

# GEDRAGSTHERAPIE

*Tijdschrift voor gedragstherapie  
& cognitieve therapie*



JAARGANG 57 **NUMMER 2** JUNI 2024

Het tijdschrift *Gedragstherapie* wordt uitgegeven door Uitgeverij Boom in opdracht van Stichting Tijdschrift voor Gedragstherapie, waarin participeren de Vereniging voor Gedragstherapie en Cognitieve Therapie en de Vlaamse Vereniging voor Gedragstherapie. Het tijdschrift verschijnt viermaal per jaar: in maart, juni, september & december

### **Doelstelling**

Tijdschrift *Gedragstherapie* heeft als doelstelling het publiceren van wetenschappelijk onderzoek (inclusief psychometrisch onderzoek), theoretische artikelen, literatuuroverzichten, case-studies en boekbesprekingen op het gebied van cognitieve gedragstherapie.

### **Redactie**

Elske Saleminck, *hoofdredacteur*  
Hermien Elgersma  
Arnold van Emmerik  
Kristof Hoorelbeke  
Jaap Lancee  
Miriam Lommen  
Margreet Oorschot  
Sara Scheveneels  
Anouk Vanden Bogaerde  
Leentje Vervoort  
Janneke Wolters, *redactiesecretaris*

### **Redactieraad**

dr. Anika Bexkens, prof. dr. Claudy Bockting,  
prof. dr. Paul Boelen, prof. dr. Theo K.  
Bouman, prof. dr. Caroline Braet, dr.  
Suzanne van Bronswijk, dr. Vincent van  
Bruggen, prof. dr. Iris Engelhard, dr. Klaske  
Glashouwer, prof. dr. Dirk Hermans, prof.  
dr. Jan de Houwer, prof. dr. Jos de Keijser, dr.  
Rianne de Kleine, em. prof. dr. Kees  
Korrelboom, prof. dr. Ernst Koster, prof. dr.  
Marije van der Lee, prof. dr. Sandra  
Mulken, prof. dr. Saskia van der Oord, dr.  
Kaya Peerdeman, prof. dr. Rudi De Raedt,  
prof. dr. Filip Raes, dr. Jeffrey Roelofs, prof.  
dr. Agnes Scholing, prof. dr. Christiaan  
Schotte, prof. dr. Anne Speckens, prof. dr.  
Philip Spinhoven, prof. dr. Johan Vlaeyen,  
dr. Marisol Voncken, prof. dr. Reinout  
Wiers, dr. Lidewij Wolters

### **Redactieadres**

Weltevreden 61  
3811 NR Amersfoort  
redactie@tijdschriftgedragstherapie.nl

### **Website**

[www.tijdschriftgedragstherapie.nl](http://www.tijdschriftgedragstherapie.nl)  
[tvg.boomtijdschriften.nl](http://tvg.boomtijdschriften.nl)

### **Uitgever**

Uitgeverij Boom  
Postbus 15970, 1001 NL Amsterdam

### **Advertenties**

[klantenservice@boompsychologie.nl](mailto:klantenservice@boompsychologie.nl)

### **Typografische verzorging**

LINE UP boek en media bv, Groningen  
(opmaak)  
René van der Vooren, Amsterdam  
(basisontwerp)

© 2024

Boom uitgevers Amsterdam/  
Stichting Tijdschrift voor Gedragstherapie

### **Aanwijzingen voor auteurs**

De auteursrichtlijnen vindt u op  
[www.tijdschriftgedragstherapie.nl](http://www.tijdschriftgedragstherapie.nl).



.....

De redactie is, samen met anderen, bezig met een aantal vernieuwingen rondom het tijdschrift. In dat kader hielden we in december 2023 een enquête over mogelijke digitalisering van het tijdschrift. In korte tijd vulden 1998 mensen die in. Dat laat ons al zien dat jullie betrokken zijn en dit een relevant onderwerp vinden, fijn! Ik kan ook al de belangrijkste resultaten delen: 12,8% heeft voorkeur voor een papieren versie, 56,7% voor een digitale versie, en 30,5% voor zowel een papieren als een digitale versie van het tijdschrift. Hier gaan we natuurlijk mee aan de slag.

Tegelijkertijd zijn we al een aantal stappen aan het zetten om kennis makkelijker te delen, meer bekendheid aan bijdragen te geven en informatie makkelijker toegankelijk te maken. Zo bieden we van elk nummer één bijdrage aan in *open access* en zijn na een jaar alle bijdragen voor iedereen toegankelijk. Ook publiceren we artikelen als *online first*-artikelen en kom je ons tegen op LinkedIn.

Het eerste artikel in dit nummer is de open access-bijdrage. Yannick Vander Zwalm en collega's geven daarin een mooi overzicht van de effectiviteit van adaptieve cognitieve training bij depressieve klachten. Het volgende artikel, van Edwin de Beurs en collega's, is een mooi voorbeeld van informatie toegankelijker maken. Het artikel is een pleidooi om universele schalen (*t*-scores en percentiel rangorde scores) te berekenen en te gebruiken tijdens de bespreking van testuitslagen met cliënten. Dit klinkt mogelijk ingewikkeld, want er moeten scores berekend worden. Het mooie is dat de auteurs een online Excelfile gemaakt hebben waarin ruwe scores ingevoerd kunnen worden die dan automatisch omgezet worden naar de universele maten. We delen zowel de link als de QR-code naar die file. De derde bijdrage is een pilotstudie van Anne Schimmel en Julie Krans over de inzet van virtual reality voorafgaand aan exposure.

Het laatste nummer van vorig jaar was een themanummer over Skinner en daarin schreven Jan De Houwer en Yannick Boddez een artikel over regelgeleid gedrag en *relational frame theory*. Dit riep verschillende vragen en gedachten op bij Erik ten Broeke en Kees Korrelboom die zij prikkelend opschreven. De auteurs van het stuk hebben hier weer een reactie op gegeven en beide stukken zijn nu als Forumbijdrage te vinden in dit nummer. De rubriek Voor u gelezen ten slotte gaat over de vraag of patiënten die psychofarmaca (medicatie) gebruiken anders reageren op psychologische behandeling voor depressie dan patiënten die deze niet gebruiken.

Elske Salemink  
Hoofdredacteur

# *Effectiviteit van een adaptieve cognitieve training bij depressieve klachten*

Een systematische review en meta-analyse<sup>1</sup>

---

YANNICK VANDER ZWALMEN, DAVID DEMEESTER,  
KRISTOF HOORELBEKE & ERNST KOSTER

## *Samenvatting*

---

De laatste jaren heeft cognitieve-controletraining (CCT) een opmars gemaakt als interventie om cognitieve problemen te remediëren en depressieve symptomen te verminderen. Een beloftevolle taak om cognitieve controle te trainen is de adaptieve Paced Auditory Serial Addition Task (aPASAT).

In deze systematische review en meta-analyse werden de effectiviteit van aPASAT-training en mogelijke moderatoren nagegaan. Hiervoor werden de online databases PsycINFO, MEDLINE, Embase, Web of Science en Cochrane Library doorzocht op wetenschappelijke studies die de effecten van een aPASAT-training op depressieve klachten en ruminatie onderzochten. Negentien studies met in totaal 1255 participanten werden geïncludeerd, waaronder studies met depressieve patiënten, voorheen depressieve mensen, risicopopulaties en gezonde participanten.

We vonden significante effecten na de training voor depressieve en ruminatieve klachten, zowel onmiddellijk na de training als op langere termijn. De effectgrootte was klein tot middelmatig. Subgroepanalyses suggereren een hogere effectiviteit in populaties van (voorheen) depressieve mensen en risicogroepen onmiddellijk na de training, in vergelijking met gezonde participanten.

*Trefwoorden: cognitieve training, cognitieve controle, adaptieve Paced Auditory Serial Addition Task, depressie, meta-analyse*

1 Dit artikel is een vertaalde en deels bewerkte versie van Vander Zwalm et al. (2023).

**Kernboodschappen voor de klinische praktijk**

- ▶ Cognitieve remediatie door middel van aPASAT-training is een geschikte interventie in de context van depressie.
- ▶ Deze training lijkt het meest zinvol voor individuen met verhoogde depressieve kwetsbaarheid, zoals depressieve patiënten, voorheen depressieve individuen en mensen met een verhoogd risico op het (her)optreden van een depressieve episode (bijvoorbeeld piekeraars).
- ▶ Meer onderzoek is noodzakelijk om te bepalen hoeveel sessies er nodig zijn om optimale effecten van aPASAT-training te bekomen, alsook om de effectiviteit van een gecombineerde behandeling met andere interventies na te gaan.

119

**INLEIDING**

.....

Depressie is een veelvoorkomende en ernstige psychiatrische aandoening, die zowel aanzienlijk individueel lijden als hoge maatschappelijke kosten met zich meebrengt (Kessler, 2012). Veelgebruikte behandelingen voor depressie omvatten farmacologische interventies, zoals antidepressieve medicatie, en psychotherapie, zoals cognitieve gedragstherapie (CGT). Dit zijn matig effectieve interventies in de acute fase (Cipriani et al., 2018; Gautam et al., 2020), maar ze slagen er vaak niet in om terugval van depressie op lange termijn te voorkomen (Lorimer et al., 2019). Depressie gaat gepaard met heel wat verstoorde cognitieve processen, waaronder verstoorde informatieverwerking en emotieregulatie (Gotlib & Joormann, 2010; LeMoult & Gotlib, 2019). Onderzoek heeft bijvoorbeeld aangetoond dat depressie geassocieerd is met cognitieve inflexibiliteit (Rose & Ebmeier, 2006) en verhoogd rumineren (Lewis et al., 2019; Nolen-Hoeksema, 2000). Interessant is dat cognitieve problemen in verband zijn gebracht met een beperkte respons op antidepressieve behandelingen (Story et al., 2008) en met ongunstige uitkomsten van CGT-behandelingen (Porter et al., 2015).

