de volgende zes dimensies van positieve gezondheid: (1) lichaamsfuncties, (2) mentaal welbevinden, (3) sociaal-maatschappelijk functioneren, (4) dagelijks functioneren, (5) kwaliteit van leven, en (6) zingeving (Huber et al., 2011, 2016). Al met al bestaat een goede gezondheid uit meer dan niet lijden aan een lichamelijke ziekte, en wat geldt voor de algemene gezondheid, geldt ook voor de *geestelijke* gezondheid. Ook psychische gezondheid werd aanvankelijk vooral afgemeten aan de mate waarin men vrij was van psychische klachten of problemen. De evaluatie van de psychiatrische gezondheidszorg is traditioneel dan ook gericht op het meten van psychiatrische symptomen: het succes van een behandeling wordt afgemeten aan de afname van psychopathologie op generieke of aandoeningsspecifieke klachtenlijsten.

Twee ontwikkelingen hebben hier verandering in gebracht: de herstelbeweging in de psychiatrie (Anthony, 1993) en de positieve psychologie (Seligman & Csikszentmihalyi, 2000). Door beide wordt gepleit voor een bredere rolopvatting van de ggz. In plaats van een nadruk op het voorkomen, tegengaan en behandelen van gebreken en symptomen die voortkomen uit een psychiatrische ziekte (oftewel een tekort aan geestelijke gezondheid), zou de ggz zich ook moeten richten op het bevorderen en verbeteren van het (geestelijk) welbevinden, in het Engels vaak aangeduid met de term *well-being*. In lijn hiermee is door de Wereldgezondheidsorganisatie (WHO; Herrman et al., 2005) geestelijke gezondheid gedefinieerd als ‘a state of well-being in which every individual realises his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community’ (p. 12). Volgens deze definitie is geestelijke gezondheid dus niet de afwezigheid van psychiatrische ziekten, maar te beschouwen als een toestand van welbevinden, die aspecten omvat als tevredenheid met het leven dat men leidt, het ervaren van positieve emoties, het realiseren van persoonlijke doelstellingen en zingeving, en het naar tevredenheid functioneren op diverse levensgebieden, zowel in sociale relaties als op werk of opleiding.

Voor het meten van positieve geestelijke gezondheid is door Keyes (2002) een meetinstrument ontwikkeld, de Mental Health Continuum. Aanvankelijk was dat een vragenlijst met 40 items – de Long Form-versie of MHC-LF –, die later werd verkort tot een vragenlijst met 14 items – de Short Form-versie of MHC-SF. Gebaseerd op uitgebreide literatuurstudies over subjectief welbevinden (Diener, 1984), optimaal psychologisch functioneren (Ryff, 1989) en optimaal maatschappelijk functioneren (Keyes, 1998), worden door Keyes en collega’s (Keyes, 2002, 2005; Keyes & Martin, 2017) drie aspecten onderscheiden: (1) emotioneel, (2) sociaal en (3) psychologisch welbevinden. Deze drie aspecten zijn met confirmatieve factoranalyse teruggevonden in de MHC-SF. De MHC-SF is in het Nederlands vertaald en de psychometrische eigenschappen zijn onderzocht door Franken en collega’s (2018), en door Lamers en collega’s (2011, 2012).

Om de vraag te beantwoorden of de MHC-SF echt iets anders meet dan het ontbreken van geestelijke gezondheid, zijn scores op de lijst vergeleken met meetinstrumenten voor psychopathologie. Volgens het twee-continuummodel (Tudor, 1996) zijn positieve geestelijke gezondheid zoals gemeten met de MHC-SF, en psychopathologie zoals gemeten met de Brief Symptom Inventory (BSI; de Beurs & Zitman, 2006) twee gerelateerde maar afzonderlijke factoren, die het beste als twee aparte dimensies kunnen worden weergegeven. De uitkomst van confirmatieve factoranalyse van MHC-SF- en BSI-schaalscores bij een representatieve steekproef uit de Nederlandse bevolking ondersteunde dit model: resultaten lieten een betere fit zien bij een model met twee gerelateerde factoren dan bij een model met een enkele

factor (Lamers et al., 2011). Ook bij patiënten is ondersteuning gevonden voor het twee-continuamodel (de Vos et al., 2018; Keyes & Martin, 2017). Hiermee werd het onderscheid tussen positieve geestelijke gezondheid en symptomen van psychopathologie bevestigd, en het belang onderstreept van het meten van positieve geestelijke gezondheid in aanvulling op het meten van psychopathologie.

Recent hebben Franken en collega's (2018) en de Vos en collega's (2018) de factorstructuur onderzocht van de MHC-SF in een klinische populatie, nu in vergelijking met de Outcome Questionnaire (OQ-45; Lambert et al., 2004). Opnieuw vonden zij steun voor het twee-continuamodel van positieve geestelijke gezondheid en psychopathologie. Een andere Nederlandse onderzoeksgroep (van Erp Taalman Kip & Hutschemaekers, 2018) vond daarentegen in een patiëntensteekproef hoge correlaties tussen de drie factoren van de MHC-SF en geringe discriminante validiteit van de MHC-SF ten opzichte van de OQ-SD-schaal. Daarmee lijkt het laatste woord over het twee-continuamodel nog niet gezegd en is het de moeite waard de validiteit van dit model verder te onderzoeken.

Het eerste doel van het huidige onderzoek (studie I) is om de psychometrische eigenschappen van de MHC-SF bij gebruik in de specialistische ggz te onderzoeken. We zullen de constructvaliditeit van de MHC-SF-schalen onderzoeken met confirmerende factoranalyse. Daarnaast onderzoeken we de betrouwbaarheid van de totale schaal, en van de subschalen voor emotioneel, psychologisch en sociaal welbevinden. Voor nader onderzoek naar het twee-continuamodel vergelijken we scores op de MHC-SF met een meetinstrument dat in de ggz traditioneel gebruikt wordt om de uitkomst van behandelingen te monitoren: de Symptomatic Distress schaal (OQ-SD) van de Outcome Questionnaire (OQ-45; Lambert et al., 1996). De OQ-45 wordt in Nederland vaak gebruikt voor *routine outcome monitoring* (ROM; de Beurs et al., 2011) bij veelvoorkomende aandoeningen, zoals depressie en angst. De OQ-45 bestaat uit 25 vragen over symptomen en klachten (Symptomatische Distress – OQ-SD), 11 vragen over functioneren in relatie tot familie, vrienden en kennissen (interpersoonlijke relaties), en 9 vragen over maatschappelijk functioneren op werk of op school (sociale rol). De OQ-SD-schaal heeft de beste psychometrische kenmerken (de Beurs et al., 2005; de Jong et al., 2007; Lambert et al., 1996): deze schaal is het betrouwbaarst en het meest gevoelig voor verandering. We zullen onderzoeken hoe goed de OQ-SD en de MHC-SF verandering tijdens de behandeling detecteren. Dit validiteitsaspect van meetinstrumenten wordt ook wel de *responsiviteit* genoemd (Mokkink et al., 2010). De twee instrumenten zijn herhaald en op hetzelfde moment afgenomen, waardoor de onderlinge samenhang van de twee instrumenten kan worden vastgesteld en het verschil in scores over de tijd direct kan worden vergeleken (de Beurs et al., 2012, 2020; de Beurs, Visser et al., 2019; Schawo et al., 2019).

Het tweede doel van het onderzoek komt in studie II aan bod en betreft het normeren van de Nederlandse MHC-SF. Gewoonlijk geeft men betekenis aan een meetresultaat op een test door de score te vergelijken met normgroepen. In een normeringstabel wordt weergegeven welke range van scores 'zeer hoog' is, welke 'hoog', welke 'bovengemiddeld', enzovoort. Zo'n normering gebeurt ten opzichte van een vergelijkingsgroep, zoals de algemene bevolking of cliënten in de ggz. Hoewel ruwe scores zo betekenis krijgen, houdt dit wel in dat men de score van een continue schaal omzet naar categorieën, waarbij vanzelfsprekend informatie verloren gaat. Er is een handleiding voor de Nederlandse versie van de MHC-SF gepubliceerd met deze normgegevens voor de algemene bevolking en diverse klinische groepen (Franken et al., 2019).

