

ARTIKEL

Psychometrische evaluatie van een beoordelaarschaal voor het meten van de behandelintegriteit bij de uitvoering van Kortdurende Psychoanalytische Steungevende Psychotherapie (KPSP)

Matthijs Vroegop, Mariëlle Hendriksen, René Kortrijk, Peter ten Klooster, Rien Van, Jack Dekker en Maarten van Dijk

Samenvatting

Kortdurende Psychoanalytische Steungevende Psychotherapie (*KPSP*) is een behandelmethode voor de behandeling van depressie die qua effectiviteit niet onder lijkt te doen voor die van Cognitieve Gedragstherapie (*CGT*). Claims met betrekking tot de non-inferioriteit ondervinden nadeel van het feit dat voor *KPSP* nog geen instrument bestaat om de

Matthijs Vroegop is GZ-psycholoog in opleiding tot klinisch psycholoog bij Dimence Groep.

E-mail ■ m.vroegop@transfore.nl

Mariëlle Hendriksen, PhD, is GZ-psycholoog, psychoanalytisch (groeps) psychotherapeut en directeur Proven Therapies in Hilversum.

Dr. Peter ten Klooster is assistent professor bij Universiteit Twente.

René Kortrijk is klinisch psycholoog bij NPI, Arkin en in eigen praktijk.

Rien Van, PhD, is psychiater, psychoanalytisch psychotherapeut, directeur behandelzaken NPI en A-opleider bij Arkin.

Jack Dekker, PhD, is bijzonder hoogleraar aan de faculteit Gedrags- en bewegingswetenschappen van de Vrije Universiteit Amsterdam.

Maarten van Dijk, PhD, is klinisch psycholoog en hoofdonderzoeker bij Dimence Groep.

Leerdoelen

Na het lezen van dit artikel:

1. bent u bekend met de methode specifieke elementen van KPSp;
2. weet u hoe deze elementen geoperationaliseerd kunnen worden in de ontwikkeling van een beoordelingsinstrument voor het meten van behandelintegriteit bij de uitvoering van KPSp;
3. bent u op de hoogte van het belang van een dergelijk instrument in het verbeteren van de behandeling van depressie middels deze behandelmethode.

behandelintegriteit te bepalen. De ontwikkeling van een dergelijk instrument wordt beschreven. Initiële items werden geformuleerd. De psychometrische kwaliteit van het instrument werd bepaald. Resultaten: De interne consistentie van het instrument is hoog, de interbeoordelaarsbetrouwbaarheid is redelijk tot goed en het lijkt mogelijk om een significant verschil aan te tonen tussen de scores van ervaren en minder ervaren therapeuten. Een betrouwbaar en valide instrument voor het meten van de behandelintegriteit bij KPSp lijkt binnen handbereik. Meer onderzoek naar de psychometrische eigenschappen van het instrument is evenwel nodig, uitgevoerd bij een grotere steekproef van met KPSp behandelde patiënten.

Inleiding

Dit artikel beschrijft de ontwikkeling van een instrument om de behandelintegriteit bij de uitvoering van KPSp voor patiënten met een depressie te meten. Ook wordt ingegaan op de opzet, uitvoer en de resultaten van een eerste onderzoek naar de psychometrische eigenschappen van dit instrument.

Depressie is een psychiatrische stoornis met een hoge prevalentie en, afhankelijk van de ernst, grote gevolgen voor patiënten en hun omgeving. Van de volwassenen tot 65 jaar krijgt in Nederland ongeveer 20% ooit in hun leven te kampen met een depressieve stoornis (Trimbos-instituut, 2016). Het lijden aan een depressie heeft grote gevolgen voor iemands persoonlijk welbevinden. Het leidt tot belangrijke beperkingen in sociaal, emotioneel en lichamelijk functioneren (Beekman e.a., 2002; Bijl & Ravelli, 2000; Greden, 2001). Het voorkomen van depressie heeft daarnaast ook grote economische gevolgen. Enerzijds maken mensen met een depressie extra gebruik van medische voorzieningen. Anderzijds ontstaan er productieverliezen in betaalde en onbetaalde arbeid door ziekteverzuim. Voor depressie zijn de jaarlijkse behandelkosten in Nederland 660 miljoen euro. Daarnaast ontstaat ook nog eens 953 miljoen euro aan kosten door ziekteverzuim in betaalde arbeid (Romijn, Ruiters & Smit, 2008). De behoefte aan goede behandelingen voor depressie is dan ook groot.

Er zijn verschillende bewezen effectieve behandelingen voor depressie beschikbaar. Bij een matige tot ernstige depressie adviseert de landelijke Multidisciplinaire Richtlijn Depressie om, in overleg met de patiënt, in eerste instantie een keuze te maken voor ofwel een vorm van psychotherapie, ofwel een behandeling met medicijnen (Spijker e.a., 2013). In de praktijk blijkt dat patiënten daarbij meestal een voorkeur hebben voor een vorm van psychotherapie (Swift e.a., 2017). In de afgelopen jaren zijn verschillende soorten psychotherapie effectief bevonden in de behandeling van depressie (Cuijpers, 2015). Deze therapieën verschillen onderling in hun verklaring voor het bestaan en het ontstaan van depressieve klachten en kennen elk andere aangrijpingspunten voor therapie. Met name Cognitieve Gedragstherapie (CGT) is veel onderzocht en deze methode wordt in de praktijk van de Nederlandse ggz ook vaak aangeboden. Echter, ook andere behandelvormen, zoals Interpersoonlijke Psychotherapie (IPT) en Kortdurende Psychodynamische Psychotherapie (KPSP) kunnen steeds meer bogen op een wetenschappelijke onderbouwing voor de werkzaamheid ervan (Picardi & Gaetano, 2014). Uit een meta-analyse van Cuijpers (2015) komt hierbij naar voren dat er tussen de verschillende bewezen effectieve behandelingen bij depressie nauwelijks onderlinge verschillen zijn in effectiviteit. Meerdere onderzoeken hebben aangetoond dat grote groepen therapeuten een voorkeur hebben voor psychodynamische psychotherapie (Norcross & Rogan, 2013). Er zijn ook aanwijzingen dat het geloof en vertrouwen in de eigen methode, ook wel *allegiance* genoemd, een positief effect heeft op de uitkomst van behandelingen (McLeod, 2009). Om die reden is het wenselijk dat ook te kiezen valt voor een vorm van psychodynamische psychotherapie bij de behandeling van depressie, zeker als deze therapie zowel door de patiënt alsook de therapeut aantrekkelijk wordt gevonden.