Een van de manieren om dergelijke cognitieve stoornissen aan te pakken, is door het gebruik van cognitieve-controletraining (CCT), dat beloftevolle resultaten heeft laten zien op onder andere ruminatie (Hoorelbeke et al., 2015; Siegle et al., 2007) en op depressieve symptomatologie (Iacoviello et al., 2014; Siegle et al., 2007). CCT vindt doorgaans plaats via een adaptieve computertaak, die vaak toegankelijk is via het internet. Daardoor kan de training gemakkelijk worden verspreid en flexibel worden ingezet. Andere voordelen van CCT ten opzichte van bestaande behandelingen voor depressie zijn de afwezigheid van bijwerkingen, in tegenstelling tot bijvoorbeeld gangbare farmacologische interventies (Motter et al., 2016). CCT kan op verschillende manieren worden geoperationaliseerd, zoals met de Eriksen flankertaak (Eriksen & Eriksen, 1974; bijvoorbeeld: Cohen et al., 2015) of de n-back-taak (Kirchner, 1958; bijvoorbeeld: Sari et al., 2016).

Een van de meest gebruikte CCT-procedures in de context van depressie is gebaseerd op de Paced Auditory Serial Addition Task (PASAT; Gronwall, 1977; Tombaugh, 2006). Tijdens de PASAT krijgen deelnemers auditieve stimuli te horen (getallen van 1 tot 9) en wordt hun gevraagd om continu te reageren met de som van de laatste twee gehoorde cijfers. Dit vereist dat deelnemers deze auditieve stimuli in hun werkgeheugen houden, de optelsom uitvoeren en eerdere (dat wil zeggen: niet langer relevante) stimuli, zoals eerder gehoorde auditieve stimuli en vorige sommen, uit het werkgeheugen verwijderen. Deze taak werd in 2007 aangepast om cognitieve controle te trainen in de context van depressie (Siegle et al., 2007) en staat sindsdien bekend als de adaptieve PASAT. Terwijl tijdens de niet-adaptieve PASAT deelnemers auditieve stimuli gepresenteerd krijgen in een vast tempo (bijvoorbeeld 3000 ms), wordt binnen de adaptieve PASAT (aPASAT) cognitieve controle getraind door het tempo van stimuluspresentatie continu bij te sturen op basis van de prestatie van de participant. Na vier opeenvolgende correcte antwoorden wordt de periode tussen twee opeenvolgende stimuli met 100 ms verlaagd, waardoor de participant de rekenkundige taak steeds sneller dient te volbrengen. Na vier opeenvolgende *incorrecte* responsen wordt de taak echter weer met 100 ms vertraagd.

Meerdere studies hebben de effecten van aPASAT-training op depressieve symptomen onderzocht. Siegle en collega's (2007) waren de eersten die een aPASAT-training in combinatie met een andere aandachtstraining gebruikten bij een steekproef van depressieve personen, waarbij de training toegevoegd werd aan *treatment as usual* (TAU). Zij observeerden een daling in depressieve symptomatologie en ruminatie in de CCT+TAU-conditie, in vergelijking met een TAU-controleconditie. In latere publicaties werden de gunstige resultaten van aPASAT-training op depressieve ruminatie (Hoorelbeke & Koster, 2017; Siegle et al., 2014) en depressieve klachten gerepliceerd, waar men ook dalingen vond in klinische zorgbehoeften na een jaar (Siegle et al., 2014). In lijn hiermee toont recent onderzoek aan dat aPASAT-training het risico op terugval in depressie in de periode van één jaar na de training reduceert (Hoorelbeke et al., 2021).

Deze bevindingen suggereren dat aPASAT-training zich richt op cognitieve stoornissen geassocieerd met depressie, waardoor CCT bijdraagt tot een reductie in depressieve klachten en preventie van terugval in depressie, al is evidentie voor de effectiviteit van aPASAT-training niet sluitend (Ferrari et al., 2021). Hoewel meerdere studies de effecten van een aPASAT-training nagingen in de context van depressie (bijvoorbeeld: Hoorelbeke & Koster, 2017; voor een overzicht, zie: Koster et al., 2017), ontbreekt meta-analytisch onderzoek naar de effectiviteit van aPASAT-training voor depressieve kwetsbaarheid. Het voordeel van het uitvoeren van een meta-analyse op één specifieke operationalisatie van CCT is dat de taakspecifieke heterogeniteit laag is, waardoor de onzekerheid die gepaard gaat met het onderzoeken van verschillende vormen van CCT in één meta-analyse wordt vermeden en de uitkomsten zo aan één vorm van CCT kunnen worden toegeschreven.

*Theoretische achtergrond*

121

De mechanismen waardoor CCT de kwetsbaarheid voor depressie beïnvloedt zijn nog niet voldoende begrepen. Modellen van ruminatie stellen dat individuen minder aandachtscontrole uitoefenen over de verwerking van emotionele, zelfrelevante informatie (Gotlib et al., 2004; Koster et al., 2011). Wanneer men geconfronteerd wordt met stress, kan negatief denken optreden. De meeste mensen zijn na enige tijd in staat om hun aandacht los te koppelen van deze negatieve gedachten en gevoelens. Wanneer het iemand echter niet lukt om de aandacht los te koppelen van irrelevante negatieve gedachten en er cycli van ruminatie ontstaan, kan dit leiden tot de ontwikkeling van een depressieve episode. Een aanhoudende negatieve stemming is een kernfactor van depressie en er is aangetoond dat ruminatie een belangrijke mediërende rol speelt bij het koppelen van cognitieve-controlestoornissen aan het heroptreden van depressieve symptomen (Demeyer et al., 2012). Door cognitieve controle te trainen, zou men mogelijk het loskoppelen van de aandacht aan het repetitief negatief denken kunnen vergemakkelijken, waardoor men minder lang in een toestand van aangehouden negatief affect blijft. In lijn met deze visie toont voorgaand onderzoek in voorheen depressieve individuen aan dat cognitieve vooruitgang na aPASAT-training predictief is voor een daling in ruminatie, wat vervolgens een verdere daling in residuele depressieve klachten voorspelt (Hoorelbeke & Koster, 2017).

In CCT-studies vertonen participanten vaak aanzienlijke heterogeniteit in behandelrespons. Verschillende factoren kunnen hierbij een rol spelen, zoals de context waarin CCT wordt uitgevoerd. Bovendien is momenteel niet helemaal duidelijk voor wie een aPASAT-training het meest effectief is om de kwetsbaarheid voor depressie te verminderen. Dit heeft mede te maken met de sterke heterogeniteit in voorgaand onderzoek inzake de gebruikte populatie, en met name de mate waarin participanten een depressieve kwetsbaarheid vertoonden. Zo werden effecten van aPASAT-training op repetitief negatief denken en depressieve klachten onder andere nagegaan in patiënten met een majeure depressie (MDD), voorheen depressieve individuen (RMD), risicopopulaties (bijvoorbeeld mensen die hoog scoren op ruminatie) en gezonde deelnemers (voor een overzicht, zie: Koster et al., 2017). Het is bijgevolg mogelijk dat de waargenomen heterogeniteit in behandeleffecten deels toe te schrijven is aan verschillen in onderzocht type populatie. Een ander gebied dat nog niet is onderzocht, is het vereiste aantal CCT-sessies om een vermindering van de kwetsbaarheid voor depressie waar te nemen. Toen Siegle en collega's (2007) de eersten waren die de aPASAT gebruikten als een CCT in combinatie met een aandachtstraining, gebruikten zij zes sessies, maar recentere studies doen typisch beroep op een hogere dosis van de aPASAT (meestal tien sessies). De effecten van een hogere dosis CCT op de kwetsbaarheid voor depressie zijn momenteel onduidelijk, omdat er nog geen dosis-responsstudies zijn uitgevoerd.

### *Huidige studie*

Bovenstaande bevindingen wijzen op het potentieel van aPASAT-training als interventie om depressieve kwetsbaarheid aan te pakken. Deze meta-analyse heeft tot doel de korte- en langetermijneffectiviteit van de aPASAT als CCT-procedure voor het aanpakken van depressieve symptomatologie en ruminatie te onderzoeken. Daarnaast zullen follow-up-analyses worden uitgevoerd om effecten van onder andere studiepopulatie en trainingsdoserings te bepalen. Een antwoord op deze vragen is noodzakelijk om een gericht klinisch gebruik van aPASAT-training te verantwoorden.