De laatste jaren wordt een alternatieve manier van betekenis geven aan scores toegepast, waarbij de scores naar een *common metric* of uniforme meetschaal worden vertaald (zie voor voorbeelden: Choi et al., 2014; Cook et al., 2015; Schalet et al., 2014, 2015; Wahl et al., 2014). Deze meer verfijnde manier van betekenis geven aan scores, waarbij alle informatie van de ruwe scores behouden blijft, bestaat uit het omzetten van ruwe scores naar T-scores (de Beurs, 2010). T-scores zijn gestandaardiseerde scores met een gemiddelde van 50 en een spreiding van 10. Hiermee geeft een T-score de afstand weer tot de gemiddelde score in standaardeenheden en kan uit de score direct worden afgelezen hoe uitzonderlijk een testresultaat is. T-scores hebben een directe relatie met percentielscores volgens de cumulatieve normaalverdeling (de Beurs, Flens et al., 2019). Voor elke T-score is zo te bepalen welke proportie van de normpopulatie dezelfde of een hogere score heeft. Bijvoorbeeld: $T = 70$ is twee standaarddeviaties verwijderd van het gemiddelde van 50 van de normgroep; zo'n score of hoger heeft slechts 2,5% van de normgroep (zie figuur D in de online bijlage op www.tijdschriftgedragstherapie.nl). Wanneer we er bovendien voor zorgen dat de T-scores normaal verdeeld zijn, worden de ruwe schaalscores omgezet naar een schaal met gelijke intervallen, waarbij rekenkundige bewerkingen als aftrekken van scores geoorloofd zijn (om bijvoorbeeld een pre-posttest-verschilscore te bepalen als maat voor verandering tijdens behandeling). Elders hebben we ervoor gepleit hierbij voortaan de algemene bevolking als vergelijkingsgroep te nemen (de Beurs, Flens et al., 2019). Dit komt overeen met internationale ontwikkelingen om de scoring van meetinstrumenten meer uniform te maken, zoals het Prosetta Stone initiatief (www.prosetta.org) en PROMIS (Cella et al., 2007). De algemene bevolking krijgt dan een gemiddelde score van $M = 50$ ($SD = 10$) en de conventie is om met de T-score de hoogte of intensiteit van het gemeten construct uit te drukken: een hoge score duidt op veel welbevinden (MHC-SF) of veel klachten (OQ-SD). Cliënten scoren dan bij aanvang van de behandeling, al naargelang hoe ernstig het welbevinden is afgenomen volgens de MHC-SF, $T = 25-35$. Of, afhankelijk van de ernst van hun klachten of problemen, scoren cliënten volgens de OQ-45

T = 65-75. Het zal even wennen zijn, maar uiteindelijk kunnen we bij klinisch en neurologisch onderzoek, en voor ROM, veel baat hebben bij het gebruik van een uniforme meetschaal als de T-score (de Beurs, Flens et al., 2019).

Het tweede doel van dit onderzoek is om klassieke normtabellen op te stellen en T-scores te bepalen volgens een Item Respons Theorie benadering. We stellen ook omzettingsformules op van ruwe scores naar T-scores, en onderzoeken of aparte normen nodig zijn voor mannen en vrouwen, en voor leeftijdsgroepen.

METHODE

.....

Respondenten

.....

Voor dit onderzoek werden twee datasets gebruikt. Ten eerste data van het LISS-panel (N = 1802) die eerder werden geanalyseerd (Lamers et al., 2011, 2012). Deze dataset betrof een representatieve steekproef uit de Nederlandse bevolking die gedurende meerdere jaren via het internet werd benaderd met zelfrapportagevragenlijsten. De gemiddelde leeftijd was $M = 47,2$ ($SD = 17,7$); de steekproef omvatte 1802 respondenten, van wie zich 913 (50,7%) als vrouwelijk hebben geïdentificeerd. Ten tweede werden gegevens van een klinische groep (N = 971) gebruikt. Deze data komen van Arkin, een ggz-instelling in Amsterdam, en werden verzameld bij PuntP, een kliniek voor specialistische ggz (waar cliënten met veelvoorkomende psychiatrische klachten zoals depressie en angst worden behandeld). De klinische data zijn niet eerder geanalyseerd. Bij PuntP wordt de MHC-SF gebruikt als een uitkomstmaat voor ROM, naast de OQ-45 als maat voor de afname van psychische problemen.

Deze gegevens werden verzameld in het kader van behandeling en kwaliteitscontrole. Volgens de Nederlandse wetgeving is geanonimiseerde analyse van deze gegevens ten behoeve van wetenschappelijk onderzoek toegestaan. De gemiddelde leeftijd in deze steekproef was $M = 36,8$ ($SD = 12,8$) en de meerderheid van de cliënten ($n = 606$; 62,4%) heeft zich als vrouwelijk geïdentificeerd. De meting werd tijdens de behandeling herhaald. Van 392 cliënten zijn gegevens gebruikt van de eerste hertest na een interval van gemiddeld $M = 133$ ($SD = 57$) dagen of 4,5 maanden.

Instrumenten

.....

MHC-SF — De MHC-SF werd ontwikkeld door Keyes, aanvankelijk als een vragenlijst met 40 items (MHC-LF), die werd verkort tot een versie, de MHC-SF, met 14 items (Keyes, 2002). Deze verkorte lijst is in het Neder-

lands vertaald door Lamers en collega's (2011). De lijst omvat drie items voor 'emotioneel welbevinden' (MHC-EWB; item 1-3; geluk, belangstelling en tevredenheid), vijf items voor 'sociaal welbevinden' (MHC-SWB; item 4-8; sociale betrokkenheid, geborgenheid, verwezenlijking, acceptatie en saamhorigheid) en zes items voor 'psychologisch welbevinden' (MHC-PWB; item 9-14; acceptatie van zichzelf, voldoende gevoel van beheersing van je omgeving, positieve interpersoonlijke relaties, persoonlijke groei, autonomie en het leven als zinvol ervaren). Vragen worden beantwoord op een zespunts-likertschaal voor frequentie ('In de afgelopen maand, hoe vaak had u het gevoel ... [dat u gelukkig was]'; 0 = nooit, tot 5 = elke dag). De vragenlijst is opgenomen in de online bijlage bij dit artikel.

OQ-45 — De OQ-45 is ontwikkeld door Lambert en collega's om te gebruiken als generiek instrument voor het meten van de uitkomst van psychologische of psychiatrische behandeling (Lambert et al., 2004). Het instrument meet de ernst van symptomen van psychopathologie (Symptomatic Distress, SD), het interpersoonlijk functioneren met familie en vrienden (Interpersoonlijke Relations, IR) en het maatschappelijk functioneren in werk of opleiding (Social Role, SR). De instructie luidt: 'Help ons begrijpen hoe u zich de afgelopen week, tot en met vandaag, hebt gevoeld (voorbeelditem: 'Ik was angstig.')

De items beschrijven gevoelens, klachten of problemen, en worden beantwoord op een likertschaal voor frequentie in de afgelopen week, die loopt van 0 = nooit tot 4 = bijna altijd. De Nederlandse vertaling is onderzocht door de Beurs en collega's (2005) en de Jong en collega's (2007). Recent zijn er door Timman en collega's (2017) normgegevens voor de algemene bevolking en diverse patiëntengroepen gepubliceerd. Op basis van deze gegevens zijn door de Beurs en collega's (submitted) T-scores voor de OQ-schalen berekend. Voor dit onderzoek beperken we ons tot de OQ-SD-schaal. De andere schalen werden niet afgenomen bij PuntP.