KPSP is een specifieke vorm van psychodynamische psychotherapie bij depressie. KPSP is een vorm van kortdurende individuele psychotherapie, waarbij in de eerste zitting alle therapieafspraken al vastgelegd worden. Met het direct vastleggen van de therapieafspraken wordt de eindigheid van de therapie vanaf het begin af aan aangekondigd en gemarkeerd, hetgeen het automatisch ook tot een onderwerp binnen de therapie maakt. Het doel van de therapie binnen KPSP is om vanuit een relationele visie op het ontstaan van depressie de symptomen te verminderen, het sociaal functioneren te verbeteren en beperkte persoonlijkheidsverandering te bewerkstelligen (De Jonghe e.a., 2013). Hierbij wordt de depressie begrepen vanuit een relationeel perspectief of psychodynamische deeltheorieën die gaan over superegopathologie, afweerstructuur, zelfpathologie of object-relatieel functioneren. KPSP is in meerdere RCT's onderzocht op effectiviteit en ook werkzaam gebleken (De Jonghe e.a., 2001, 2004; Dekker e.a., 2005, 2008). Volgens Driessen en anderen (2010) blijkt KPSP ook een gelijkwaardig alternatief voor CGT te zijn. In een *randomized controlled trial* (RCT) van Driessen en anderen (2013) is de non-inferioriteit van zestien zittingen KPSP aan zestien zittingen CGT aangetoond. Deze RCT werd uitgevoerd in de setting waar KPSP ook werd ontwikkeld. Als de resultaten van deze RCT zouden worden gerepliceerd in een onafhankelijke setting, dan zou KPSP voldoen aan de kwalificatie-eisen voor een empirisch ondersteunde behandeling volgens de criteria van Chambless & Hollon (1998). Op dit moment gebeurt dat in de D*Phase-studie, een studie die is opgestart vanuit een samenwerking tussen de Dimence Groep en een aantal collega's van Arkin (Migglioli e.a., 2021);

en waarbij KPSp dus opnieuw in een RCT met CGT vergeleken wordt. Bevestigen de resultaten ook de eerdere uitkomsten, dan zou dit zeker extra motivatie kunnen geven voor een besluit om KPSp, als specifieke variant van KPP, ook als een van de eerstekeuzeinterventies in de multidisciplinaire behandelrichtlijn voor depressie op te nemen.

Een belangrijke beperking van de eerder beschreven non-inferioriteitsstudie naar KPSp van Driessen en anderen (2013) is evenwel dat de behandelintegriteit van CGT en KPSp behandelingen in deze RCT niet gemeten zijn. Behandelintegriteit wordt vaker niet gemeten in studies die een vorm van kortdurende psychodynamische psychotherapie vergeleken met CGT (Bhar & Beck, 2009). In de studie van Driessen en anderen (2013) werd er wel controle op de behandelintegriteit uitgevoerd, in de vorm van supervisie, waarbij audio-opnames van de sessie werden beluisterd en besproken. De behandelintegriteit werd echter niet onafhankelijk gescoord aan de hand van een betrouwbaar en valide instrument. Daarom valt niet met zekerheid te zeggen of de onderzochte behandelmethoden uit het onderzoek van Driessen en anderen (2013) ook daadwerkelijk zijn uitgevoerd zoals bedoeld. De claim met betrekking tot non-inferioriteit van KPSp ten opzichte van CGT ondervindt hier nadeel van. Andere, niet-methodespecifieke factoren, zouden overeenkomstige effecten van de vergeleken behandelmethoden kunnen verklaren. Uit onderzoek blijkt dat zogeheten niet-methodespecifieke factoren stevig van invloed zijn op de uitkomst van een therapie (Chatoor & Kurpnick, 2001). Het meten van de behandelintegriteit is dus van belang voor het aantonen van non-inferioriteit.

De behandelintegriteit definieert de mate waarin de therapeut de theoriespecifieke methode en technische interventies tijdens de behandeling toepast en dit competent doet (Nezu & Nezu, 2008). Om behandelintegriteit te meten is een instrument nodig met goede psychometrische kwaliteiten waarmee de behandeling kan worden beoordeeld door een getrainde beoordelaar. Dit moet een instrument zijn waarin de belangrijkste methodespecifieke elementen zijn geoperationaliseerd. Met betrekking tot de minimaal vereiste aspecten van de psychometrische kwaliteit dient het instrument voldoende interne consistentie te vertonen. Wenselijk is ook dat het instrument goede van minder goed uitgevoerde behandelingen kan onderscheiden (Messick, 1980). Hierbij zou gekeken kunnen worden naar de mate waarin gemiddelde scores van *novice*- en *expert*-therapeuten van elkaar verschillen. Het gaat dan om zogenaamde 'Known group validiteit', dat wil zeggen de mate waarin een instrument onderscheid kan maken tussen twee groepen waarvan vooraf bekend is dat deze van elkaar verschillen (Davidson, 2014).

Waar voor het meten van behandelintegriteit van CGT bij depressie meetinstrumenten beschikbaar zijn (Vallis, Shaw & Dobson, 1986) is een dergelijk instrument voor KPSp nog niet voorhanden. Dit artikel doet verslag van de ontwikkeling van een instrument voor het meten van de behandelintegriteit binnen de behandeling van depressie met KPSp en de stappen die daarbij gevolgd zijn, om daarmee in dit hiaat te voorzien. Ook doen we daarbij verslag van de resultaten van een eerste verkennend onderzoek naar de psychometrische eigenschappen van dit instrument. Eerdere studies waarin dergelijke meetinstrumenten zijn ontwikkeld voor andere vormen van psychotherapie hebben daarbij als blauwdruk voor het te volgen proces dienst gedaan (Ogrodniczuk & Piper, 1999; Crane e.a., 2013).

Dit onderzoek is een deelproject van de D*Phase-psychotherapiestudie waarin nogmaals getoetst wordt of KPSp non-inferieur is aan CGT. Ook wordt in deze studie gekeken naar mogelijke prescriptieve variabelen voor behandeling alsook de meest effectieve strategie voor het omgaan met aanvankelijke behandelnon-response. In een eerste behandelfase van acht weken en zestien therapie sessies wordt daarbij de non-inferioriteit van KPSp aan CGT onderzocht. Ook wordt onderzocht of misschien specifieke patiëntengroepen meer voordeel zouden kunnen hebben van ofwel de eerste, ofwel de tweede vorm van psychotherapie. Patiënten worden in het kader van het onderzoek *at random* verdeeld over de twee condities. Na het doorlopen van deze eerste fase wordt bepaald of de depressieve symptomen zijn afgenomen. Wanneer dit niet het geval is, wordt in dit onderzoek in een tweede behandelfase gekeken wat de invloed is van ofwel doorgaan met dezelfde therapeut en dezelfde therapie methode, switchen enkel van therapeut of switchen van therapeut *en* therapie methode (Miggiels e.a., 2021).

Method

Er worden twee stappen onderscheiden in het ontwikkelproces van het beoordelingsinstrument voor het meten van behandelintegriteit bij KPSp in de behandeling van depressie.

Stap 1: Itemgeneratie en -selectie; de twee belangrijkste methodespecifieke elementen van KPSp: Adequate Psychoanalytische Steun (APS) en het hanteren van het juiste bespreekniveau werden geoperationaliseerd.