### METHODE

#### *Zoekstrategie*

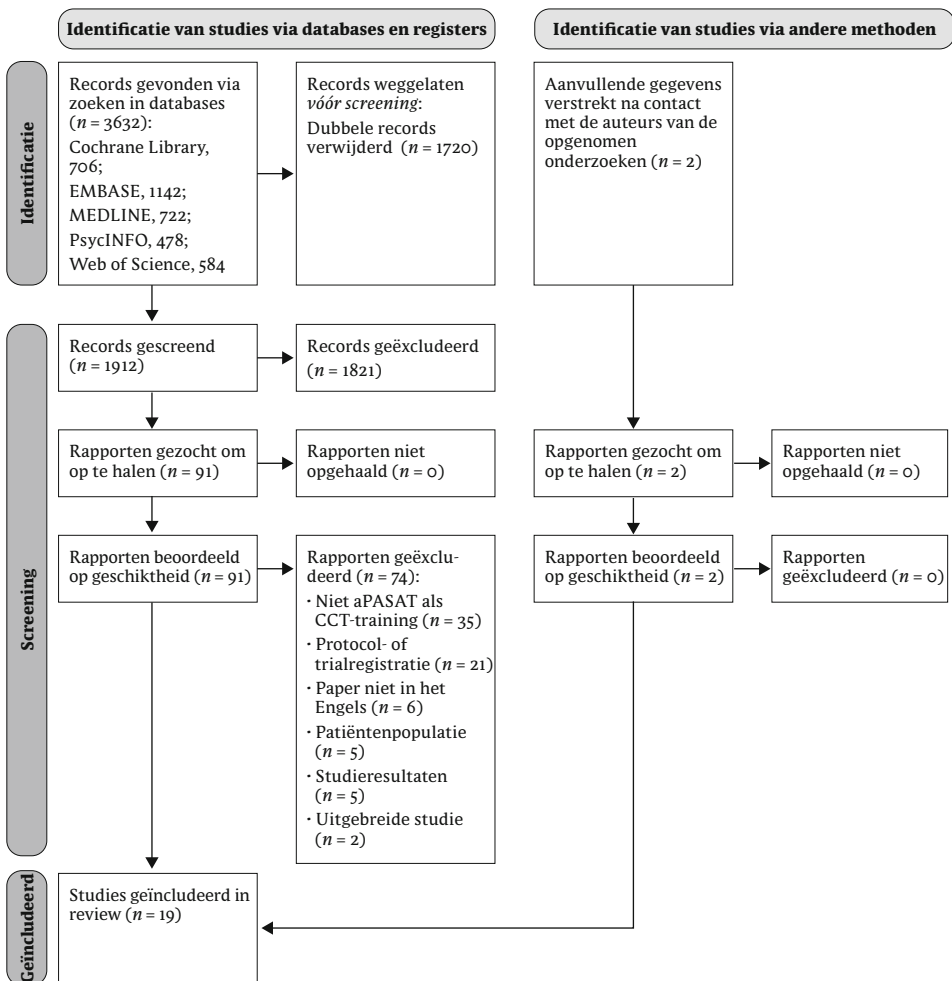
De zoektocht naar studies werd eerst uitgevoerd op 29 maart 2021 en herhaald op 28 september 2021. Er werd gezocht in de volgende elektronische databases: PsycINFO, MEDLINE, Embase, Cochrane Library en Web of Science, waarbij de volgende kernwoorden werden gebruikt om te zoeken op titel en abstract: ('cognitive control therapy' OF 'CCT' OF 'cognitive control training' OF 'cognitive control task' OF 'neurocognitive training' OF 'cognitive training' OF 'executive control training' OF 'working memory training' OF 'cognitive emotional training' OF 'cognitive remediation' OF 'neurobehavioral therapy' OF 'PASAT' OF 'Paced Auditory Serial Addition Task') EN (depress\* OF 'negative repetitive thinking' OF 'repetitive negative thinking' OF ruminat\* OF 'negative mood' OF 'brooding'). Om de impact van publicatiebias te verminderen, kwamen ook doctoraatsthesisen in aanmerking die aPASAT-training gebruikten en die de impact op depressieve klachten of ruminatie evalueerden.

#### *Inclusie- en exclusiecriteria*

Studies werden geïnccludeerd indien ze: (1) in het Engels waren geschreven, (2) een experimentele manipulatie van cognitieve controle hadden op basis van de aPASAT, en (3) de effecten van een aPASAT-training nagingen op ruminatie of depressieve klachten in een gezonde steekproef, risicogroep (bijvoorbeeld bij subklinische symptomen van depressie of verhoogde ruminatiescores), depressieve groep of voorheen depressieve (*remitted depressed*, RMD) populatie. Studies met een primaire focus op angststoornissen, verslavingsproblematiek, neurologische letsels of psychotische stoornissen werden geëxcludeerd.

### Studieselectie en data-extractie

Twee onafhankelijke onderzoekers screenen de studies eerst op titel en abstract, en nadien op volledige tekst. Bij verschillen werd een derde onderzoeker als arbiter ingeschakeld om een definitieve keuze te maken. De systematische zoektocht resulteerde in 3632 resultaten, waarvan 19 studies werden behouden. Het zoek- en selectieproces is weergegeven in een PRISMA-flowdiagram (zie figuur 1). Om de consistentie tussen de twee onafhankelijke beoordelaars te bepalen, werd een interbeoordelaarsbetrouwbaarheidsanalyse met behulp van Cohens kappa uitgevoerd.



FIGUUR 1 PRISMA flowdiagram

### Methodologische kwaliteitsbeoordeling

De Downs and Black checklist (Downs & Black, 1998), een methode om het risico op bias en om de interne en externe validiteit na te gaan, werd in deze studie door twee beoordelaars onafhankelijk van elkaar gebruikt om de kwaliteit van de geïncludeerde studies te beoordelen. Wanneer de twee reviewers verschilden van mening over het risico op bias, werd een derde reviewer als arbiter aangesteld, zoals aanbevolen door de Agency for Healthcare Research and Quality (Viswanathan et al., 2018).

De aanwezigheid van publicatiebias werd nagegaan aan de hand van een *funnel plot* voor de primaire uitkomst, namelijk depressieve symptomatologie, bij de postmeting. Verder werd Eggers test (Egger et al., 1997) uitgevoerd om de asymmetrie van de funnel plot te onderzoeken en werd Duval en Tweedie's *trim-and-fill*-procedure (Duval & Tweedie, 2000) gehanteerd om de schatting van de effectgrootte te controleren nadat rekening was gehouden met mogelijke publicatiebias.

### Statistische analyses

De analyses werden uitgevoerd in R (versie 4.0.5; R Core Team, 2021). Voor elk van de geselecteerde studies werden Hedges' *g* effectgrootten berekend voor depressieve symptomatologie en ruminatie. Hedges' *g* werd verkozen boven Cohens *d*, omdat deze een correctie biedt voor positieve bias in kleine steekproeven. De effectgrootten werden geïnterpreteerd volgens de aanbevelingen van Cohen (dat wil zeggen: we hanteerden 0,2 als 'klein', 0,5 als 'gemiddeld' en 0,8 als *cut off* voor 'grote' effecten; Cohen, 1988). Bij studies met meerdere condities werd de CCT-groep vergeleken met de meest vergelijkbare controlegroep.

Om gepoolde effectgrootten te berekenen werden *random effects*-modellen uitgevoerd. Een subgroepanalyse onderzocht de heterogeniteit als gevolg van de verschillende populatietypen, waarbij de volgende indeling werd gemaakt: gezond, risico (bijvoorbeeld met subklinische niveaus van depressieve symptomatologie), depressief (MDD) of voorheen depressief (remitted depressed, RMD). De indeling naar studiepopulatie werd gedaan door twee onafhankelijke beoordelaars en een derde beoordelaar fungeerde als arbiter. Een meta-regressieanalyse werd uitgevoerd om de associatie tussen het aantal aPASAT-sessies en de kwetsbaarheid voor depressie te onderzoeken.

Tot slot werden *forest plots* gemaakt voor zowel de primaire uitkomstmaat (depressieve symptomatologie) als de secundaire uitkomstmaat (ruminatie) voor de post- en follow-upmetingen. De follow-upmetingen varieerden sterk in tijd: van 2 weken tot 12 maanden ( $M = 3,6$  maanden,  $SD = 4,0$  maanden). De klinische significantie werd verduidelijkt door de gepool-

de effectgrootte van de primaire uitkomst bij de postmeting om te zetten in een *number needed to treat* (NNT), berekend met de Kraemer & Kupfer-methode (Kraemer & Kupfer, 2006).

## RESULTATEN

.....

### *Studiekenmerken*

.....

In deze meta-analyse werden 19 studies opgenomen met in totaal 1255 deelnemers. Er was een hoge interbeoordelaarsbetrouwbaarheid tussen de twee onafhankelijke onderzoekers voor de bepaling van de inclusie van studies ( $\kappa = 0,86$ ). In de geïnccludeerde studies was ongeveer 72% van de deelnemers vrouw, met een gemiddelde leeftijd van 37,9 jaar ( $SD = 12,7$ ). De steekproefgrootten die werden gebruikt voor de berekening van effectgrootten varieerden van  $n = 9$  (Segrave et al., 2014) tot  $n = 213$  (Hoorelbeke et al., 2022b), met een mediaan van  $n = 49$  ( $IQR = 50$ ). Van de 19 geïnccludeerde studies richtten 7 (36,8%) zich op een majeure depressieve populatie, 4 (21,1%) op voorheen depressieve individuen, 5 op een overige risicopopulatie (26,3%) en 3 onderzochten effecten van aPASAT-training in een gezonde steekproef (15,8%). Tabel 1 bevat de kenmerken van de geïnccludeerde studies. 'Inclusief N' in de tabel verwijst naar het aantal deelnemers voor wie gegevens werden gerapporteerd die beschikbaar waren voor de berekening van effectgrootten, waarbij alleen vergelijkbare groepen in aanmerking werden genomen.

In totaal werden 57 effectgrootten berekend: 34 voor de primaire uitkomstmaat depressieve symptomatologie (18 voor nameting, 16 voor follow-up), 23 voor de secundaire uitkomstmaat ruminatie (14 voor nameting, 9 voor follow-up). Wanneer studies meerdere groepen hadden, werd de meest vergelijkbare of meest conservatieve controlegroep gebruikt. Zo hadden Lass en collega's (2021) drie groepen: CCT, actieve controle en wachtlijstcontrole: voor de berekening van de effectgrootten werden de CCT en actieve controlegroepen gebruikt. Tabel 2 geeft een overzicht van de gerapporteerde gegevens en de berekende Hedges' *g* effectgrootten.

### *Risico op bias*

.....