Statistische analyse

Studie I — De betrouwbaarheid van de MHC-SF-schalen werd afgemeten aan de interne consistentie (Cronbachs alfa). De constructvaliditeit werd onderzocht met confirmerende factoranalyse (het pakket lavaan in R), waarbij de fit van verschillende modellen werd vergeleken: een single factor model, een 3-factor model en een bi-factor model (drie factoren en een algemene factor). Als criteria voor acceptabele fit gelden: RMSEA \leq 0,08, SRMR \leq 0,08, CFI \geq 0,90, TLI \geq 0,90 en een χ^2/df ratio \leq 5; goede fit is: RMSEA \leq 0,06 (90% CI \leq 0,06), SRMR \leq 0,05, CFI \geq 0,95, TLI \geq 0,95 en een $\chi^2/df \leq$ 3 (Hu & Bentler, 1999).

In de klinische dataset werd de constructvaliditeit onderzocht door scores van cliënten en de algemene bevolking te vergelijken (concurrente vali-

diteit). De correlatie tussen de MHC-SF-schalen en de gelijktijdig ingevulde OQ-SD voor voormetingsgegevens en voor test-hertest-verschilscores werd vastgesteld om de convergente validiteit te onderzoeken. De responsiviteit van de MHC-SF werd onderzocht door scores van herhaalde metingen met de MHC-SF (totaalscore en schaalscores) te vergelijken met de OQ-SD met variantieanalyse voor herhaalde metingen. Voor beide instrumenten drukken we het test-hertest-verschil uit in Cohen's d ($M_{\text{pretest}} - M_{\text{retest}} / SD_{\text{pooled}}$).

Studie II — Ter normering van de MHC-SF hebben we sekse- en leeftijdsverschillen onderzocht, en op basis van de frequentieverdeling van scores op de MHC-SF normtabellen opgesteld voor de algemene bevolking en voor cliënten van de specialistische ggz met veelvoorkomende stoornissen. Waar de resultaten er aanleiding toe gaven, zijn er aparte normen opgesteld voor mannen en voor vrouwen in vier leeftijdsklassen (populatie: 18-29, 30-49, 50-64, 65+).

We berekenden ook T-scores. Ruwe scores van een bevolkingssteekproef en van een klinisch sample werden geanalyseerd met R (mirt) om factor-scores (theta's) te bepalen (Chalmers, 2012). Dit zijn *expected-a-posteriori* (EAP-)scores, die berusten op zowel de combinatie van de antwoordoptie die de respondent heeft gekozen (de itemscores), als kenmerken van het item (itemparameters voor de zwaarte van het item). De itemparameters werden voor de bevolking gefixeerd op een latent gemiddelde $M = 0$ en $SD = 1$. De theta's komen dan in de vorm van standaardcores of Z-scores. Vermenigvuldigen van theta's met 10 en er 50 bij optellen geeft T-scores. We presenteren een oversteektabel en oversteekfiguur voor ruwe scores naar T-scores en formules om T-scores uit ruwe somscores te berekenen. Hiertoe is voor elke mogelijke ruwe schaalscore de gemiddelde T-score in de totale onderzoeksgroep bepaald. Vervolgens is met *curve fitting* (non-lineaire regressieanalyse in R) de beste vergelijking gezocht voor het beschrijven van de relatie tussen ruwe scores en T-scores. Wanneer er verschillen zijn in gemiddelde score tussen mannen en vrouwen, of tussen leeftijdsgroepen, is het zaak deze formules ook apart te bepalen voor sekse en leeftijdsgroepen. De gevonden formules geven uiteindelijk op basis van ruwe schaalscores 'berekende T-scores' die de *theta-based* T-scores zo goed mogelijk benaderen. Met de formule kan de ruwe schaalscore van een individuele respondent worden omgezet naar een T-score zonder gebruik te hoeven maken van een grote dataset en IRT-software. We inspecteerden de overeenstemming tussen theta-based T-scores en berekende T-scores visueel met Bland-Altman plots en door de Intraclass Correlatie coëfficiënt (ICC) tussen beide T-scores vast te stellen.

RESULTATEN

Studie I: Betrouwbaarheid

In tabel 1 zijn de betrouwbaarheid (interne consistentie) en de inter-item-correlaties voor de MHC-SF-schalen voor de klinische steekproef van PuntP weergegeven. Tevens zijn de correlatiecoëfficiënten (product moment correlatie r) weergegeven tussen MHC-SF en OQ-SD bij de pretest, de her-test en voor het pre-hertestverschil. Alle r 's wijken statistisch significant af van 0 ($p < 0,05$) en zijn negatief (veel klachten gaat samen met minder welbevinden).

TABEL 1 *Interne consistentie, gemiddelde inter-itemcorrelatie en range en samenhang (correlatie) met OQ-SD bij voormeting, her-test en voor het test-hertestverschil*

Schaal	MHC-SF Totaal	EWB	SWB	PWB
Aantal items	14	3	5	6
Cronbachs alfa	0,91	0,85	0,76	0,83
r_{ii}	0,41	0,66	0,38	0,45
r_{range}	0,21-0,68	0,63-0,68	0,26-0,52	0,35-0,53
Correlatie met de OQ-SD:				
Pretest ($n = 971$)	-0,65	-0,65	-0,50	-0,60
Hertest ($n = 391$)	-0,71	-0,69	-0,58	-0,69
Test-hertestverschil ($n = 391$)	-0,68	-0,65	-0,48	-0,63

Noot. EWB = Emotioneel Welbevinden; MHC-SF Totaal = positieve geestelijke gezondheid; OQ-SD = Symptomatic Distress schaal van de OQ-45; PWB = Psychologisch Welbevinden; SWB = Sociaal Welbevinden.

De interne consistentie van de subschalen was goed; alleen de schaal voor Sociaal Welbevinden bleef iets achter bij de rest. Het patroon van correlaties van de MHC-SF met de OQ-SD was conform de meetpretentie van de MHC-SF (convergente validiteit): hoge negatieve correlaties voor de totaalscore, emotioneel welbevinden en psychologisch welbevinden; iets lagere samenhang met sociaal welbevinden.

Concurrente validiteit

140

In tabel 2 staan gemiddelde en spreiding van scores op de MHC-SF van de algemene bevolking en van cliënten in behandeling bij de specialistische ggz. De gemiddelde score verschilde significant voor alle schaalcores van de MHC-SF volgens een *t*-test voor onafhankelijke steekproeven. Scores van cliënten waren aanzienlijk lager dan van de algemene bevolking, een bevinding die de concurrente validiteit van de MHC-SF verder ondersteunt.

TABEL 2 *Vergelijking van scores tussen de algemene bevolking en cliënten, van mannen en vrouwen, en van twee leeftijdsgroepen*

Steekproef	Bevolking (n = 1802)		Cliënten (n = 971)		t(2771)	Cohens d
	M	SD	M	SD		
MHC_TOT	2,98	0,85	1,78	0,96	33,60***	1,86
EWB	3,66	0,96	1,79	1,14	45,61***	2,50
SWB	2,32	1,01	1,58	1,02	18,53***	1,04
PWB	3,18	1,00	1,95	1,09	29,90***	1,66
Geslacht	Man (n = 1254)		Vrouw (n = 1519)		t(2771)	Cohens d
	M	SD	M	SD		
MHC_TOT	2,62	1,02	2,51	1,09	2,81**	0,15
EWB	3,08	1,33	2,94	1,38	2,85**	0,15
SWB	2,13	1,03	2,00	1,11	3,13**	0,17
PWB	2,80	1,16	2,71	1,21	1,86	0,10
Leeftijd	≤ 40 (n = 1339)		≥ 41 (n = 1434)		t(2771)	Cohens d
	M	SD	M	SD		
MHC_TOT	2,47	1,06	2,64	1,05	4,41***	0,24
EWB	2,80	1,35	3,19	1,34	7,63***	0,41
SWB	1,98	1,05	2,14	1,09	3,78***	0,20
PWB	2,71	1,20	2,79	1,18	1,98*	0,11

Noot. EWB = Emotioneel Welbevinden; MHC_TOT = positieve geestelijke gezondheid; PWB = Psychologisch Welbevinden; SWB = Sociaal Welbevinden.

* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Constructvaliditeit

141

De confirmerende factoranalyse van MHC-SF-schalen wees uit dat een bi-factor model met drie factoren (emotioneel, sociaal en psychologisch welbevinden) en een algemene factor (positieve geestelijke gezondheid) het beste bij de data paste en aan de meeste criteria voor goede fit voldeed [$\chi^2(57) = 278,8$ CFI = 0,96; TLI = 0,94; RMSEA = 0,06 (CI = 0,06-0,07); SRMR = 0,03]. Dit model had een betere fit dan een 3-factor model [$\chi^2(74) = 492,3$ CFI = 0,93; TLI = 0,91; RMSEA = 0,08 (CI = 0,07-0,08); SRMR = 0,04] en een 1-factor model [$\chi^2(77) = 793,0$ CFI = 0,88; TLI = 0,86; RMSEA = 0,10 (CI = 0,09-0,10); SRMR = 0,06]. De constructvaliditeit van de MHC-SF met drie factoren en een algemene factor is hiermee ondersteund.

Responsiviteit

Variantieanalyse voor herhaalde metingen gaf hoofdeffecten voor tijd en instrument, en een interactie-effect voor instrument bij tijd. Vooral het laatste is belangrijk: een significant interactie-effect impliceert dat er verschil is tussen schaalscores op beide instrumenten in verandering over de tijd, oftewel verschil in responsiviteit. De variantieanalyse liet een significant interactie-effect zien [$F(4,1560) = 2,75, p = 0,027, \eta_p^2 = 0,007$], wat aangeeft dat er verschil was in responsiviteit tussen de schalen. Geplande paarsgewijze vergelijkingen tussen de OQ-SD en MHC-SF totaalscore en subschaalscores wezen uit dat dit vooral is toe te schrijven aan grotere responsiviteit van de OQ-SD ten opzichte van de Sociaal Welbevinden schaal [$F(1,390) = 5,15; p = 0,024; \eta_p^2 = 0,013$; de andere contrasten $p > 0,086, \eta_p^2 < 0,008$]. Om de grootte van de verandering over de tijd volgens de OQ-SD en de MHC-SF te vergelijken werden de scores gestandaardiseerd op het gemiddelde en de pooled SD van voor- en nameting (zie tabel 3). De gemiddelde *afname* in score op de OQ-SD in standaardeenheden was $d = 0,52$, de *toename* in score op de MHC-SF schalen was iets geringer (op zijn hoogst $d = 0,41$ voor psychologisch welbevinden, voor sociaal welbevinden $d = 0,29$).

Samenvattend was er alleen een significant verschil in responsiviteit tussen de SWB-schaal en de OQ-SD-schaal. Over het geheel genomen verschillen de MHC-SF-schalen en de OQ-SD-schaal niet veel in responsiviteit.

TABEL 3 *Pretest- en hertestgemiddelden en gestandaardiseerd verschil bij herhaald gemeten cliënten (N = 391)*

142

	Pretest		Hertest		ES
	M	SD	M	SD	
OQ-SD	55,73	14,72	48,06	16,94	0,52
MHC-SF Totaal	1,84	0,96	2,22	1,07	0,40
EWB	1,88	1,16	2,27	1,21	0,34
SWB	1,63	1,02	1,92	1,13	0,29
PWB	2,00	1,08	2,45	1,16	0,42

Noot. ES = $M_{\text{pretest}} - M_{\text{herstest}} / SD_{\text{pretest}}$; EWB = Emotioneel Welbevinden; MHC-SF Totaal = positieve geestelijke gezondheid; OQ-SD = Symptomatic Distress schaal van de OQ-45; PWB = Psychologisch Welbevinden; SWB = Sociaal Welbevinden.

Studie II: Normering

De gemiddelde score op de MHC-SF van mannen en vrouwen uit de bevolking en mannelijke en vrouwelijke cliënten bij voormeting verschilde: mannen scoorden hoger dan vrouwen (zie tabel 2). Ook was er een effect van leeftijd op de MHC-SF-score: met het toenemen van de leeftijd nam de score op de MHC-SF toe, maar een uitzondering vormde de jongste leeftijdsgroep. We vinden dit effect zowel bij de algemene bevolking als bij de cliënten van de specialistische ggz (zie figuur A in de online bijlage). Dit rechtvaardigt aparte normen voor mannen en vrouwen, en voor leeftijdsgroepen.

Tabel 4 biedt een klassieke normeringstabel met ranges van scores, en hun betekenis voor de respondenten uit de bevolkingssteekproef en voor cliënten in de gespecialiseerde ggz, zowel ongeacht geslacht, als apart voor mannen en vrouwen. De tabel geeft de normwaarden voor ruwe scores op de MHC-SF-totaalschaal, en op de schalen voor emotioneel (EWB), sociaal (SWB) en psychologisch (PWB) welbevinden in zeven niveaus. Normtabel- len voor leeftijdsgroepen zijn te verkrijgen bij de eerste auteur.

TABEL 4 Normtabel MHC-SF voor 7 niveaus (5%, 15%, 20%, 20%, 20%, 15%, 5% van de populatie)

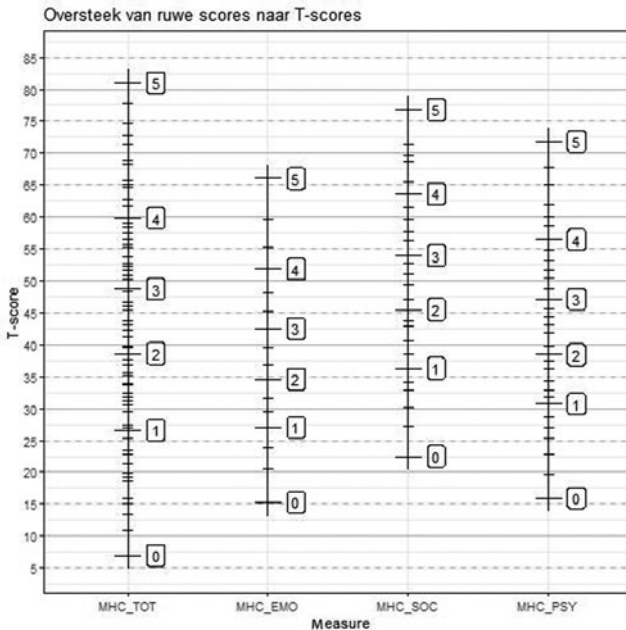
Bevolking	Allen (N = 1662)				Mannen (N = 828)				Vrouwen (N = 834)			
	Totaal	EWB	SWB	PWB	Totaal	EWB	PWB	SWB	Totaal	EWB	PWB	SWB
Zeer laag	0,00- 1,42	0,00- 1,99	0,00- 0,59	0,00- 1,32	0,00- 1,42	0,00- 1,66	0,00- 0,49	0,00- 1,16	0,00- 1,42	0,00- 1,99	0,00- 0,59	0,00- 1,44
Laag	1,43- 2,28	2,00- 2,99	0,60- 1,39	1,33- 2,32	1,43- 2,20	1,67- 2,99	0,50- 1,39	1,17- 2,32	1,43- 2,28	2,00- 2,99	0,60- 1,39	1,45- 2,46
Beneden- gemiddeld	2,29- 2,78	3,00- 3,66	1,40- 1,99	2,33- 2,99	2,21- 2,78	3,00- 3,66	1,40- 1,99	2,33- 2,99	2,29- 2,82	3,00- 3,66	1,40- 1,99	2,47- 3,16
Gemiddeld	2,79- 3,28	3,67- 3,99	2,00- 2,59	3,00- 3,49	2,79- 3,28	3,67- 3,99	2,00- 2,59	3,00- 3,49	2,83- 3,28	3,67- 3,99	2,00- 2,59	3,17- 3,66
Boven- gemiddeld	3,29- 3,70	4,00- 4,32	2,60- 3,19	3,50- 3,99	3,29- 3,70	4,00- 4,32	2,60- 3,19	3,50- 3,99	3,29- 3,70	4,00- 4,32	2,60- 3,19	3,67- 3,99
Hoog	3,71- 4,28	4,33- 5,00	3,20- 3,99	4,00- 4,66	3,71- 4,20	4,33- 5,00	3,20- 3,99	4,00- 4,66	3,71- 4,28	4,33- 5,00	3,20- 3,99	4,00- 4,66
Zeer hoog	4,29- 5,00		4,00- 5,00	4,67- 5,00	4,21- 5,00		4,00- 5,00	4,67- 5,00	4,29- 5,00		4,00- 5,00	4,67- 5,00