Stap 2: Bepalen van de psychometrische eigenschappen van het instrument; een eerste verkenning van de psychometrische eigenschappen van dit nieuwe instrument, op grond van data, verzameld bij een beperkt aantal uitgevoerde KPSp-behandelingen. Hierbij is gebruikgemaakt van twee verschillende beoordelaars die aan de hand van het instrument de mate van de behandelintegriteit hebben gescoord.

Stap 1 – Itemgeneratie en -selectie

Een van de belangrijkste methodespecifieke factoren van KPSp is APS. Een tweede belangrijke methodespecifieke factor is de mate waarin rekening wordt gehouden met de verschillende bespreekniveaus, zoals die binnen KPSp onderscheiden worden (De Jonghe e.a., 2013, 2014). Binnen de ontwikkeling van het meetinstrument voor het meten van de behandelintegriteit zullen de twee belangrijkste methodespecifieke elementen van KPSp geoperationaliseerd moeten worden. Voor een uitgebreide beschrijving van deze twee methodespecifieke onderdelen van de behandeling verwijzen wij naar het *Handboek KPSp* (De Jonghe e.a., 2014; Dekker e.a., 2014).

Adequate psychoanalytische steun (APS) als eerste belangrijke methodespecifieke factor is daarbij volgens de bedenkers van KPSp de adequate bevrediging van die ontwikkelingsbehoeften, waarvan te weinig aanwezig was in de vroege kindertijd (Dekker e.a., 2014; Dewald, 1994; Hendriksen, 2015; Summers & Barber, 2010). Van APS is volgens de KPSp-methode sprake als er aan vier voorwaarden voldaan is: 1) De therapeut beoogt

empathisch te valideren (Lachmann, 2008; Schechter, 2007), 2) er wordt validatie ervaren door de patiënt (Lachmann, 2008; Schechter, 2007), 3) de steun is progressiebevorderend (Dewald, 1994) en 4) de steun leidt tot objectrelationele dissonantie (Greenberg, 1986; Strachey, 1934).

Naast APS kent KPSF dus nog een tweede, belangrijk methodespecifiek element. Zoals aangegeven is de primaire focus van KPSF de relationele context van de onderliggende pathologie. Het gaat daarbij om relaties in het heden, om relaties van vroeger en om de relatie die iemand met zichzelf heeft. Bij de uitvoering van KPSF worden daarbij acht bespreekniveaus onderscheiden (De Jonghe e.a., 2013, 2014): klachten en symptomen, levensomstandigheden, interpersoonlijke problemen, interpersoonlijk patroon, eigen inbreng in het patroon, interne interpersoonlijke relaties, intrapersoonlijke relatie en overdracht. Het is de bedoeling om binnen de behandeling zoveel mogelijk aan te sluiten bij het niveau van klachtenbeleving van de patiënt (De Jonghe e.a., 2013, 2014).

Middels de methodespecifieke techniek van CARE worden hierbij concrete, actuele, emotioneel geladen interpersoonlijke interacties besproken op het passende bespreekniveau. CARE staat voor: Concreet, Actueel, Relationeel en Emotioneel. Wanneer deze vier elementen bij elkaar gebracht en geëxploreerd zijn, wordt vervolgens het verband onderzocht met de depressieve klachten. Dit kan het inzicht in het ontstaan en de instandhouding van de depressie verdiepen en zo progressie bevorderen (De Jonghe e.a., 2014; Dekker e.a., 2014). Voor deze twee methodespecifieke elementen en de genoemde techniek van het CARE uitvragen van interpersoonlijke interacties waarbij de patiënt somberheid ervaren heeft, werden verschillende items opgesteld. Hiervoor werd allereerst gebruikgemaakt van het *Handboek KPSF* (De Jonghe e.a., 2014), maar ook werd gebruikgemaakt van de uitkomsten van een aanvullende literatuurstudie en consensusbijeenkomsten van een expert-panel met senior KPSF-behandelaren. Het iteratieve proces van de verschillende expert-panel-bijeenkomsten diende daarbij als trainingssessies in het gebruik van het instrument. Na elke bijeenkomst werden de items en antwoordcategorien verfijnd. Daarnaast werd een scoringsinstructie geschreven teneinde de eenduidigheid van de scoring verder te vergroten. Het constructieproces leverde de KPSF-adherentieschaal, zoals weergegeven in [figuur 1](#):

Stap 2 – Bepalen van de psychometrische eigenschappen van het instrument

De interne consistentie van de uiteindelijke items werd bepaald. Interne consistentie is een maat voor de mate waarin items van een instrument één dan wel meerdere begrippen meten. Een veelgebruikte index is Cronbach's alpha (Cronbach, 1951).

Het instrument met verschillende items werd vervolgens ook getoetst op interbeoordelaarsbetrouwbaarheid. Interbeoordelaarsbetrouwbaarheid is een maat voor in hoeverre er overeenstemming is tussen verschillende beoordelaars, en dus in hoeverre scores over observatoren/beoordelaars kunnen worden gegeneraliseerd (Shrout, 1979). De methode uit het onderzoek van Ogrodniczuk en Piper (1999) diende daarbij als leidraad. Dit betekent dat om de interbeoordelaarsbetrouwbaarheid te bepalen in de huidige studie gebruik werd gemaakt van twee getrainde expert-beoordelaars. De beoordelaars hebben, net als in de studie van Ogrodniczuk en Piper (1999), vijftig behandeltrajecten beoordeeld aan de hand

Figuur 1 KPSP-adherentieschaal

1. APS; afdoende steunende interventies

1. Slecht
2. Onvoldoende
3. Voldoende
4. Goed

2. APS; Cliënt voelt zich gesteund

1. Slecht
2. Onvoldoende
3. Voldoende
4. Goed

3. APS; Steun is progressiebevorderend

1. Slecht
2. Onvoldoende
3. Voldoende
4. Goed

4. CARE

5. Slecht
6. Onvoldoende
7. Voldoende
8. Goed

Care: Concreet, Actueel, Relationeel, Emotioneel, Link depressie

Wat is het bespreekniveau?