De beoordeling van het risico op bias tussen de twee onafhankelijke beoordelaars kwam sterk overeen. Van de 19 onderzochte studies resulteerden er slechts drie in een verschillende beoordelingscategorie, die elk slechts één punt verschilden op de Downs and Black checklist. Een derde beoordelaar nam de uiteindelijke beslissing over deze drie punten. Het risico op bias in de onderzochte studies was in het algemeen beperkt. Rapportage, blinding en randomisatie waren meestal voldoende. Slechts 12 van de 19 studies

TABEL 1 Karakteristieken van de geïncludeerde studies

| Studies                    | Land       | Doel-<br>groep | N (inclu-<br>sief N) | Percenta-<br>ge vrouw | Aantal<br>sessies | Uitkomst-<br>maten | Follow-up<br>(maanden) | Controle-<br>groep    | Gemiddelde<br>leeftijd (SD) | RoBa-<br>sessment |
|----------------------------|------------|----------------|----------------------|-----------------------|-------------------|--------------------|------------------------|-----------------------|-----------------------------|-------------------|
| Brunoni et al.,<br>2014    | Brazilië   | MDD            | 40 (17)              | 30,8                  | 10                | BDI-II, HDRS       | 0,5                    | Within-group          | 41,5 (10,6)                 | Excellent         |
| Calkins et al.,<br>2015    | USA        | Risico         | 56 (48)              | 54,2                  | 3                 | BDI-II             | /                      | Niet-actieve controle | 35,7 (14,7)                 | Goed              |
| Ferrari et al.,<br>2021    | Duits-land | MDD            | 115 (115)            | 62,6                  | 10                | BDI-II, RSQ        | 12                     | Actieve controle      | 51,3 (7,4)                  | Goed              |
| Hoorelbeke et al.,<br>2015 | België     | Risico         | 53 (47)              | 92,5                  | 10                | BDI-II, RRS        | 1                      | Actieve controle      | 20,6 (2,1)                  | Goed              |
| Hoorelbeke et al.,<br>2016 | België     | Gezond         | 61 (61)              | 85,3                  | 10                | VAS                | /                      | Actieve controle      | 21,4 (2,5)                  | Goed              |
| Hoorelbeke & Koster, 2017  | België     | RMD            | 68 (68)              | 66,2                  | 10                | BDI-II, RRS        | 3                      | Actieve controle      | 47,0 (11,5)                 | Excellent         |
| Hoorelbeke et al., 2021    | België     | RMD            | 92 (82)              | 64,1                  | 10                | BDI-II, RRS        | 12                     | Actieve controle      | 45,8 (12,8)                 | Goed              |
| Hoorelbeke et al., 2022a   | België     | RMD            | 92 (84)              | 64,1                  | 10                | VAS                | /                      | Actieve controle      | 45,8 (12,8)                 | Goed              |
| Hoorelbeke et al., 2022b   | België     | Risico         | 213 (213)            | 73,0                  | 10                | DASS, PTQ          | 1                      | Within-group          | 43,1 (13,3)                 | Goed              |
| Lass et al., 2021          | USA        | Risico         | 72 (47)              | 75,0                  | 6                 | BDI, RRS           | /                      | Actieve controle      | 21,8 (5,7)                  | Goed              |

| Studies                    | Land      | Doelgroep | N (inclusief N) | Percentage vrouw | Aantal sessies | Uitkomstmaten      | Follow-up (maanden) | Controlegroep         | Gemiddelde leeftijd (SD) | RoB assessment |
|----------------------------|-----------|-----------|-----------------|------------------|----------------|--------------------|---------------------|-----------------------|--------------------------|----------------|
| Moshier & Otto, 2017       | USA       | MDD       | 34 (34)         | 52,0             | 4              | BDI-II, MADRS, RRS | 1                   | Niet-actieve controle | 35,6 (14,6)              | Goed           |
| Segrave et al., 2014       | Australië | MDD       | 27 (18)         | 37,0             | 5              | BDI-II             | 0,75                | Niet-actieve controle | 40,4 (14,5)              | Goed           |
| Siegle et al., 2014        | USA       | MDD       | 51 (43)         | 69,8             | 6              | BDI-II, RSQ        | /                   | Niet-actieve controle | 39,5 (10,4)              | Volledende     |
| Sommer & Plewnia, 2021     | Duitsland | MDD       | 51 (17)         | 58,8             | 12             | BDI, MADRS         | 3                   | Within-group          | 34,6 (13,7)              | Volledende     |
| Van den Bergh et al., 2020 | België    | Risico    | 102 (89)        | 79,7             | 10             | DASS, PTQ          | 2                   | Actieve controle      | 20,78 (2,3)              | Goed           |
| Vanderhasselt et al., 2015 | Brazilië  | MDD       | 37 (14)         | 79,0             | 10             | BDI-II, RRS        | /                   | Within-group          | 41,0 (11,5)              | Goed           |
| Vanderhasselt et al., 2020 | België    | Gezond    | 41 (41)         | 53,9             | 10             | RRS                | 1,5                 | Actieve controle      | 70,3 (3,8)               | Volledende     |
| Vervaeke et al., 2020      | België    | Gezond    | 147 (147)       | 74,1             | 10             | DASS, RRS          | /                   | Actieve controle      | 21,4 (4,5)               | Goed           |
| Vervaeke et al., 2021      | België    | RMD       | 68 (68)         | 64,7             | 10             | BDI-II, DASS, RRS  | 3, 6                | Actieve controle      | 39,9 (17,6)              | Goed           |

Noot. Geïnccludeerde N refereert naar de steekproef die werd gebruikt om de effectgrootten te berekenen in deze meta-analyse. BDI = Beck Depression Inventory; BDI-II = Beck Depression Inventory - 2nd edition; DASS = Depression Anxiety Stress Scales - depressie subschaal; HDRS = Hamilton Depression Rating Scale; MADRS = Montgomery Asberg Depression Rating Scale; MDD = Major Depressive Disorder; PTQ = Perseverative Thinking Questionnaire; RMD = Remitted depressed; RoB = Risk of Bias; RRS = Ruminative Response Scale; RSQ = Response Style Questionnaire; VAS = Visual Analogue Scale.

TABEL 2 Data en berekende effectgrootten

| Studie                  | Uitkomst-maat         | Meting | N CCT | Gemiddel-de CCT | SD CCT | N con-trole | Gemiddelde controle | SD con-trole | Hedges' g (SE) | Tijdstip  |
|-------------------------|-----------------------|--------|-------|-----------------|--------|-------------|---------------------|--------------|----------------|-----------|
| Brunoni et al., 2014    | Depressieve symptomen | BDI-II | 15    | 23,00           | 12,00  | 15          | 34,00               | 8,00         | 0,98 (0,29)    | Post      |
| Brunoni et al., 2014    | Depressieve symptomen | HDRS   | 17    | 20,00           | 9,00   | 17          | 27,00               | 5,70         | 1,11 (0,35)    | Post      |
| Brunoni et al., 2014    | Depressieve symptomen | BDI-II | 15    | 24,00           | 11,40  | 15          | 34,00               | 8,00         | 1,00 (0,31)    | Follow-up |
| Brunoni et al., 2014    | Depressieve symptomen | HDRS   | 17    | 20,00           | 8,70   | 17          | 27,00               | 5,70         | 1,14 (0,35)    | Follow-up |
| Calkins et al., 2015    | Depressieve symptomen | BDI-II | 24    |                 |        | 24          |                     |              | 0,72 (0,30)    | Post      |
| Ferrari et al., 2021    | Depressieve symptomen | BDI-II | 56    | 1728            | 9,50   | 59          | 19,40               | 11,58        | 0,20 (0,19)    | Post      |
| Ferrari et al., 2021    | Ruminatie             | RSQ    | 56    | 50,23           | 10,24  | 59          | 51,25               | 13,48        | 0,08 (0,19)    | Post      |
| Hoorelbeke et al., 2015 | Depressieve symptomen | BDI-II | 25    | 12,12           | 11,33  | 22          | 9,41                | 7,84         | -0,27 (0,29)   | Follow-up |
| Hoorelbeke et al., 2015 | Ruminatie             | RRS    | 25    | 48,88           | 11,27  | 22          | 50,95               | 9,88         | 0,19 (0,29)    | Post      |
| Hoorelbeke et al., 2015 | Ruminatie             | RRS    | 20    | 50,05           | 10,16  | 17          | 51,24               | 8,74         | 0,12 (0,33)    | Follow-up |
| Hoorelbeke et al., 2016 | Depressieve symptomen | VAS    | 29    | 10,44           | 10,09  | 32          | 8,12                | 6,65         | -0,27 (0,26)   | Post      |