Cliënten	Allen (N = 971)				Mannen (N = 365)				Vrouwen (N = 606)			
	Totaal	EWB	SWB	PWB	Totaal	EWB	PWB	SWB	Totaal	EWB	PWB	SWB
Zeer laag	0,00- 0,42	0,00- 0,66	0,00- 0,19	0,00- 0,32	0,00- 0,42	0,00- 0,66	0,00- 0,19	0,00- 0,32	0,00- 0,37	0,00- 0,66	0,00- 0,59	0,00- 0,32
Laag	0,43- 0,92		0,20- 0,59	0,33- 0,99	0,43- 0,99		0,20- 0,79	0,33- 0,99	0,38- 0,85			0,33- 0,99
Beneden- gemiddeld	0,93- 1,42	0,67- 1,32	0,60- 1,19	1,00- 1,49	1,00- 1,49	0,67- 1,32	0,80- 1,19	1,00- 1,66	0,86- 1,42	0,67- 1,32	0,60- 0,99	1,00- 1,49
Gemiddeld	1,43- 1,99	1,33- 1,99	1,20- 1,79	1,50- 2,16	1,50- 1,99	1,33- 1,99	1,20- 1,79	1,67- 2,16	1,43- 1,92	1,33- 1,99	1,00- 1,59	1,50- 2,16
Boven- gemiddeld	2,00- 2,56	2,00- 2,66	1,80- 2,39	2,17- 2,82	2,00- 2,56	2,00- 2,66	1,80- 2,39	2,17- 2,82	1,93- 2,56	2,00- 2,66	1,60- 2,39	2,17- 2,82
Hoog	2,57- 3,56	2,67- 3,99	2,40- 3,39	2,83- 3,99	2,57- 3,54	2,67- 3,99	2,40- 3,53	2,83- 3,99	2,57- 3,63	2,67- 3,99	2,40- 3,39	2,83- 3,99
Zeer hoog	3,57- 5,00	4,00- 5,00	3,40- 5,00	4,00- 5,00	3,55- 5,00	4,00- 5,00	3,54- 5,00	4,00- 5,00	3,64- 5,00	4,00- 5,00	3,40- 5,00	4,00- 5,00

Noot. EWB = Emotioneel Welbevinden; PWB = Psychologisch Welbevinden; SWB = Sociaal Welbevinden.

Conversie naar T-scores

Tabel A in de online bijlage geeft een overzicht van de conversieformules om ruwe scores te converteren naar genormaliseerde T-scores. Volgens de formule $0,66559x^3 - 4,8103x^2 + 21,251x + 10,0$ (waarbij x staat voor de ruwe score) in tabel A geeft een ruwe score van 1,00 een T-score van 21,7; een ruwe score van 3,15 geeft een T-score van 50,0. Figuur 1 geeft met oversteekwaarden de vertaling weer van ruwe scores naar T-scores.



FIGUUR 1 Oversteekfiguur met alle mogelijke ruwe scores op de vier schalen van de MHC-SF en hun positie ten opzichte van T-scores. MHC_EMO = Emotioneel Welbevinden; MHC_PSY = Psychologisch Welbevinden; MHC_SOC = Sociaal Welbevinden; MHC_TOT = positieve geestelijke gezondheid.

Tabel 5 is een oversteektabel van ruwe scores naar T-scores. Tevens worden percentielscores gegeven. Een cliënt met een score van 0,43 op de MHC-SF-totaalschaal heeft een T-score van 18,3 en een percentielscore ten opzichte van de bevolking (P_n) < 1 (minder dan 1% van de algemene bevolking heeft zo'n lage score) en een $P_{cl} = 5,4$ (score behoort tot de laagste 5% van de klinische populatie uit de specialistische ggz).

TABEL 5 Oversteektabel van ruwe score naar berekende T-score en naar percentielscores (algemene bevolking en s-ggz-cliënten) voor de MHCSF-schalen

TOT				EWB			
RS	T	PR_n	PR_cl	RS	T	PR_n	PR_cl
0,00	10,0	0,0	0,4	0,00	15,8	0,1	3,3
0,14	12,9	0,1	1,6	0,33	19,9	0,4	9,9
0,29	15,8	0,2	3,2	0,67	23,6	0,6	17,9
0,43	18,3	0,3	5,4	1,00	26,8	1,4	28,0
0,57	20,7	0,4	8,9	1,33	29,6	2,8	38,6
0,71	22,9	0,7	12,9	1,67	32,3	4,3	49,4
0,86	25,2	1,4	17,8	2,00	34,7	6,8	60,5
1,00	27,2	2,0	23,0	2,33	37,1	10,3	69,8
1,14	29,0	2,8	28,6	2,67	39,6	14,5	77,5
1,29	30,9	3,7	33,4	3,00	42,1	21,8	84,0
1,43	32,5	4,7	39,5	3,33	44,9	31,8	88,9
1,57	34,1	6,0	45,4	3,67	48,0	42,6	92,6
1,71	35,6	8,2	51,7	4,00	51,5	59,4	95,8
1,86	37,2	10,6	56,7	4,33	55,5	76,4	98,0
2,00	38,6	13,4	61,0	4,67	60,3	85,3	99,0
2,14	40,0	17,4	66,6	5,00	65,6	94,1	99,7
2,29	41,5	20,8	72,9				
2,43	42,8	25,3	77,3				
2,57	44,2	29,9	80,5				
2,71	45,6	35,8	84,0				
2,86	47,1	41,9	86,3				
3,00	48,5	47,7	88,4				
3,14	50,0	53,6	89,9				
3,29	51,6	60,3	91,7				
3,43	53,2	67,2	93,5				
3,57	54,9	73,9	95,0				
3,71	56,7	79,6	96,1				
3,86	58,7	84,9	96,8				
4,00	60,7	89,1	97,5				
4,14	62,8	93,1	98,2				
4,29	65,2	95,4	98,8				
4,43	67,7	97,4	99,2				
4,57	70,2	98,6	99,7				
4,71	73,0	99,4	99,8				
4,86	76,1	99,8	100,0				
5,00	79,2	99,9	100,0				

Noot. EWB = Emotioneel Welbevinden; PR_n = Percentielscore algemene bevolking; PR_cl = Percentielscore cliënten; PWB = Psychologisch Welbevinden; SWB = Sociaal Welbevinden; RS = ruwe score; T = T-score.