1. klachten en symptomen
2. levensomstandigheden
3. interpersoonlijke problemen
4. interpersoonlijk patroon
5. eigen inbreng in het patroon
6. interne interpersoonlijke relaties
7. intrapersoonlijke relaties
8. overdracht

5. Therapeut houdt met zijn interventies rekening met het bespreekniveau

1. Slecht
2. Onvoldoende
3. Voldoende
4. Goed

6. Is er sprake van verstoring tegenoverdracht?

1. ja
2. nee

van het ontwikkelde instrument. Voor de analyse werd gebruikgemaakt van een tweeweg random-effectmodel met absolute overeenstemming voor enkelvoudige en gemiddelde metingen. De waarde voor overeenstemming bij enkelvoudige metingen geeft aan wat de interbeoordelaarsbetrouwbaarheid is wanneer er gebruikgemaakt wordt van de scores van één beoordelaar bij een beoordeling van de behandelintegriteit. De waarde voor overeenstemming bij gemiddelde metingen geeft aan wat de interbeoordelaarsbetrouwbaarheid is wanneer er gebruikgemaakt wordt van de gemiddelde scores van twee beoordelaars bij een beoordeling van de behandelintegriteit (McGraw & Wong, 1996). Volgens Fleiss (1986) betekent een $icc < 0,40$ een slechte overeenstemming tussen beoordelaars, tussen $0,40$ en $0,75$ een redelijke tot goede mate van overeenstemming en $> 0,75$ een sterke overeenstemming tussen beoordelaars. Om een zekere mate van homogeniteit van de te onderzoeken audio-opnames na te streven en om een vergelijking te kunnen maken tussen therapeuten is enkel gebruikgemaakt van behandeltrajecten uit de eerste fase van de D*Phase-studie. In de tweede fase van het onderzoek betreft de behandeling uitsluitend patiënten die na de eerste fase van behandeling non-responder zijn gebleken, waardoor potentieel een andere van invloed zijnde variabele geïntroduceerd wordt (Mohr e.a., 1990). In de tweede behandel fase kunnen deze patiënten daarbij ook doorstromen naar een onderzoekconditie waarbij dezelfde therapeut doorgaat met dezelfde behandel methode. Dat kan een bias opleveren, omdat dat invloed kan hebben op de behandelintegriteit. De behandelaar weet namelijk dat hij in de eerste fase een behandeling heeft gegeven die onvoldoende resultaat heeft gehad. Dit zou de neiging kunnen vergroten om andere interventies toe te voegen. Het wordt dan een meer eclectische therapievorm in plaats van zuiver KPSP.

Als laatste werd ten behoeve van de psychometrische eigenschappen van het instrument ook de constructvaliditeit in de vorm van de zogenaamde 'known group'-validiteit bepaald. De 'known group'-validiteit is de mate waarin een instrument onderscheid kan maken tussen twee groepen waarvan vooraf bekend is dat deze van elkaar verschillen (Davidson, 2014); bij dit onderzoek de groepen novice- en expert-behandelaren. Er kan bij novices in vergelijking met experts van uitgegaan worden dat deze verschillen op het gebied van de behandelintegriteit. Novices hebben immers minder training en supervisie genoten in de uitvoering van de methodespecifieke interventies. Deze manier van psychometrische toetsing is te vergelijken met de methode van Crane en anderen (2013). In de desbetreffende studie naar de ontwikkeling en validatie van een meetinstrument voor het toetsten van mindfulnessinterventies was de hypothese dat behandelaren met minder dan één jaar ervaring lager zouden score op de behandelintegriteit dan behandelaren met meer dan twee jaar ervaring. Bij de vergelijking van die twee groepen (resp. $N = 20$ en $N = 23$) vonden zij inderdaad dat de groep met meer dan twee jaar ervaring significant hoger scoorde op behandelintegriteit. Gezien het beperkter aantal behandelaren binnen de huidige studie ($N = 18$) zal ook bij dezelfde behandelaar op verschillende ervaringsniveaus de mate van behandelintegriteit worden bepaald om zo aan de benodigde vijftig samples te komen. Dat houdt in dat eenzelfde behandelaar voor kan komen zowel in de groep novices als in de groep experts, met als verschil zijn of haar ervaring met de methode op dat moment en verschil in hoeveelheid genoten supervisie. Alle behandelaren binnen onderhavig onderzoek hebben hierbij een driedaagse basistraining gevolgd in het geven van KPSP voor de behandeling van depressie, alvorens ze konden instromen als therapeut binnen het onderzoek.

Daarna volgden zij supervisie, eens per twee weken door een KPSp-supervisor. Dit supervisietraject liep gedurende de gehele betrokkenheid bij de overkoepelende studie door. De indeling met betrekking tot het onderscheid novice versus expert is hierbij uiteindelijk als volgt tot stand gekomen: de groep novices bestond uit behandelaren die startten na de basistraining en nog minder dan twintig sessies supervisie hadden gevolgd. De groep experts bestond uit behandelaren die al meer dan twintig sessies supervisie hadden gevolgd. Twee expert-beoordelaars hebben het beoordelingsinstrument voor onderhavig onderzoek uiteindelijk gebruikt om, onafhankelijk van elkaar, vijftig gesprekseenheden (geluidsfragmenten van ongeveer vijftien minuten) te beoordelen. Beoordelaar 1 is een GZ-psycholoog/psychoanalytisch (groeps)psychotherapeut, trainer en supervisor KPSp. Beoordelaar 2 is een klinisch psycholoog/psychoanalytisch psychotherapeut en supervisor KPSp. Beiden zijn coauteur van dit artikel. Het betrof fragmenten van vijftig verschillende cliënten, behandeld door in totaal achttien verschillende behandelaren (tabel 1). De betreffende sessie (1 t/m 16) waaruit het fragment werd geselecteerd, werd random bepaald middels de willekeurige selectiefunctie in Excel. Het fragment kwam daarnaast altijd uit het tweede kwartier van een sessie van 45 minuten. Dit zijn de fragmenten met doorgaans de meeste behandelinhoud.

Met een t-test voor onafhankelijke steekproeven werd getoetst of de groepen novices en experts significant van elkaar verschillen in de score op de behandelintegriteit. Er werd vooraf geen poweranalyse gedaan om te bepalen hoeveel metingen er nodig zijn om een moderaat tot groot verschil aan te tonen tussen de groepen novices en experts, in de mate van de behandelintegriteit. De steekproefgrootte stond namelijk vast, aangezien de looptijd van de studie direct verbonden was met de duur van de opleiding tot klinisch psycholoog van de eerste auteur. De onderzoeksgroep bestond hierbij uiteindelijk uit achttien KPSp-therapeuten: vier mannen en veertien vrouwen. Gemiddelde leeftijd: 38 jaar (sd = 8). Opleidingsniveau: veertien gezondheidszorgpsychologen, drie klinisch psychologen en een basispsycholoog, waarbij de therapeuten begonnen als novice binnen het onderzoek en bij voldoende lange participatie binnen het onderzoek, en na meer dan twintig supervisiezingingen te hebben gevolgd, in veel gevallen uitgroeiden tot zogenaamde expert-therapeut.

Resultaten

Stap 1 – Itemgeneratie en -selectie

De uiteindelijke versie van het instrument bestaat uit zes items, te vinden in [figuur 1](#). Items 1, 2 en 3 gaan over APS, item 4 over het toepassen van CARE, item 5 over het hanteren van het juiste bespreekniveau en item 6 over verstorende tegenoverdracht. Om de afname te vergemakkelijken zijn er voor de items 4 en 5 ondersteunende ‘checklists’ toegevoegd voor respectievelijk de elementen van CARE en voor de verschillende bespreekniveaus.

De eerste 5 items worden gescoord op een vierpunts likertschaal (1 = slecht, 2 = onvoldoende, 3 = voldoende, 4 = goed). Er is gekozen om geen neutrale score toe te voegen om zo een keuze te forceren en validiteit te vergroten (Bartam, 2007). Item 6, over verstorende

tegenoverdracht, wordt dichotoom gescoord (ja of nee). Het expert-panel was van mening dat verstorende tegenoverdracht die een duidelijk negatief effect heeft op de kwaliteit van APS, er wel of niet is.