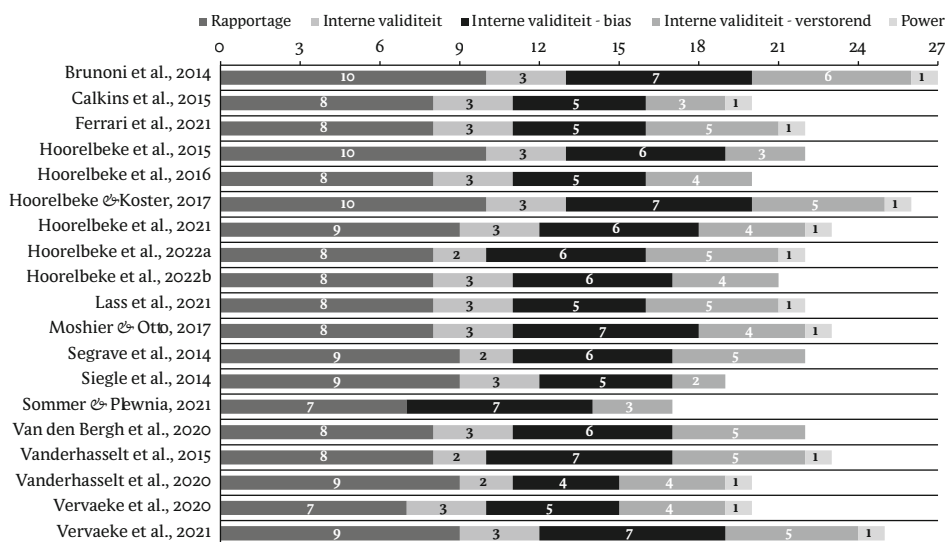
| Studie                    | Uitkomst-maat         | Meting | N CCT | Gemiddel-de CCT | SD CCT | N con-trole | Gemiddelde controle | SD con-trole | Hedges' g (SE) | Tijdstip  |
|---------------------------|-----------------------|--------|-------|-----------------|--------|-------------|---------------------|--------------|----------------|-----------|
| Hoorelbeke et al., 2016   | Ruminatie             | VAS    | 29    | 28,05           | 14,50  | 32          | 27,07               | 12,76        | -0,07 (0,26)   | Post      |
| Hoorelbeke & Koster, 2017 | Depressieve symptomen | BDI-II | 34    | 5,38            | 7,10   | 34          | 9,29                | 9,37         | 0,47 (0,25)    | Post      |
| Hoorelbeke & Koster, 2017 | Depressieve symptomen | BDI-II | 34    | 4,50            | 5,10   | 34          | 9,29                | 7,28         | 0,75 (0,25)    | Follow-up |
| Hoorelbeke & Koster, 2017 | Ruminatie             | RRS    | 34    | 34,71           | 10,01  | 34          | 44,29               | 12,58        | 0,83 (0,25)    | Post      |
| Hoorelbeke & Koster, 2017 | Ruminatie             | RRS    | 34    | 30,06           | 6,72   | 34          | 40,35               | 14,06        | 0,92 (0,26)    | Follow-up |
| Hoorelbeke et al., 2021   | Depressieve symptomen | BDI-II | 43    | 9,51            | 8,75   | 39          | 11,38               | 10,16        | 0,20 (0,22)    | Follow-up |
| Hoorelbeke et al., 2021   | Ruminatie             | RRS    | 43    | 45,95           | 14,44  | 39          | 48,90               | 13,30        | 0,21 (0,22)    | Follow-up |
| Hoorelbeke et al., 2022a  | Depressieve symptomen | VAS    | 46    | 10,36           | 10,51  | 38          | 15,06               | 16,82        | 0,34 (0,22)    | Post      |
| Hoorelbeke et al., 2022a  | Ruminatie             | VAS    | 46    | 22,15           | 17,34  | 38          | 29,47               | 21,47        | 0,38 (0,22)    | Post      |
| Hoorelbeke et al., 2022b  | Depressieve symptomen | DASS   | 213   | 9,29            | 8,80   | 213         | 12,60               | 9,32         | 0,29 (0,06)    | Post      |
| Hoorelbeke et al., 2022b  | Ruminatie             | PTQ    | 213   | 34,98           | 9,08   | 213         | 39,43               | 8,22         | 0,42 (0,06)    | Post      |
| Hoorelbeke et al., 2022b  | Depressieve symptomen | DASS   | 213   | 9,27            | 9,10   | 213         | 12,60               | 9,32         | 0,33 (0,06)    | Follow-up |

| Studie                      | Uitkomst-<br>maat        | Meting | N<br>CCT | Gemiddel-<br>de CCT | SD<br>CCT | N con-<br>trole | Gemiddelde<br>controle | SD con-<br>trole | Hedges' g (SE) | Tijdstip  |
|-----------------------------|--------------------------|--------|----------|---------------------|-----------|-----------------|------------------------|------------------|----------------|-----------|
| Hoorelbeke et al.,<br>2022b | Ruminatie                | PTQ    | 213      | 31,77               | 9,15      | 213             | 39,43                  | 8,22             | 0,79 (0,07)    | Follow-up |
| Moshier & Otto, 2017        | Depressieve<br>symptomen | BDI-II | 21       | 23,10               | 10,80     | 13              | 18,60                  | 13,30            | -0,37 (0,36)   | Post      |
| Moshier & Otto, 2017        | Depressieve<br>symptomen | BDI-II | 21       | 22,00               | 11,80     | 13              | 18,20                  | 14,60            | -0,29 (0,35)   | Follow-up |
| Moshier & Otto, 2017        | Depressieve<br>symptomen | MADRS  | 21       | 19,76               | 11,10     | 13              | 17,00                  | 9,90             | -0,25 (0,35)   | Post      |
| Moshier & Otto, 2017        | Depressieve<br>symptomen | MADRS  | 21       | 20,00               | 9,80      | 13              | 1750                   | 11,70            | -0,23 (0,35)   | Follow-up |
| Moshier & Otto, 2017        | Ruminatie                | RRS    | 21       | 54,60               | 12,20     | 13              | 48,90                  | 15,20            | -0,42 (0,36)   | Post      |
| Moshier & Otto, 2017        | Ruminatie                | RRS    | 21       | 52,00               | 11,20     | 13              | 44,70                  | 14,30            | -0,57 (0,21)   | Follow-up |
| Segrave et al., 2014        | Depressieve<br>symptomen | BDI-II | 9        | 23,44               | 12,37     | 9               | 25,33                  | 10,05            | 0,16 (0,47)    | Post      |
| Segrave et al., 2014        | Depressieve<br>symptomen | BDI-II | 9        | 15,33               | 10,37     | 9               | 27,78                  | 8,61             | 1,24 (0,52)    | Follow-up |
| Siegle et al., 2014         | Depressieve<br>symptomen | BDI-II | 23       | 24,13               | 12,14     | 20              | 30,55                  | 11,31            | 0,54 (0,31)    | Post      |
| Siegle et al., 2014         | Ruminatie                | RSQ    | 23       | 49,70               | 9,52      | 20              | 59,80                  | 9,87             | 1,02 (0,33)    | Post      |
| Sommer & Plewnia,<br>2021   | Depressieve<br>symptomen | BDI    | 17       | 1735                | 754       | 17              | 2559                   | 6,68             | 1,08 (0,29)    | Post      |

| Studie                     | Uitkomst-maat         | Meting | N CCT | Gemiddel-de CCT | SD CCT | N con-trole | Gemiddelde controle | SD con-trole | Hedges' g (SE) | Tijdstip  |
|----------------------------|-----------------------|--------|-------|-----------------|--------|-------------|---------------------|--------------|----------------|-----------|
| Sommer & Plewnia, 2021     | Depressieve symptomen | MADRS  | 17    | 18,35           | 8,85   | 17          | 29,0                | 7,45         | 1,41 (0,36)    | Post      |
| Sommer & Plewnia, 2021     | Depressieve symptomen | BDI    | 17    | 14,73           | 8,41   | 17          | 25,59               | 6,68         | 1,26 (0,30)    | Follow-up |
| Sommer & Plewnia, 2021     | Depressieve symptomen | MADRS  | 17    | 15,87           | 8,94   | 17          | 29,0                | 7,45         | 2,07 (0,47)    | Follow-up |
| Van den Bergh et al., 2020 | Depressieve symptomen | DASS   | 47    | 6,10            | 5,50   | 42          | 5,80                | 5,10         | -0,06 (0,21)   | Post      |
| Van den Bergh et al., 2020 | Depressieve symptomen | DASS   | 47    | 5,90            | 5,00   | 42          | 5,80                | 4,90         | -0,02 (0,21)   | Follow-up |
| Van den Bergh et al., 2020 | Ruminatie             | PTQ    | 47    | 35,90           | 10,10  | 42          | 37,10               | 9,80         | 0,12 (0,21)    | Post      |
| Van den Bergh et al., 2020 | Ruminatie             | PTQ    | 47    | 33,90           | 9,90   | 42          | 35,30               | 9,80         | 0,14 (0,21)    | Follow-up |
| Vanderhasselt et al., 2015 | Ruminatie             | RRS    | 14    | 58,57           | 13,25  | 14          | 66,00               | 8,44         | 0,67 (0,30)    | Post      |
| Vanderhasselt et al., 2020 | Ruminatie             | RRS    | 21    | 35,29           | 9,54   | 20          | 34,89               | 11,17        | -0,04 (0,31)   | Post      |
| Vanderhasselt et al., 2020 | Ruminatie             | RRS    | 21    | 33,52           | 10,29  | 20          | 35,82               | 14,75        | 0,18 (0,31)    | Follow-up |
| Vervaeke et al., 2020      | Depressieve symptomen | DASS   | 74    | 18,74           | 5,78   | 73          | 19,89               | 6,56         | 0,19 (0,17)    | Post      |

| Studie                | Uitkomst-<br>maat        | Meting | N<br>CCT | Gemiddel-<br>de CCT | SD<br>CCT | N con-<br>trole | Gemiddelde<br>controle | SD con-<br>trole | Hedges' g (SE) | Tijdstip  |
|-----------------------|--------------------------|--------|----------|---------------------|-----------|-----------------|------------------------|------------------|----------------|-----------|
| Vervaeke et al., 2020 | Ruminatie                | RRS    | 74       | 41,58               | 11,93     | 73              | 42,11                  | 10,76            | 0,05 (0,17)    | Post      |
| Vervaeke et al., 2021 | Depressieve<br>symptomen | BDI-II | 34       | 9,00                | 5,70      | 34              | 9,00                   | 6,50             | 0,00 (0,24)    | Follow-up |
| Vervaeke et al., 2021 | Depressieve<br>symptomen | DASS   | 34       | 6,50                | 7,90      | 34              | 5,90                   | 7,90             | -0,08 (0,24)   | Post      |
| Vervaeke et al., 2021 | Depressieve<br>symptomen | DASS   | 34       | 4,50                | 4,80      | 34              | 7,70                   | 8,50             | 0,46 (0,25)    | Follow-up |
| Vervaeke et al., 2021 | Depressieve<br>symptomen | DASS   | 34       | 4,80                | 5,40      | 34              | 7,60                   | 7,60             | 0,42 (0,25)    | Follow-up |
| Vervaeke et al., 2021 | Ruminatie                | RRS    | 34       | 50,50               | 10,30     | 34              | 51,90                  | 12,22            | 0,12 (0,24)    | Post      |
| Vervaeke et al., 2021 | Ruminatie                | RRS    | 34       | 46,10               | 12,00     | 34              | 45,90                  | 12,50            | -0,02 (0,24)   | Follow-up |
| Vervaeke et al., 2021 | Ruminatie                | RRS    | 34       | 48,10               | 12,80     | 34              | 47,50                  | 11,40            | -0,05 (0,24)   | Follow-up |
| Lass et al., 2021     | Depressieve<br>symptomen | BDI    | 24       | 17,08               | 10,22     | 23              | 18,26                  | 9,54             | 0,12 (0,29)    | Post      |
| Lass et al., 2021     | Ruminatie                | RRS    | 24       | 58,08               | 12,69     | 23              | 56,35                  | 10,70            | 0,14 (0,29)    | Post      |