De T-scoreconversie was succesvol (de gemiddelde score op de diverse schalen was $M = 50$; $SD \approx 9,3$) in de algemene bevolking. Scheefheid en gepiektheid in ruwe scores zijn afgenomen, wat leidde tot scores die bij benadering normaal verdeeld zijn. In figuur B in de online bijlage is dit geïllustreerd met een histogram voor de frequentieverdeling van scores voor en na conversie voor de MHC-SF-totaalscore. De met een conversieformule berekende T-scores kwamen goed overeen met de theta-based T-scores uit IRT-analyses. De Intra Class Correlatie tussen op IRT gebaseerde theta's en berekende T-scores is bepaald en was hoog ($ICC > 0,99$). De overeenkomst tussen beide soorten scores werd ook met Bland-Altman plot geïnspecteerd. Daarmee kan over de gehele breedte van de schaal nagegaan worden of er discrepantie is in beide scores. IRT-gebaseerde T-scores en berekende T-scores bleken ook hier goed overeen te komen (zie figuur C in de online bijlage). Berekende T-scores waren iets hoger dan op theta gebaseerde T-scores, maar 95% van de verschillen lagen tussen -2,7 en 2,5 T-punten.

DISCUSSIE

.....

De belangrijkste bevindingen van het onderzoek naar de psychometrie van de MHC-SF zijn: ondersteuning voor de validiteit gezien de resultaten van factoranalyse, de correlatie met de OQ-SD, en het verschil in score tussen cliënten en de algemene bevolking. Hiermee zijn de gunstige psychometrische eigenschappen van de MHC-SF die eerder werden gerapporteerd (Franken et al., 2019; Lamers et al., 2011, 2012) gerepliceerd in het huidige onderzoek. Nieuw is de bevinding dat de responsiviteit van de MHC-SF iets lager was dan de OQ-SD, met name voor sociaal welbevinden. De omzetting naar T-scores was succesvol en levert bruikbare normgegevens op.

Uit de vergelijking tussen verschillende factormodellen kwam naar voren dat de bi-factor oplossing met een algemene factor voor positieve psychologische gezondheid, en drie factoren voor emotioneel, sociaal en psychologisch welbevinden, het beste paste bij de data. Goede fit voor het bi-factor model is ook gerapporteerd door Lamers en collega's (2011) voor de algemene bevolking. In een studie van Longo en collega's (2017) werd de MHC-SF onderzocht en vergeleken tussen meerdere landen bij bevolkingssteekproeven door middel van Exploratory Structural Equation Modeling (ESEM). De onderzoekers vonden een betere fit van bi-factor dan 3-factor modellen zonder totaalscore. ESEM werd door Silverman en collega's (2018) gebruikt in een klinische steekproef en liet daar ook de beste fit zien voor het bi-factor model. ESEM is een veelbelovende techniek om verder onderzoek te doen naar de factorstructuur van de MHC-SF bij klinische groepen, waar doorgaans hogere correlaties tussen de schalen onderling gevonden worden.

De hoge betrouwbaarheid (interne consistentie) van de MHC-SF-totaalscore en de schaalcores is in overeenstemming met eerdere bevindingen

van internationaal onderzoek (Donnelly et al., 2019) en van Nederlands onderzoek (de Vos et al., 2018; Franken et al., 2018; Kennes et al., 2020; Lamers et al., 2011; van Erp Taalman Kip & Hutschemaekers, 2018).

Er werden hoge negatieve correlaties gevonden tussen de MHC-SF-schalen en de OQ-SD, wat ten dele te verklaren is door overeenkomst in iteminhoud met de positief geformuleerde items in de OQ-SD. Toch is de hoge negatieve correlatie opvallend, mede omdat in beide instrumenten een verschillende periode wordt voorgesteld waarop de vragen betrekking hebben (de afgelopen week bij de OQ-SD en de afgelopen maand bij de MHC-SF). Tegelijk laten de correlatiecoëfficiënten zien dat positieve psychische gezondheid niet hetzelfde is als lage symptomatische distress. Een correlatie van 0,60-0,70 geeft zo'n 40% gedeelde variantie weer en beide constructen hebben dus ook veel unieks. We vonden min of meer dezelfde correlaties bij verschillcores. Dit duidt er eveneens op dat er weliswaar samenhang is, maar ook diversiteit in verandering van score: bij sommige respondenten was sprake van convergentie in afname van psychopathologie en toename van positieve psychische gezondheid, terwijl bij andere deze uitkomsten juist divergeerden. Divergentie van uitkomsten op beide concepten – symptomen en welbevinden – kan aanknopingspunten bieden voor differentiatie in behandeling, zoals meer of juist minder focus op symptomatologie. De OQ-45 bevat nog twee schalen: de OQ-IR-schaal voor problemen met interpersoonlijk functioneren (interpersoonlijke relaties) en de OQ-SR-schaal voor functioneren op werk of school (sociale rol). Deze twee schalen werden echter niet afgenomen bij PuntP, waardoor de samenhang van deze concepten met welbevinden niet kon worden onderzocht, wat een beperking van het onderzoek is.

Het onderzoek bracht een sekseverschil en een leeftijdsverschil aan het licht. Mannen hadden een hogere score voor welbevinden dan vrouwen, en de score neemt toe met de leeftijd, met uitzondering van de jongste leeftijdsgroep, die relatief hoog scoort in welbevinden. De toename in welbevinden bij hogere leeftijd was vooral sterk aanwezig bij vrouwen. Een mogelijke verklaring voor het sekseverschil kan zijn dat (jonge) vrouwen het zwaarder hebben onder de huidige maatschappelijke omstandigheden. In vergelijking met mannen hebben vrouwen geringere carrièrekansen, een lager inkomen (bij gelijke functie), moeten ze hun baan vaak combineren met het uitvoeren van de meerderheid van de huishoudelijke taken in een gezin met jonge kinderen, enzovoort. In de internationale literatuur wordt het sekseverschil in welbevinden ook wel de *gender gap* genoemd. Hier is veel over gepubliceerd, zowel in Nederland (Verdonk & de Rijk, 2008) als internationaal (Nolen-Hoeksema & Rusting, 1999; Western & Tomaszewski, 2016). Het leeftijdsverschil is in overeenstemming met eerdere bevindingen van Stone en collega's (2010), die als verklaringen noemen: toegenomen wijsheid en emotionele intelligentie, minder herinneringen aan negatieve gebeurtenissen en een toegenomen vermogen om eigen emoties te reguleren

bij oudere leeftijdsgroepen. Al deze verklaringen gaan uit van het toenemen van welbevinden bij ouder worden als een intrapsychische ontwikkeling. Een alternatieve mogelijkheid is een meer sociologische verklaring: de samenleving wordt al maar complexer en er worden steeds hogere eisen aan individuen gesteld, waardoor opeenvolgende generaties steeds slechter af zijn en jongere respondenten minder welbevinden rapporteren dan oudere respondenten. Alleen met een analyse van longitudinale data van welbevinden over een langere tijdsspanne kunnen we nagaan welke verklaring het meest valide is.

Een beperking van dit onderzoek is dat we bij het vaststellen van de responsiviteit de twee meetinstrumenten alleen vergeleken hebben met een enkele hertest, die was uitgevoerd in de beginfase van de behandeling. Het is mogelijk dat toename van positieve psychische gezondheid en afname van psychopathologie niet synchroon verlopen, maar dat eerst symptomen afnemen en pas later het welbevinden toeneemt (Lamers et al., 2012). Dat zou ook de wat grotere responsiviteit van de OQ-SD in vergelijking met de MHC-SF verklaren. In de toekomst is het interessant om langere series van meetgegevens op beide instrumenten te vergelijken. Voorts kan een vergelijking tussen de groep met divergerende uitkomsten en de groep met convergerende uitkomsten aanwijzingen opleveren voor de achtergrond van (de)synchroniteit in beloop. Te denken valt aan demografische, sociaal-economische of klinische kenmerken, of aan verschillen in wijze van behandeling.