Bij een eerste analyse van de variantie in antwoorden op de verschillende items, bleken de scores op item 6 met betrekking tot verstorende tegenoverdracht een zeer lage variantie te hebben. Over de vijftig audiofragmenten werd door beide beoordelaars in 39 gevallen gescoord dat er geen sprake was van verstorende tegenoverdracht. Verdere analyse was niet nuttig, omdat de spreiding zo beperkt was en dit item niet erg gevoelig leek voor het meten van verschillen of verandering. Item 6 werd derhalve niet verder meegenomen in de analyses. De variantie voor de items 1 t/m 5 was groter. Alle verdere analyses zijn uitsluitend uitgevoerd met deze items.

Stap 2 – Controle op de psychometrische eigenschappen

Statische analyses voor de betrouwbaarheid en validiteit van de items 1 t/m 5 werden uitgevoerd met SPSS versie 26 (IBM Corp., 2019).

Aan de hand van de scores van de twee beoordelaars werd de interne consistentie van de vijf items voor beide beoordelaars bepaald. Deze is berekend voor de items 1 t/m 5 omdat deze op dezelfde schaal zijn gescoord. De alphacoëfficiënten waren voor beide beoordelaars over deze items 0,90. Dit is een indicatie voor een zeer hoge interne consistentie (Cicchetti, 1994).

Voor de somscores van de items 1 t/m 5 is de interbeoordelaarsbetrouwbaarheid berekend door de intraklasse correlatiecoëfficiënt (icc) te bepalen met een 95% betrouwbaarheidsinterval (Shrout, 1979). De icc voor enkelvoudige metingen voor de items 1 t/m 5 was 0,42 en voor gemiddelde metingen was de icc 0,59. Beide uitkomsten vallen in de categorie 'redelijk tot goede overeenstemming'. Aanvullend zijn voor de interbeoordelaarsbetrouwbaarheid op itemniveau gewogen kwadratische kappacoëfficiënten voor categoriale data berekend met een betrouwbaarheidsinterval van 95% (Fleiss & Cohen, 1973). Gewogen kappacoëfficiënten tussen 0,21 tot 0,40 duiden op een lage overeenstemming, tussen 0,41 en 0,60 op een middelmatige overeenstemming en tussen 0,61 en 0,80 op een substantiële overeenstemming (Landis & Koch, 1977). Items 1, 2 en 3 vielen in de categorie 'lage overeenstemming' met waarden van respectievelijk 0,35, 0,37 en 0,25. De gewogen kappacoëfficiënten van de items 4 en 5 vielen in de categorie 'moderate overeenstemming' met waarden van respectievelijk 0,47 en 0,50. Het weglaten van het item met de laagste

Tabel 1 Gewogen kappa en ICC enkelvoudige en gemiddelde metingen

	Gewogen kappa	ICC enkelvoudige metingen	ICC gemiddelde metingen
Item 1	0,35	-	-
Item 2	0,37	-	-
Item 3	0,25	-	-
Item 4	0,47	-	-
Item 5	0,50	-	-
Totaalscore	-	0,42	0,59

gewogen kappacoëfficiënt (item 3) had vrijwel geen effect op de ICC van de somscore van de dan overgebleven items 1, 2, 4 en 5. De ICC over enkelvoudige metingen voor de items 1, 2, 4 en 5 samen was 0,44 en voor gemiddelde metingen was de ICC 0,62.

Het verschil in totaalscores van experts en novices werden voor beide groepen vergeleken om 'known groups'-validiteit te bepalen. Allereerst werden de gemiddelde totaalscores van de twee beoordelaars voor de groepen experts (n = 22) en novices (N = 28) vergeleken. Vervolgens werden apart de totaalscores van beide beoordelaars voor de groepen experts en novices vergeleken.

De gemiddelde totaalscore van beide beoordelaars samen bleek niet significant hoger voor de groep experts (M = 14,98, SD = 3,27), $t(48) = 1,334$ $p = 0,188$ dan voor de groep novices (M = 13,86, SD = 2,67). De effectgrootte was $d = 0,38$, hetgeen correspondeert met een klein tot middelgroot verschil (Cohan, 1992).

De totaalscore van beoordelaar 1 van de groep experts (M = 14,82, SD = 3,95) was niet significant hoger dan de gemiddelde totaalscore van de groep novices (M = 14,54, SD = 3,51), $t(48) = 0,267$ $p = 0,790$. De totaalscore van beoordelaar 2 van de groep experts (M = 15,14, SD = 3,83) bleek wel significant hoger dan de gemiddelde totaalscore van de groep novices (M = 13,18, SD = 2,72, $t(48) = 2,111$ $p = 0,040$).

Tabel 2 Gemiddelde totaalscores op de adherentieschaal KPSP van experts en novices

	Experts (N = 22) Mean (SD)	Novices (N = 28) Mean (SD)	t (df = 48)	p*
Beoordelaar 1 en 2	14,98 (3,95)	13,86 (2,67)	1,334	0,188
Beoordelaar 1	14,82 (3,95)	14,54 (3,51)	0,267	0,790
Beoordelaar 2	14,14 (3,83)	13,18 (2,72)	2,111	0,040*

* Tweezijdig.

Het statistisch onderscheidingsvermogen ('power') bij deze steekproefgrootte was 0,25. Dat betekent dat het een *underpowered sample* betrof. De kans is dus reëel dat in de populatie bestaande verschillen door een te kleine steekproefomvang niet zijn gedetecteerd.

De post-hoc poweranalyse liet zien dat wanneer aan de hand van het instrument een moderaat verschil aangetoond zou moeten worden (*effect size* 0,5), met een hoge power (0,8) (Cohen, 1992), er voor zowel de groep novices als experts, elk 64 sessies beoordeeld zouden moeten zijn.

Discussie

Dit artikel heeft het proces beschreven van de constructie van een beoordelingsinstrument voor het meten van behandelintegriteit van KPSP bij de behandeling van depressie: de KPSP-adherentieschaal. De psychometrische kwaliteit van vijf van de zes items blijkt hoopgevend. De interne consistentie is hoog, de interbeoordelaarsbetrouwbaarheid is

redelijk tot goed en er zijn aanwijzingen dat aan de hand van dit nieuwe instrument significant verschil aan te tonen is tussen scores van de groepen experts en novices. In het navolgende bespreken we eerst de psychometrische bevindingen en doen we suggesties voor vervolgonderzoek. Daarna zullen we het belang van deze studie plaatsen in een bredere context van het hedendaagse psychotherapieprocesonderzoek.