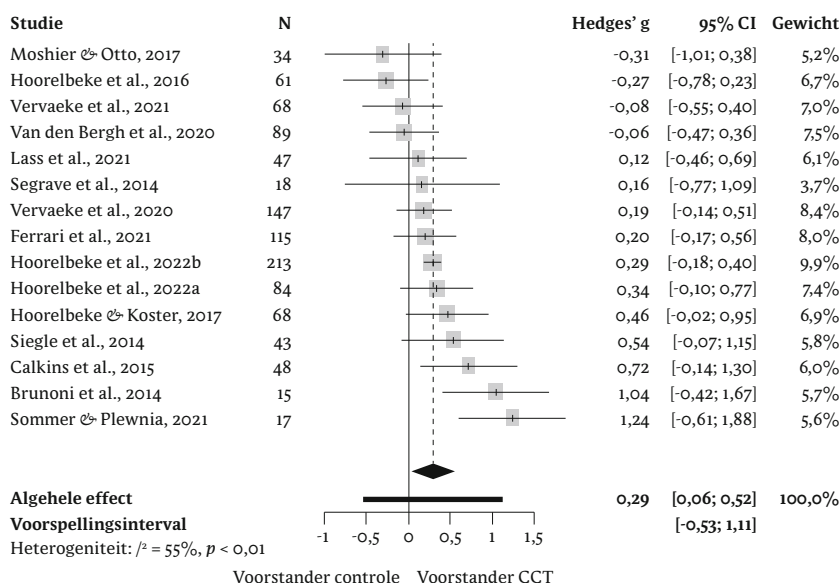
*Noot.* Voor Calkins et al., 2015 werden geen gemiddelden en standaardafwijkingen gerapporteerd. De effectgrootte werd berekend van de gerapporteerde *t*-test. BDI = Beck Depression Inventory; BDI-II = Beck Depression Inventory - 2nd edition; CCT = Cognitive Control Training; DASS = Depression Anxiety Stress Scales – depression subscale; HDRS = Hamilton Depression Rating Scale; MADRS = Montgomery Asberg Depression Rating Scale; PTQ = Perseverative Thinking Questionnaire; RRS = Ruminative Response Scale; RSQ = Response Style Questionnaire; VAS = Visual Analogue Scale

FIGUUR 2 *Downs and Black checklist – Samenvatting*

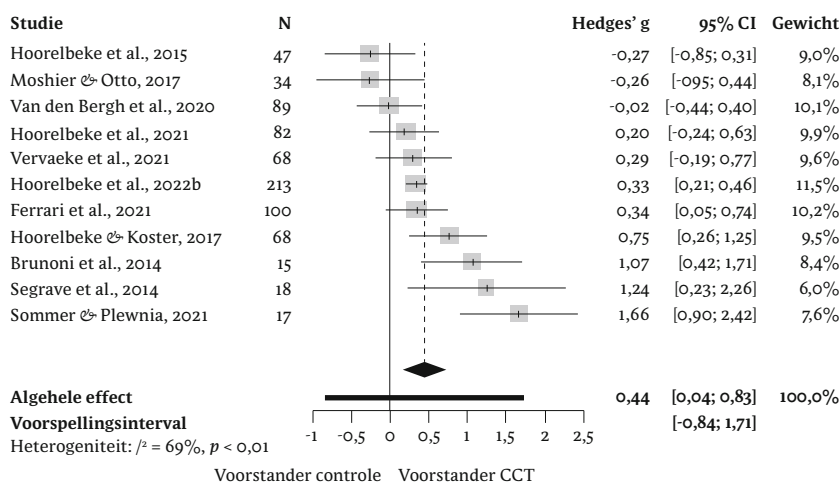
voerden een statistische poweranalyse uit, waarbij zij in hun powerberekeningen meestal uitgingen van een matige tot hoge effectgrootte. Mogelijke bijwerkingen werden in de aPASAT-literatuur niet uitgebreid behandeld. Omdat voor een CCT-procedure zeer weinig bijwerkingen worden verwacht, is hier mogelijk in eerder onderzoek weinig aandacht aan besteed. Concluderend kan gesteld worden dat geen enkele studie in de categorie 'slecht' scoorde, drie studies scoorden 'voldoende' (15,8%), 14 studies scoorden 'goed' (73,7%) en twee studies scoorden 'excellent' (10,5%). Voor een samenvatting van de beoordeling van het risico op bias, zie figuur 2.

### *Effecten van aPASAT-training op depressieve symptomatologie*

Vijftien studies onderzochten de effecten van aPASAT-training op depressieve symptomatologie onmiddellijk na de training, wat resulteerde in een gepoolde effectgrootte  $g = 0,29$  met een 95% betrouwbaarheidsinterval (CI) [0,06, 0,52]. De gepoolde effectgrootte was significant ( $p = 0,018$ ) en de heterogeniteit (dat wil zeggen: de hoeveelheid variatie tussen de studies die niet toe te schrijven is aan kans) was gemiddeld ( $\tau^2 = 0,13$ ,  $I^2 = 54,7\%$ , 95% CI [18,8%, 74,7%]). Elf studies gingen tevens langetermijneffecten na van aPASAT-training op depressieve klachten, wat resulteerde in een gepoolde effectgrootte  $g = 0,44$ , 95% CI [0,04, 0,83]. De gepoolde effectgrootte voor de follow-up was eveneens significant ( $p = 0,033$ ) en de heterogeniteit was aanzienlijk ( $\tau^2 = 0,28$ ,  $I^2 = 69,4\%$ , 95% CI [42,9%, 83,6%]). Figuren 3 en 4 tonen de forest plots voor depressieve symptomen voor respectievelijk de



FIGUUR 3 Forest plot voor depressieve symptomatologie – posttraining (CCT = cognitieve-controltraining)

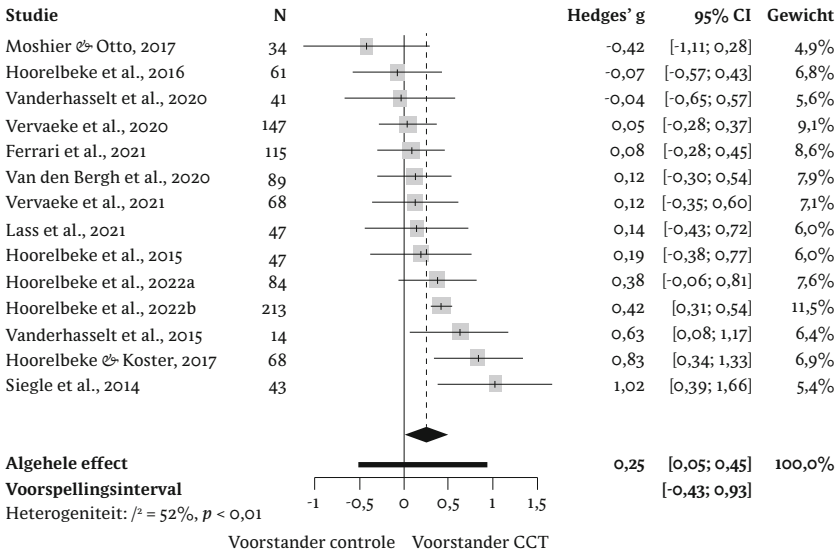


FIGUUR 4 Forest plot voor depressieve symptomatologie – follow-up (CCT = cognitieve-controltraining)

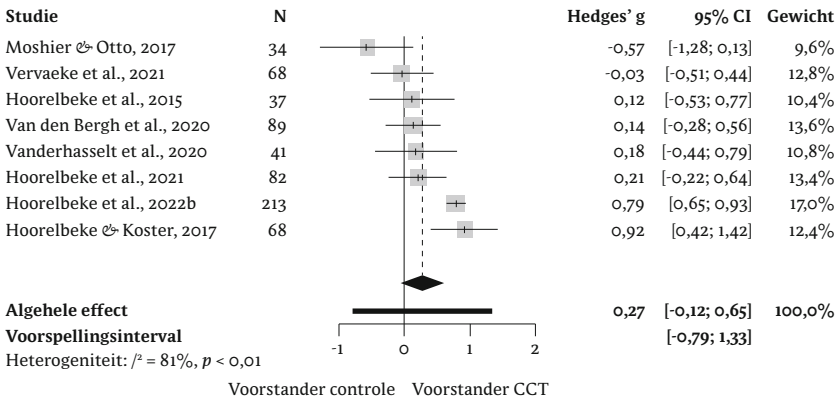
postmeting en de follow-upmeting. Deze wijzen op kleine, maar significante effectgrootten van aPASAT-training op depressieve symptomatologie.