Het onderzoek leverde bruikbare normgegevens om de scores op de MHC-SF af te zetten tegen de algemene bevolking en de klinische populatie van cliënten, die worden behandeld in de specialistische ggz. Er werd een sekseverschil gevonden (mannen rapporteren meer welbevinden) en er was een verband met leeftijd (naarmate men ouder wordt, neemt welbevinden toe, met uitzondering van de jongste leeftijdsgroep onder de 21 jaar). Voor de interpretatie van de score kan het nuttig zijn om de ruwe score te vergelijken met de meest passende leeftijdsgroep met hetzelfde geslacht als de respondent, of om een leeftijds- en geslachtscongruente T-score conversieformule te kiezen.

De beschikbaarheid van een omvangrijke dataset met meetresultaten uit een bevolkingssteekproef heeft ons in staat gesteld om T-scores te berekenen die een verfijnde interpretatie van testresultaten mogelijk maken. Dit laat uitspraken toe als: 'Uw T-score is 32,5, wat lager is dan 95% van alle vrouwen uit uw leeftijdsgroep', of: 'Uw score is hoger dan 65% van de vrouwen met dezelfde leeftijd die een behandeling beginnen in de specialistische ggz.' Met een klassieke normtabel, zoals weergegeven in tabel 4, wordt betekenis gegeven aan ruwe scores in zeven niveaus, van 'zeer laag' tot 'zeer hoog'; met een conversie naar T-scores blijft het continue karakter van de ruwe score behouden. In de loop van de tijd zal meer 'gevoel' ontstaan bij gebruikers van ROM-gegevens (cliënten en hun behandelaars) voor de betekenis van T-scores. Gezien de grote diversiteit aan meetinstrumenten en bijbehoren-

de diversiteit aan testresultaten in ruwe schaalscores, kan het gebruik van een uniforme meetschaal zoals de T-score de onderlinge communicatie over ROM-metingen tussen professionals onderling, maar ook tussen de behandelaar en de cliënt, zeer ten goede komen. Cliëntvriendelijke communicatie bij het uitleggen van de testcores is uiterst belangrijk voor de optimale benutting van ROM. In de online bijlage zijn twee figuren (D en E) opgenomen die kunnen helpen bij de uitleg wat een T-score betekent.

Edwin de Beurs is als senior onderzoeker verbonden aan de afdeling onderzoek van Arkin en is als hoogleraar verbonden aan de Universiteit Leiden, Faculteit Sociale Wetenschappen, Sectie Klinische Psychologie.

Sjoerd Kosterman was ten tijde van de uitvoering van het onderzoek als directeur behandelzaken verbonden aan PuntP, de s-ggz kliniek van Arkin. Nu is hij als zelfstandige werkzaam als trainer, coach en organisatieadviseur.

Suzanne Anten schreef haar masterthesis over dit onderzoek aan de Universiteit Leiden, Sectie Klinische Psychologie. Nu is zij als trainee verbonden aan OGD ict-diensten.

Ernst Bohlmeijer is als hoogleraar en hoofd van de afdeling Technology, Human and Institutional Behaviour verbonden aan de Universiteit Twente.

Gerben J. Westerhof is als hoogleraar en voorzitter van de sectie Psychologie, Gezondheid en Technologie verbonden aan de Universiteit Twente.

Correspondentieadres: Prof. dr. E. de Beurs, Arkin, afdeling onderzoek, Klaprozenweg 111, 1033 NN Amsterdam. E-mail: edwin.de.beurs@arkin.nl.

Summary *Psychometric evaluation of the Mental Health Continuum – Short Form (MHC-SF): Construct validity, responsiveness, norms, and T-scores*

The Mental Health Continuum - Short Form (MHC-SF) was developed to measure positive mental health. This includes emotional, social and psychological well-being. The aim of this study was to investigate the reliability and validity of the MHC-SF. Construct validity and concurrent validity were discussed. We also examined the responsiveness of the MHC-SF by comparing change on the MHC-SF with change on the Symptomatic Distress scale of the Outcome Questionnaire (OQ-SD). Finally, we provide norm data and T-scores for the MHC-SF. The psychometric characteristics of the MHC-SF were examined with data from a population sample (N = 1802) and a sample of clients treated in specialist mental healthcare (N = 971). The results supported the previously found factor structure with a general positive mental health factor, and three subscale scores for emotional, social and psychological well-being. The responsiveness of the MHC-SF was comparable to the OQ-SD, except for change in social well-being, which lagged significantly behind the OQ-SD. This makes the MHC-SF a useful addition to a ROM set, but we should not expect quite the same changes on the MHC-SF as found with the OQ-SD scale. With standardized T-scores, measurement results can be interpreted more easily: the T-score directly indicates how far a test result is removed from the population mean and thus how exceptional the test result is. Compared to traditional norm tables, a T-score allows a more refined way to report

on the outcome of a test. T-scores thus facilitate communication about the measurement results with the client or with colleagues.

152

Keywords *positive psychology, wellbeing, psychometrics, responsiveness, T-scores*

Literatuur

- Anthony, W. A. (1993). Recovery from mental illness: The guiding vision of the mental health service system in the 1990s. *Psychosocial Rehabilitation Journal*, *16*, 11-23.
- Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., Ader, D., Fries, J. F., Bruce, B., & Rose, M. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap cooperative group during its first two years. *Medical Care*, *45*, S3-S11.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*, 1-29.
- Choi, S. W., Schalet, B., Cook, K. F., & Cella, D. (2014). Establishing a common metric for depressive symptoms: Linking the BDI-II, CES-D, and PHQ-9 to PROMIS Depression. *Psychological Assessment*, *26*, 513-527.
- Cook, K. F., Schalet, B. D., Kallen, M. A., Rutsohn, J. P., & Cella, D. (2015). Establishing a common metric for self-reported pain: Linking BPI Pain Interference and SF-36 Bodily Pain Subscale scores to the PROMIS Pain Interference metric. *Quality of Life Research*, *24*, 2305-2318.
- de Beurs, E. (2010). De genormaliseerde T-score, een 'euro' voor testuitslagen. *Maandblad Geestelijke Volksgezondheid*, *65*, 684-695.
- de Beurs, E., Barendregt, M., Flens, G., van Dijk, E., Huijbrechts, I., & Meerding, J. W. (2012). Vooruitgang in de behandeling meten: Een vergelijking van vragenlijsten voor zelfrapportage. *Maandblad Geestelijke Volksgezondheid*, *67*, 259-270.
- de Beurs, E., den Hollander-Gijsman, M., Buwalda, V., Trijsburg, W., & Zitman, F. G. (2005). De Outcome Questionnaire (OQ-45): Een meetinstrument voor meer dan alleen psychische klachten. *De Psycholoog*, *40*, 53-63.
- de Beurs, E., den Hollander-Gijsman, M. E., van Rood, Y. R., van der Wee, N. J., Giltay, E. J., van Noorden, M. S., van der Lem, R., van Fenema, E., & Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands: Practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical Psychology & Psychotherapy*, *18*, 1-12.
- de Beurs, E., Flens, G., & Williams, G. (2019). Meetresultaten interpreteren in de klinische psychologie: Een aantal voorstellen. *De Psycholoog*, *54*, 10-23.
- de Beurs, E., Oudejans, S., & Terluin, B. (Submitted). A common measurement scale for scores from self-report instruments in mental health care: T scores with a normal distribution. *European Journal of Psychological Assessment*.
- de Beurs, E., Thomaes, K., Kronemeijer, H., & Dekker, J. J. M. (2020). Een vergelijking van de responsiviteit van de PTSS Checklist voor DSM-5 (PCL-5) en de Outcome Questionnaire (OQ-45). *Tijdschrift voor Psychiatrie*, *62*, 448-456.
- de Beurs, E., Vissers, E., Schoevers, R., Carlier, I. V. E., van Hemert, A. M., & Meesters, Y. (2019). Comparative responsiveness of generic versus disorder-specific instruments for depression: An assessment in three