Bij de initiële analyse van de items bleek dat er bij een van de zes items sprake was van een zeer geringe spreiding. Het betrof item 6, het item waarmee de aanwezigheid van versturende tegenoverdracht werd gescoord. Weinig antwoordmogelijkheden betekent per definitie een lagere spreiding. In volgende versies van het instrument zou ervoor gekozen kunnen worden het construct te scoren op een zelfde schaal als de andere items, een vierpuntschaal. Dit kan voor meer spreiding zorgen en psychometrische eigenschappen van de lijst als geheel verbeteren. Er zou dan onderzocht kunnen worden of dit item ook bij het onderliggende construct hoort, door wederom de interne consistentie te bepalen.

De interne consistentie van de vijf overgebleven items was uitstekend. Dat suggereert dat deze items een samenhangend concept meten. Dit is in lijn met de theoretische aanname dat APS een techniek is in de therapeutische relatie, waarbij de therapeut resoneert met de meer of minder bewuste behoefte van de patiënt, deze adequaat bevredigt, in die zin dat het progressie bevordert door middel van CARE. KPSF gaat ervan uit dat dit vervolgens relationele dissonantie oftewel libidineuze dan wel agressieve schuring veroorzaakt hetgeen vervolgens kan leiden tot de internalisatie van een nieuwe relatie: de therapeutische relatie. Wanneer deze ervaring goed genoeg is, zal dit een positief effect hebben op de intrapersonlijke relatie. KPSF-therapeuten gebruiken de techniek van APS sinds deze als zodanig is geformuleerd en geoperationaliseerd in het *Handboek KPSF* (De Jonghe e.a., 2014). Het is echter voor het eerst dat een psychometrische samenhang dit specifieke concept ondersteunt. Het belang van steungevende interventies binnen een veilige therapeutische relatie is een kenmerkend element in focale psychodynamische psychotherapie voor depressie (Leichsenring & Schauenburg, 2014). Echter, voor zover wij weten, is de concrete operationalisering van het concept steun, zoals KPSF dat beschrijft, nieuw.

Volgens de normen geformuleerd door Fleis (1986) was de interbeoordelaarsbetrouwbaarheid over de overgebleven vijf items eveneens goed. Zowel de interbeoordelaarsbetrouwbaarheid voor enkelvoudige metingen als voor gemiddelde metingen van twee beoordelaars vielen in de categorie 'redelijke tot goede overeenstemming'. Hierbij moet worden opgemerkt dat de interbeoordelaarsbetrouwbaarheid voor enkelvoudige metingen net in de categorie 'redelijk tot goede overeenstemming' viel. Er zou daarom gekozen kunnen worden om te werken met de gemiddelde scores van twee beoordelaars. Dit is arbeidsintensiever, maar zorgt waarschijnlijk voor meer valide schatting van de bereikte behandelintegriteit, zoals gemeten met dit instrument.

Op itemniveau viel de interbeoordelaarsbetrouwbaarheid van de afzonderlijke items in de categorieën 'enigszins overeenstemming' en 'gemiddelde overeenstemming'. Item 3 'steun is progressiebevorderend' had de laagste waarde. Het weglaten van item 3 had echter zeer beperkte invloed op de totale interbeoordelaarsbetrouwbaarheid van de dan overgebleven items. Theoretisch gezien is het feit of steun al dan niet progressiebevorderend is, de uitkomst van APS. Het is denkbaar dat het daarom zo laag scoort: bij de andere twee items meten we het proces. Mogelijk is dat makkelijker te beoordelen. Het is aan te raden

om beoordelaars in het vervolg meer te trainen in de beoordeling van dit item en het te behouden voor de adherentieschaal.

Wij adviseren in het algemeen dat beoordelaars intensief getraind worden in het gebruik van dit instrument. We zagen in de trainingssessies van experts dat de overeenstemming na iedere sessie toenam, zonder dat er tijdens de drie bijeenkomsten een plafondeffect leek op te treden. De reden hiervoor zou erin kunnen liggen dat APS een construct omvat dat werkelijk een theoretisch geheel is. De items zijn wel los van elkaar te scoren, maar hangen in hoge mate met elkaar samen. Bovendien is het zo dat APS afhankelijk is van de egosterkte van de patiënt. APS is technisch anders bij een vermoede borderline-persoonlijkheidsorganisatie dan bijvoorbeeld bij een neurotische organisatie (Hendriksen, 2015). Het scoren en beoordelen van APS vraagt dus om de capaciteit hier in zeer korte tijd kennis van te nemen, een klinische hypothese te vormen en vervolgens de interventies van de therapeut en de reacties van de patiënt op deze wijze te toetsen en te beoordelen. Binnen de trainingen is het dan ook aan te raden regelmatig elkaars beoordelingen te toetsen en te vergelijken, in het bijzijn van een KPSP-expert. Zo werd dit ook gedaan in de studie naar KPSP versus KPSP plus farmacotherapie (De Jonghe e.a., 2004) om te zorgen voor betrouwbare afname van de *Hamilton Depression Rating Scale* (HDRS; Hamilton, 1967) als uitkomstmaat. In deze studie werd de scoring van de onderzoeksassistenten wekelijks met een expert-KPSP-rater geoefend en getoetst op betrouwbaarheid.

Een beperking van deze studie is de beperkte steekproefomvang die gebruikt is. Uiteindelijk bleek hierbij ook dat de steekproef te klein was om overtuigende uitspraken te kunnen doen over de constructvaliditeit, in de vorm zoals die onderzocht is. Met de huidige steekproefomvang werd namelijk geen significant verschil gevonden in de scores tussen experts en novices wanneer de gemiddelde scores van twee beoordelaars werden geanalyseerd. De gevonden effect size correspondeerde met een klein tot moderaat verschil (Cohen, 1992). Er lijkt dus wel een tendens te zijn dat er kleine verschillen zijn tussen beide groepen. Echter, alleen uit de scores van beoordelaar 2, bleek een *significant* verschil tussen de scores van de twee groepen. De scores van beoordelaar 1 voor de twee groepen verschilde niet significant. Een mogelijke verklaring kan zijn dat beoordelaar 1 vanaf het begin als trainer en als supervisor betrokken is geweest bij de overkoepelende studie. Zij kent alle deelnemende therapeuten. Zij heeft ze getraind en gesuperviseerd, wat kan leiden tot bias. In vervolgonderzoek zou het een aanbeveling zijn om, mits praktisch uitvoerbaar, beoordelaars te gebruiken die minder of geen banden hebben met de lopende studie en deelnemende behandelaren. In verband met de lage power kan ook een toevalstreffer niet uitgesloten worden (Cremers, Wager & Yarkoni, 2017). Bij vervolgonderzoek is een grotere sample nodig in de vorm van meer behandeltrajecten. Wij hebben hierbij na de validiteitsbepaling nog een post-hocanalyse gedaan, om gericht adviezen te kunnen geven met het oog op de toekomstige doorontwikkeling van dit instrument en de gewenste sample size. Deze post-hoc poweranalyse laat zien dat wanneer aan de hand van het instrument een moderaat verschil aangetoond zou moeten worden, met een hoge power, er voor zowel de groep novices als experts, elk 64 sessies beoordeeld zouden moeten worden.