Effecten van aPASAT-training op ruminatie

De secundaire uitkomstmaat, ruminatie, liet vergelijkbare resultaten zien. De gepoolde effectgrootte voor de 14 studies bij de postmeting was significant ( $g = 0,25$ , 95% CI = [0,05, 0,45],  $p = 0,018$ ) en de heterogeniteit was matig ( $\tau^2 = 0,09$ ,  $I^2 = 52,4\%$ , 95% CI [12,5%, 74,1%]). Bij follow-up was Hedges'  $g$  niet langer significant ( $g = 0,27$ , 95% CI = [-0,12, 0,65],  $p = 0,142$ ) voor de acht geïnccludeerde studies. De heterogeniteit was aanzienlijk ( $\tau^2 = 0,16$ ,  $I^2 = 81,4\%$ , 95% CI [64,4%, 90,3%]). Figuren 5 en 6 geven de verkregen forest plots weer voor ruminatie bij respectievelijk de postmeting en follow-up.



FIGUUR 5 Forest plot voor ruminatie – posttraining (CCT = cognitieve-controletraining)



FIGUUR 6 Forest plot voor ruminatie – follow-up (CCT = cognitieve-controletraining)

### *Meta-regressie op het aantal trainingssessies*

We voerden een meta-regressie uit om na te gaan of het aantal trainingssessies dat gebruikt werd binnen de verschillende studies een modererend effect had op de uitkomstmaten. Het aantal gebruikte aPASAT-sessies had geen significant effect op de gepoolde effectgrootte, wat betreft zowel depressieve symptomatologie als ruminatie (depressieve symptomen op de postmeting:  $\gamma = 0,02$ ,  $SE = 0,04$ ,  $p = 0,680$ ; depressieve symptomen bij follow-up:  $\gamma = 0,07$ ,  $SE = 0,08$ ,  $p = 0,471$ ; ruminatie bij postmeting:  $\gamma = 0,3$ ,  $SE = 0,05$ ,  $p = 0,621$ ; ruminatie bij follow-up:  $\gamma = 0,16$ ,  $SE = 0,08$ ,  $p = 0,097$ ). Dit betekent dat er geen effect werd gevonden van een verschillend aantal trainingssessies op de effectiviteit van aPASAT-training.

### *Subgroepanalyse op populatietype*

We voerden tevens een subgroepanalyse uit om te onderzoeken of de effecten van de aPASAT-training verschillend zijn voor verschillende groepen. Voor depressieve symptomen gemeten onmiddellijk na de training werd geen significant effect van de populatie waargenomen. Daarentegen toonde een random-effectmodel een significant verschil tussen populatietypen voor ruminatie, wat wijst op een hoger gemiddeld effect van aPASAT-training in MDD, RMD en risicopopulaties direct na de training in vergelijking met een gezonde populatie (tabel 3). Voor zowel depressieve symptomen als rumineren werd bij follow-up geen significant effect van de populatie waargenomen ( $ps > 0,11$ ).

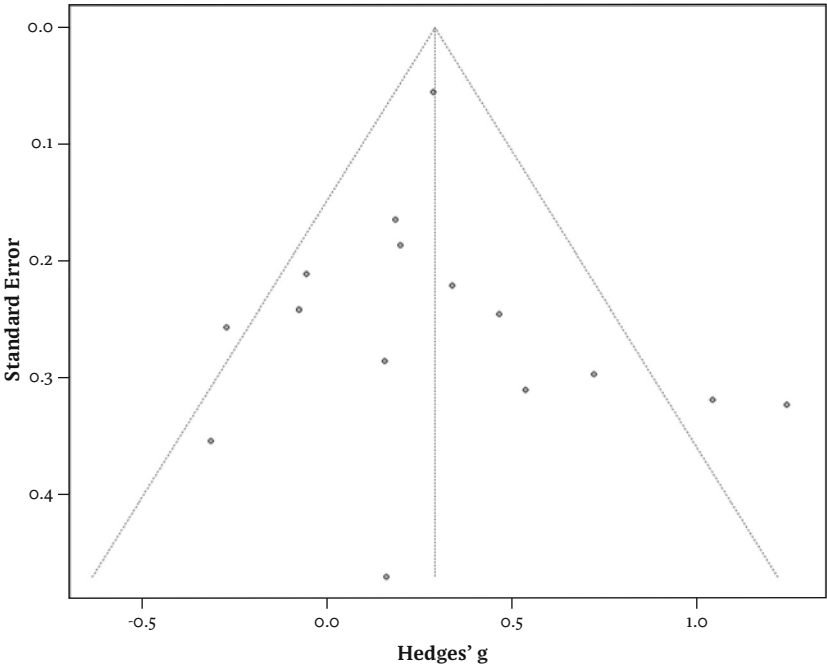
### *Publicatiebias*

Aanwijzingen voor publicatiebias werden onderzocht met behulp van visuele inspectie van de funnel plot, Eggers test en de Duval en Tweedie's trim-and-fill-methode, uitgevoerd op de primaire uitkomstmaat bij de postmeting. De funnel plot vertoonde geen asymmetrie of indicatie van publicatiebias (figuur 7). Eggers test was niet significant ( $p = 0,850$ ). Duval en Tweedie's trim-and-fill-procedure verwijderde geen studies en voegde er ook geen toe, waardoor de gepoolde effectgrootte niet veranderde. Er werden geen aanwijzingen voor publicatiebias gevonden, wat wijst op een succesvolle en volledige zoekstrategie.

TABEL 3 Subgroepanalyses op populatietype

| Depressieve symptomatologie – post |          |                  |               |          |                       |          |
|------------------------------------|----------|------------------|---------------|----------|-----------------------|----------|
| Populatie                          | <i>k</i> | Hedges' <i>g</i> | 95% CI        | $\tau^2$ | <i>I</i> <sup>2</sup> | <i>p</i> |
| MDD                                | 6        | 0,49             | [-0,12, 1,10] | 0,24     | 69,3%                 | 0,524    |
| RMD                                | 3        | 0,25             | [-0,45, 0,94] | 0,04     | 26,4%                 |          |
| Risico                             | 4        | 0,25             | [-0,22, 0,72] | 0,06     | 40,1%                 |          |
| Gezond                             | 2        | 0,00             | [-2,84, 2,84] | 0,05     | 54,9%                 |          |
| Ruminatie – post                   |          |                  |               |          |                       |          |
| Populatie                          | <i>k</i> | Hedges' <i>g</i> | 95% CI        | $\tau^2$ | <i>I</i> <sup>2</sup> | <i>p</i> |
| MDD                                | 4        | 0,33             | [-0,64, 1,30] | 0,29     | 74,3%                 | < 0,001  |
| RMD                                | 3        | 0,44             | [-0,44, 1,32] | 0,08     | 52,4%                 |          |
| Risico                             | 4        | 0,33             | [0,09, 0,57]  | < 0,01   | 2,1%                  |          |
| Gezond                             | 3        | 0,00             | [-0,16, 0,16] | < 0,01   | 0,0%                  |          |

Noot. MDD = Major Depressive Disorder; RMD = RMD = Remitted depressed.



FIGUUR 7 Funnel plot publicatiebias

*Klinische significantie*

Om de gepoolde effectgrootte voor klinische relevantie gemakkelijker te interpreteren, werd het gestandaardiseerde gemiddelde verschil van depressieve symptomen na de training omgezet in een number needed to treat (NNT) met behulp van de Kraemer & Kupfer-methode. De gepoolde effectgrootte van de primaire uitkomstmaat, depressieve symptomatologie, op de postmeting  $g = 0,29$  komt overeen met een NNT van 6,15. Dit betekent dat ongeveer zes patiënten behandeld dienen te worden met aPASAT-training om één patiënt te laten verbeteren.

## DISCUSSIE

Depressie gaat typisch gepaard met cognitieve klachten, zoals moeite met aandacht, geheugen en concentratie. Deze klachten blijven vaak aanwezig na het opklaren van een depressieve episode en houden een risico in op terugval. In deze meta-analyse onderzochten we de effectiviteit van aPASAT-training op depressieve kwetsbaarheid. Eerst werd de algemene effectiviteit van deze specifieke CCT nagegaan op depressieve symptomen als primaire uitkomst en vervolgens op ruminatie als secundaire uitkomstmaat. We maakten daarbij steeds onderscheid tussen enerzijds onmiddellijke effecten van aPASAT-training (posttraining) en anderzijds de langetermijneffecten ervan (follow-up). De resulterende effectgrootte onmiddellijk na de training wijst op een klein maar significant effect op zowel depressieve klachten als ruminatie. Voor depressieve klachten bleef de effectgrootte significant op follow-up. Hoewel in dezelfde grootteorde, werd geen significant effect teruggevonden op ruminatie bij follow-up.

Een mogelijke verklaring hiervoor is dat minder studies langetermijneffecten van aPASAT-training in kaart brachten. Bovendien maakte een aantal van deze studies gebruik van relatief beperkte steekproefgrootten, wat in combinatie met uitval van participanten tijdens een follow-upperiode evaluatie van langetermijneffecten bemoeilijkt. Dit resulteerde mogelijk in beperkte power voor deze laatste analyses. Tot slot observeerden we sterke heterogeniteit inzake de duur van opvolging van participanten. Zo vond in sommige studies de follow-upmeting enkele weken na beëindiging van de training plaats, terwijl andere studies effecten tot één jaar na de training in kaart brachten.