- longitudinal datasets. *Depression and Anxiety*, 36, 93-102.
- de Beurs, E., & Zitman, F. G. (2006). De Brief Symptom Inventory (BSI): De betrouwbaarheid en validiteit van een handzaam alternatief voor de SCL-90. *Maandblad Geestelijke Volksgezondheid*, 61, 120-141.
- de Jong, K., Nugter, M. A., Polak, M. G., Wagenborg, J. E. A., Spinhoven, P., & Heiser, W. J. (2007). The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation. *Clinical Psychology & Psychotherapy*, 14, 288-301.
- de Vos, J. A., Radstaak, M., Bohlmeijer, E. T., & Westerhof, G. J. (2018). Having an eating disorder and still being able to flourish? Examination of pathological symptoms and well-being as two continua of mental health in a clinical sample. *Frontiers in Psychology*, 9, 2145.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95, 542-575.
- Donnelly, A., O'Reilly, A., Dolphin, L., O'Keeffe, L., & Moore, J. (2019). Measuring the performance of the Mental Health Continuum – Short Form (MHC-SF) in a primary care youth mental health service. *Irish Journal of Psychological Medicine*, 36, 201-205.
- Franken, K., de Vos, S., Westerhof, G., & Bohlmeijer, E. (2019). *De Mental Health Continuum Short-Form: Het meten van welbevinden als uitkomst van behandeling gericht op geestelijke gezondheid van volwassenen met psychische klachten*. Universiteit Twente.
- Franken, K., Lamers, S. M., ten Klooster, P. M., Bohlmeijer, E. T., & Westerhof, G. J. (2018). Validation of the Mental Health Continuum-Short Form and the dual continua model of well-being and psychopathology in an adult mental health setting. *Journal of Clinical Psychology*, 74, 2187-2202.
- Herrman, H., Saxena, S., & Moodie, R. (2005). *Promoting mental health: Concepts, emerging evidence, practice*. World Health Organization.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55.
- Huber, M., Knottnerus, J. A., Green, L., van der Horst, H., Jadad, A. R., Kromhout, D., Leonard, B., Lorig, K., Loureiro, M. I., van der Meer, J. W. M., Schnabel, P., Smith, R., van Weel, C., & Smid, H. (2011). How should we define health? *British Medical Journal*, 343, 235-237.
- Huber, M., van Vliet, M., & Boers, I. (2016). Heroverweeg uw opvatting van het begrip 'gezondheid'. *Nederlands Tijdschrift voor de Geneeskunde*, 160, 1-5.
- Kennes, A., Peeters, S., Janssens, M., Reijnders, J., Lataster, J., & Jacobs, N. (2020). Psychometric evaluation of the Mental Health Continuum – Short Form (MHC-SF) for Dutch adolescents. *Journal of Child and Family Studies*, 29, 3276-3286.
- Keyes, C. L. M. (1998). Social well-being. *Social Psychology Quarterly*, 121-140.
- Keyes, C. L. M. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior*, 43, 207-222.
- Keyes, C. L. M. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology*, 73, 539-548.
- Keyes, C. L. M., & Martin, C. C. (2017). The complete state model of mental health. In M. Slade, L. Oades, & A. Jarden (Eds.), *Well-being, recovery and mental health* (pp. 75-85). Cambridge University Press.
- Lambert, M. J., Burlingame, G. M., Umphress, V. J., Hanssen, N., Vermeersch, D., Clause, G., & Yanchar, S. C. (1996). The reliability and validity of the

- Outcome Questionnaire. *Clinical Psychology & Psychotherapy*, 3, 249-258.
- Lambert, M., Morton, J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R., & Burlingame, G. (2004). *Administration and scoring manual for the OQ-45.2 (Outcome Questionnaire)*. American Professional Credentialing Services.
- Lamers, S. M. A., Glas, C. A., Westerhof, G. J., & Bohlmeijer, E. T. (2012). Longitudinal evaluation of the Mental Health Continuum – Short Form (MHC-SF). *European Journal of Psychological Assessment*, 28, 290-296.
- Lamers, S. M. A., Westerhof, G. J., Bohlmeijer, E. T., ten Klooster, P. M., & Keyes, C. L. M. (2011). Evaluating the psychometric properties of the Mental Health Continuum – Short Form (MHC-SF). *Journal of Clinical Psychology*, 67, 99-110.
- Longo, Y., Jovanović, V., Sampaio de Carvalho, J., & Karaš, D. (2017). The general factor of well-being: Multinational evidence using bifactor ESEM on the Mental Health Continuum – Short Form. *Assessment*, 27, 596-606.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63, 737-745.
- Nolen-Hoeksema, S., & Rusting, C. L. (1999). Gender differences in well-being. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: Foundations of hedonic psychology* (pp. 330-350). Russel Sage Foundation.
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57, 1069.
- Schalet, B. D., Cook, K. F., Choi, S. W., & Cella, D. (2014). Establishing a common metric for self-reported anxiety: Linking the MASQ, PANAS, and GAD-7 to PROMIS Anxiety. *Journal of Anxiety Disorders*, 28, 88-96.
- Schalet, B. D., Revicki, D. A., Cook, K. F., Krishnan, E., Fries, J. F., & Cella, D. (2015). Establishing a common metric for physical function: Linking the HAQ-DI and SF-36 PF subscale to PROMIS Physical Function. *Journal of General Internal Medicine*, 30, 1517-1523.
- Schawo, S. J., Carlier, I. V. E., van Hemert, A. M., & de Beurs, E. (2019). Measuring treatment outcome in patients with anxiety disorders: A comparison of the responsiveness of generic and disorder-specific instruments. *Journal of Anxiety Disorders*, 64, 55-63.
- Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, 55, 5-14.
- Silverman, A. L., Forgeard, M., Beard, C., & Björgvinsson, T. (2018). Psychometric properties of the Mental Health Continuum – Short Form in a psychiatric sample. *Journal of Well-Being Assessment*, 2, 57-73.
- Stone, A. A., Schwartz, J. E., Broderick, J. E., & Deaton, A. (2010). A snapshot of the age distribution of psychological well-being in the United States. *Proceedings of the National Academy of Sciences*, 107, 9985-9990.
- Timman, R., de Jong, K., & de Neve-Enthoven, N. (2017). Cut-off scores and clinical change indices for the Dutch Outcome Questionnaire (OQ-45) in a large sample of normal and several psychotherapeutic populations. *Clinical Psychology & Psychotherapy*, 24, 72-81.
- Tudor, K. (1996). *Mental health promotion: Paradigms and practice*. Routledge.

- van Erp Taalman Kip, R. M., & Hutsche-
maekers, G. J. M. (2018). Health,
well-being, and psychopathology in a
clinical population: Structure and
discriminant validity of Mental
Health Continuum Short Form
(MHC-SF). *Journal of Clinical Psy-
chology, 74*, 1719-1729.
- Verdonk, P., & de Rijk, A. (2008). Loop-
baansucces en welbevinden van Ne-
derlandse werknemers M/V. *Gedrag
& Organisatie, 21*, 451-474.
- Wahl, I., Löwe, B., Bjorner, J. B., Fischer,
F., Langa, G., Voderholzer, U., Aita,
S. A., Bergemann, N., Brähler, E., &
Rose, M. (2014). Standardization of
depression measurement: A common
metric was developed for 11 self-
report depression measures. *Journal
of Clinical Epidemiology, 67*, 73-86.
- Western, M., & Tomaszewski, W. (2016).
Subjective wellbeing, objective well-
being and inequality in Australia.
PloS one, 11, e0163345.