Een andere beperking van deze studie is dat ten behoeve van bepaling van de constructvaliditeit, vanwege het beperkte beschikbare aantal behandelaren, eenzelfde behandelaar

zowel als novice als expert kan voorkomen. Dit afhankelijk van ervaring opgedaan tijdens het doorlopen van behandeltrajecten binnen de studie. Dit zou de grootte van het gevonden verschil tussen novices en expert-therapeuten in scores op het instrument kunnen drukken. In het vervolg is het wenselijk dat behandelaren enkel in één categorie voorkomen om zo een zuiverder sample te hebben. Daarom is een van de belangrijkste adviezen voor vervolgonderzoek dan ook het gebruik van een grotere sample aan therapeuten, die bij start van het onderzoek misschien ook al meer verschillen in mate van deskundigheid op het gebied van het toepassen van KPSP. Met een dergelijke sample kan de nu waargenomen tendens (kleine tot moderate verschillen tussen beide groepen), verder worden onderzocht.

Conclusie

Met deze studie is een wetenschappelijk en klinisch relevante stap gezet in de ontwikkeling van een betrouwbaar en valide instrument voor het meten van de behandelintegriteit bij KPSP. Het theoretische concept Adequate Psychoanalytische Steun krijgt hiermee een eerste psychometrische onderbouwing. Dit is klinisch van groot belang, omdat het element steun in focale psychodynamische behandelingen voor depressie weliswaar van groot belang is, de operationalisatie ervan is lastig gebleken. Met deze samenhangende items kunnen we therapeuten nog beter trainen, handvatten geven en zichzelf laten toetsen op wat ze doen, zodat ze zichzelf blijvend kunnen verbeteren. Dat is hard nodig om het succespercentage van psychotherapie, en dus van KPSP, te verbeteren. In combinatie met de vijf reeds uitgevoerde RCT's waarin gekeken is naar de effectiviteit van KPSP (De Jonghe e.a., 2001, 2004; Dekker e.a., 2005, 2008; Driesen e.a., 2013) kan dit de status van KPSP als serieus alternatief in de behandeling van depressie nog verder verstevigen en daarmee het palet aan mogelijkheden voor de behandeling van deze ernstige aandoening vergroten. Nog belangrijker misschien wel: de beschikbaarheid van een psychometrisch deugdzzaam instrument voor het meten van de behandelintegriteit van KPSP zou ook het vertrekpunt kunnen vormen voor onderzoek naar de relatie tussen de kwaliteit van uitvoering van de therapie en de behandeluitkomst. Dergelijk onderzoek kan daarmee mogelijk helpen licht te schijnen op het raadsel van de specifieke werkingsmechanismen van onder andere deze vorm van behandeling. Bij een aantoonbare relatie tussen de kwaliteit van uitvoering van de therapie en de behandeluitkomst kan dit instrument bijdragen aan nog effectievere behandeling van depressie middels KPSP. Dit bijvoorbeeld als het instrument ook voor opleiding- en trainingsdoeleinden ingezet wordt en gebruikt wordt om te borgen dat behandelintegriteit van uitgevoerde KPSP-behandelingen, voldoende groot is.

DANKBETUIGING

Tijdens een intensieve periode van vier jaar heb ik als eerste auteur onderzoek mogen doen waarin de ontwikkeling van een instrument voor het meten van behandelintegriteit bij de uitvoering van Kortdurende Steungevende Psychoanalytische Psychotherapie voor depressie centraal stond. Ik heb met behulp van zeer bevlogen en zeer bekwame mensen die bij dit artikel als coauteur vermeld staan een eerste mooie stap mogen zetten in de ontwikkeling van een dergelijk instrument. Zonder de steun, de raad, het geduld en de inzet van deze mensen was het niet gelukt. Ik hecht er dan ook waarde aan om hen via deze weg expliciet te bedanken.

LITERATUUR

- Bhar, S.S., & Beck, A.T. (2009). Treatment integrity of studies that compare short-term psychodynamic psychotherapy with cognitive-behaviour therapy. *Clinical Psychology: Science and Practice, 16*, 370-378.
- Bartam, D. (2007). Increasing validity with forced-choice criterion measurement formats. *International Journal of Selection and Assessment, 15*, 263-72.
- Beekman, A.T., Penninx, B.W., Deeg, D.J., Beurs, E. de, Geerling, S.W., & Tilburg, W. van (2002). The impact of depression on the well-being, disability and use of services in older adults: a longitudinal perspective. *Acta Psychiatrica Scandinavica, 105*, 20-27.
- Bijl, R.V. & Ravelli, A. (2000). Current and residual functional disability associated with psychopathology: findings from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Psychological Medicine, 30*, 657-668.
- Chatoor, I., & Krupnick, J. (2001). The role of non-specific factors in treatment outcome of psychotherapy studies. *European Child & Adolescent Psychiatry, 10 (Suppl 1)*, 19-25.
- Chambless, D.L., & Hollon, S.D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7-18.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- Crane, R.S., Eames C., Kuyken, W., Hastings, R.P., Williams, J.M.G., Bartley, T., e.a. (2013). Development and validation of the Mindfulness-Based Interventions – Teaching Assessment Criteria (MBI:TAC). *Assessment, 20*, 1-8.
- Cronbach, L.J., (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.
- Cuijpers, P. (2015). Psychotherapies for adult depression: recent developments. *Current Opinion in Psychiatry, 28*, 24-29.
- Cremers, H.R., Wager, T.D., & Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PLOS ONE, 12*, 1-20.
- Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 64*, 284-290.
- Davidson, M. (2014). *Known-Groups Validity. Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht: Springer.

- Dekker, J.J.M., Hendriksen, M., Kool, S., Bakker, L., Driessen, E., Jonghe, F.E.R.E.R. de, e.a. (2014). Growing evidence for psychodynamic therapy for depression. *Contemporary Psychoanalysis*, 50, 131-155.
- Dekker, J.J.M., Molenaar, P., Kool, S., Aalst, G. van, Peen, J., & Jonghe, F.E.R.E.R. de (2005). Dose-effect relations in time-limited combined psycho-pharmacological treatment for depression. *Psychological Medicine*, 35, 47-58.
- Dekker, J.J.M., Koelen, J.A., Van, H.L., Schroeven, R.A., Peen, J., Hendriksen, M., e.a., (2008). Speed of action: The relative efficacy of short psychodynamic supportive psychotherapy and pharmacotherapy in the first 8 weeks of a treatment algorithm for depression. *Journal of Affective Disorders*, 1, 183-188
- Dewald, P.A. (1994). *The supportive and active psychotherapies. A dynamic approach*. Northvale, New Jersey, Londen: Jason Aronson Inc.,
- Driessen, E., Cuijpers, P., Maat, S.C.M. de, Abbass, A.A., Jonghe, F. de, & Dekker, J.M. (2010). The efficacy of short-term psychodynamic psychotherapy for depression: A meta-analysis. *Clinical Psychology Review*, 30, 25-36.
- Driessen, E., Van, H.L., Don, F.J., Peen, J., Kool, S., Westra, D., e.a. (2013) The efficacy of cognitive-behavioural therapy and psychodynamic therapy in the outpatient treatment of major depression: A randomized clinical trial. *The Journal of American Psychiatry*, 9, 1041-1050.
- Fleiss, J.L. (1986). *The design and analysis of clinical experiments*. New York: John Wiley & Sons.
- Fleiss, J.L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33, 613-619.
- Greenberg, J.R. (1986). Theoretical models and the analyst's neutrality. *Contemporary Psychoanalysis*, 22, 87-106.
- Greden, J.F. (2001). The burden of disease for treatment-resistant depression. *Journal of Clinical Psychiatry*, 62, 26-31.
- Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, 6, 278-296.
- Hendriksen, M. (2015). Adequate Psychoanalytische Steun voor patiënten met narcisme en comorbide depressie. *Tijdschrift voor Psychoanalyse*, 21, 285-294.
- IBM Corp. (2019). *IBM SPSS Statistics for Windows, Version 26.0*. Armonk, NY: IBM Corp.
- Jonghe, F.E.R.E.R. de, Kool, S., Aalst, G. van, Dekker, J.J.M., & Peen, J. (2001). Combining psychotherapy and antidepressants in the treatment of depression. *Journal of Affective Disorders*, 64, 217-229.
- Jonghe, F.E.R.E.R. de, Maat, S. de, Van, H.L., Hendriksen, M., Kool, S., Aalst, G. van e.a. (2013). Short-term Psychoanalytic Supportive Psychotherapy for Depressed Patients. *Psychoanalytic Inquiry*, 33, 1-11.
- Jonghe, F.E.R.E.R. de, Hendriksen, M., Maat, S. de, Dekker, J.J.M., Dil, L., Noteboom, A. e.a. (2014). *Handboek Kortdurende Psychoanalytische Steun gevende Psychotherapie*. Amsterdam: Benecke.
- Jonghe, F.E.R.E.R. de, Hendriksen, M., Aalst, G. van, Kool, S., Peen, J., Van, H.L., e.a. (2004). Psychotherapy and combined therapy (pharmacotherapy plus psychotherapy) in the treatment of depression. *British Journal of Psychiatry*, 185, 37-45.
- Lachmann, F.M. (2008). *Transforming narcissism. Reflections on empathy, humor and expectations*. New York, London: The Analytic Press.

- Landis, J.R., & Koch G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Leichsenring, F. & Schauenburg, H. (2014). Empirically supported methods of short-term psychodynamic therapy in depression - towards an evidence-based unified protocol. *Journal of Affective Disorders*, 169, 128-143.
- McGraw, K.O., & Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30-46.
- McLeod, B.D. (2009). Understanding why therapy allegiance is linked to clinical outcomes. *Clinical Psychology: Science and Practice*, 16, 69-72.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.
- Miggels, M.F., Klooster, P. ten, Bremer-Hoeve, S., Dekker, J.J.M., Huibers, M.J.M., Reefhuis, E., e.a. (2021). The D*Phase-study: a pragmatic two-phased, randomised controlled non-inferiority trial addressing non-response; comparing cognitive behavioural therapy and short-term psychodynamic supportive psychotherapy for major depression. *BMC Psychiatry*, 21, 233.
- Mohr, D.C., Beutler, L.E., Engle, D., Shoham-Salomon, V., Bergan, J., Kaszniak, A.W. e.a. (1990). Identification of patients at risk for nonresponse and negative outcome in psychotherapy. *Journal of Consulting and Clinical Psychology*, 58, 622-628.
- Nezu, A., & Nezu, C. (2008). *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychosocial interventions*. New York: Oxford University Press.
- Norcross, J.C. & Rogan, J.D. (2013). Psychologists conducting psychotherapy: Current practices and historical trends among division 29 members. *Psychotherapy*, 50, 490-495.
- Ogrodniczuk, J.S. & Piper, W.E. (1999). Use of transference interpretations in dynamically oriented individual psychotherapy for patients with personality disorders. *Journal of Personality Disorders*, 13, 297-311.
- Picardi, A. & Gaetano, P. (2014). Psychotherapy of mood disorders. *Clinical Practice & Epidemiology in Mental Health*, 10, 140-158.
- Romijn, G.A., Ruiter, M. & Smit, F. (2008). *Meer effect met depressiepreventie? Strategieën voor publieksvoorlichting, vroegherkenning en terugvalpreventie*. Utrecht: Trimbos-instituut.
- Schechter, M. (2007). The patient's experience of validation in psychoanalytic treatment. *Journal of the American Psychoanalytic Association*, 55, 105-130.
- Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychology Bulletin*; 86, 420-428.
- Spijker, J., Bockting, C., Meeuwissen, J.A.C., Vliet, I. M. van, Emmelkamp, P.M.G., Hermens, e.a. (2013). *Multidisciplinaire richtlijn Depressie (Derde revisie): Richtlijn voor de diagnostiek, behandeling en begeleiding van volwassen patiënten met een depressieve stoornis*. Utrecht: Trimbos-instituut.
- Summers, R.F., & Barber, J.P. (2010). *Psychodynamic therapy: a guide to evidence-based practice*. New York, London: Guilford Press.
- Strachey, J. (1934). The nature of the therapeutic action of psychoanalysis. *The International Journal of Psychoanalysis*, 15, 127-159.

Swift, J.K., Greenberg, R.P., Tompkins, K.A., & Parkin, S.R. (2017). Treatment refusal and premature termination in psychotherapy, pharmacotherapy, and their combination: A meta-analysis of head-to-head comparisons. *Psychotherapy, 54*, 47-57.

Trimbos-instituut (2016). Depressie, enkele cijfers. Online van <https://www.trimbos.nl/aanbod/webwinkel/product/inf016-depressie-enkele-cijfers>.

Vallis, T.M., Shaw, B.F., & Dobson, K.S. (1986). The Cognitive Therapy scale: psychometric properties. *Journal of Consulting and Clinical Psychology, 54*, 381-385.

Abstract

Psychometric evaluation of an instrument to measure adherence in the practice of Short-term Psychoanalytic Supportive Psychotherapy (KPSP)

Introduction: Short-term Psychoanalytic Supportive Psychotherapy (SPSP) is a treatment method for depression, comparable to Cognitive Behavioral Therapy (CBT) in effectiveness. Claims regarding non-inferiority are hindered by the fact that there is no instrument available to determine the adherence of SPSP. This article will explore the development of such an instrument. Method: Initial items were formulated. The instrument's psychometric quality was determined. Results: The internal consistency found for the SPSP adherence scale is high and the inter-rater reliability is reasonable to good. Furthermore, it seems possible to demonstrate a significant difference between scores of more experienced and less experienced therapists. Conclusion: A reliable and valid instrument to measure the adherence for SPSP seems possible. However, more research into the psychometric features of the instrument is necessary, preferable with a larger sample.