De geobserveerde effectgrootten voor aPASAT-training binnen deze meta-analyse stemmen overeen met eerdere meta-analyses die effecten van cognitieve training voor depressie in kaart brachten (bijvoorbeeld: Launderson et al., 2021; Motter et al., 2016). Deze rapporteerden eveneens kleine, significante effecten van cognitieve training, zij het niet specifiek voor de

aPASAT. Vanuit klinisch oogpunt is de aanwezigheid van dergelijke effecten bijzonder interessant, gegeven de relatief lage intensiteit van de training en de sterke heterogeniteit in gebruikte follow-upperiodes. Hoewel de effecten klein zijn, lijken ze dus duurzaam op langere termijn. Eerder onderzoek wees in deze context zelfs op een toename in klinische verbetering na CCT voor sommige uitkomstmaten (Hoorelbeke & Koster, 2017).

In tegenstelling tot de verwachtingen suggereren de resultaten geen effect van het *aantal trainingssessies*. Dit kan mogelijk verklaard worden door de relatief beperkte variabiliteit inzake het aantal gebruikte trainingssessies over de verschillende studies heen. Er werden slechts drie studies gevonden die vijf of minder sessies gebruikten, terwijl de meest recente studies 10 of 12 sessies hanteerden. Verder hebben eerdere studies vaak gebruikgemaakt van een multifactoriale training, waarin de aPASAT werd gecombineerd met een aandachtstraining (bijvoorbeeld: Calkins et al., 2015; Siegle et al., 2007, 2014). Hierdoor is het effect van deze gecombineerde training niet louter toe te schrijven aan aPASAT-trainingsspecifieke verbeteringen, wat mogelijk kan helpen verklaren waarom er in deze meta-regressie geen effect werd gevonden van trainingsdosis. In lijn hiermee resulteerde een sensitiviteitsanalyse waarbij studies met vijf of minder CCT sessies geëxcludeerd werden niet in betekenisvol verschillende resultaten. De vraag naar de optimale trainingsdosis blijft bijgevolg een belangrijke klinische vraag, die verdere toetsing vergt. Toekomstig onderzoek kan zich richten op het benodigde aantal sessies voor duurzame effecten, alsook op de optimale timing voor het doorlopen van deze sessies (bijvoorbeeld wanneer depressieve klachten sterk aanwezig zijn). De rol van boostersessies die kunnen worden doorlopen op cruciale momenten kan tevens worden onderzocht, waarbij een meer individuele benadering de effectiviteit van CCT voor het aanpakken van depressieve kwetsbaarheid zou kunnen vergroten.

In tegenstelling tot onze initiële hypothese vond een subgroepanalyse geen effect van *populatietype* op mate van verbetering in depressieve klachten na aPASAT-training, wat wil zeggen dat er geen verschillen in effectiviteit werden gevonden tussen depressieve populaties, voorheen depressieve populaties, risicopopulaties en gezonde populaties. Voor de secundaire uitkomstmaat, ruminatie, vonden we wel significante verschillen op basis van populatietype. Die wezen op een groter effect onmiddellijk na de training in MDD, RMD en risicopopulaties, in tegenstelling tot gezonde individuen. Dit is in lijn met onze hypothese dat populaties die typisch meer uitgesproken cognitieve klachten rapporteren meer baat hebben bij cognitieve training. Daarbij is het mogelijk onvoldoende om zich blind te staren op één of enkele klinische karakteristieken bij het inschatten van effecten van aPASAT-training voor de individuele patiënt. Recent onderzoek waarbij gebruikgemaakt werd van latente profielanalyse toont bijvoorbeeld aan dat effecten van aPASAT-training gemodereerd worden door hoe men zich verhoudt tot een combinatie van verschillende psychologische variabelen (een 'profiel'),

waaronder mate van ervaren cognitieve klachten, maar ook diverse indicatoren voor emotieregulatie, ernst van internaliserende symptomatologie en bredere indicatoren (Hoorelbeke et al., 2022b).

Om vergelijking van de klinische effectiviteit van aPASAT-training met andere gangbare interventies voor depressie mogelijk te maken, werd de gepoolde effectgrootte van depressieve symptomatologie onmiddellijk na de training ( $g = 0,29$ ) geconverteerd in een NNT van 6,15. Dit betekent dat ongeveer zes individuen behandeld dienen te worden met CCT om één patiënt te laten verbeteren. Dit is van een vergelijkbare orde van grootte als eerder geobserveerde effecten van gangbare behandelingen voor depressie, zoals antidepressieve medicatie ( $g = 0,30$ , NNT = 5,95; Khan & Brown, 2015) of psychotherapie ( $g = 0,25$ , NNT = 7,13; Cuijpers et al., 2013). In de klinische praktijk gebeurt het vaak dat een combinatie van antidepressieve medicatie en psychotherapie wordt toegepast. Een recente meta-analyse die deze combinatie onderzocht vond een gepoolde effectgrootte van  $g = 0,43$ , wat overeenkomt met een NNT van 4,19 (Cuijpers et al., 2014).

CCT lijkt in te spelen op een aantal processen die een (toenemende) kwetsbaarheid voor depressie inhouden en die minder vatbaar zijn voor verandering door gangbare interventies. In combinatie met de mogelijkheid tot groot-schalige en laagdrempelige online implementatie van CCT (bijvoorbeeld als huiswerkopdracht binnen, voorafgaand aan of volgend op een cognitief-gedragstherapeutische behandeling) maakt dit CCT tot een beloftevolle interventie voor het aanpakken van depressieve kwetsbaarheid. Daarbij zijn we van mening dat deze interventie beschouwd dient te worden als complementair aan andere gangbare interventies, eerder dan als een op zich staande interventie. Eerste studies wijzen bijvoorbeeld op de meerwaarde van CCT in combinatie met mindfulness om repetitief negatief denken aan te pakken (Course-Choi et al., 2017). Ander onderzoek waarin CCT gecombineerd werd met een gedragstherapeutische interventie voor depressie vond dan weer geen toegevoegde waarde van CCT boven op gedragsactivatie (Moshier & Otto, 2017). Daarbij dient echter opgemerkt te worden dat in voorgaand onderzoek typisch gebruikgemaakt werd van kleine steekproefgrootten, die niet toelaten om de op basis van deze meta-analyse te verwachten relatief kleine doch betekenisvolle effecten van CCT op ruminatie of depressie vast te stellen. Toekomstig onderzoek dient dus nog uit te wijzen of de combinatie van CCT met andere behandelingen in de context van depressie tot synergie zou kunnen leiden (Van den Bergh et al., 2018).

## CONCLUSIE

.....

Deze meta-analyse wijst op bescheiden doch gunstige effecten van aPASAT-training op depressieve klachten en ruminatie, die stabiel lijken op langere termijn. Gegeven de relatief lage kosten van online interventies en de hoge

toegankelijkheid ervan, kan klinische implementatie van aPASAT-training worden overwogen als strategie om cognitieve kwetsbaarheid voor depressie aan te pakken. Toekomstig onderzoek dient daarbij echter nog na te gaan voor wie en onder welke omstandigheden aPASAT-training optimaal ingezet kan worden.

*Deze studie werd gepreregistreerd op PROSPERO met identificatienummer CRD42021245971. Data en code om de resultaten te reproduceren kan gevonden worden op Open Science Framework (OSF): <https://osf.io/dhqfk>.*

*Dit onderzoek werd gefinancierd door een beurs voor Toegepast Biomedisch onderzoek met een primair Maatschappelijke finaliteit (TBM) door het Fonds voor Wetenschappelijk Onderzoek – Vlaanderen (FWO), toegekend aan het PrevenD 2.0 project (Tooo72oN). KH werd gefinancierd door het FWO (12J9722N).*

**Yannick Vander Zwalm, David Demeester, Kristof Hoorelbeke en Ernst H. W.**

**Koster** zijn verbonden aan de Vakgroep Experimenteel-Klinische en Gezondheidspsychologie, Universiteit Gent.

*Correspondentieadres:* Yannick Vander Zwalm, Universiteit Gent, Faculteit Psychologie en Pedagogische Wetenschappen, Henri Dunantlaan 2, 9000 Gent, België.

E-mail: [y.vander.zwalm@ugent.be](mailto:y.vander.zwalm@ugent.be)

**Summary** *Effectiveness of adaptive cognitive training for depression vulnerability: A systematic review and meta-analysis*

Cognitive control training (CCT), and the adaptive Paced Auditory Serial Addition Task (aPASAT) in particular, shows high potential as an intervention to remediate cognitive complaints and depressive symptoms. This systematic review and meta-analysis evaluated the effectiveness of aPASAT-training and potential moderators. For this purpose, we searched PsycINFO, MEDLINE, Embase, Web of Science and Cochrane Library for studies evaluating effects of aPASAT-training on depressive symptomatology and rumination.

Nineteen studies (N = 1255 participants) were included in the current meta-analysis, among which studies evaluating effects of aPASAT-training on depressive symptomatology and rumination in individuals suffering from major depressive disorder, remitted depressed individuals, other at-risk groups, and healthy participants. We observed small, significant effects of CCT on depressive symptomatology and rumination immediately following training and at follow-up. Subgroup analyses suggest aPASAT-training to be more effective for (remitted) depressed patients and individuals belonging to at-risk groups, compared to healthy individuals.

**Keywords** *cognitive training, cognitive control, adaptive Paced Auditory Serial Addition Task, depression, meta-analysis*

## Referenties

142

- Brunoni, A., Boggio, P., De Raedt, R., Benseñor, I., Lotufo, P., Namur, V., Valiengo, L., & Vanderhasselt, M. (2014). Cognitive control therapy and transcranial direct current stimulation for depression: A randomized, double-blinded, controlled trial. *Journal of Affective Disorders*, 162, 43-49. <https://doi.org/10.1016/j.jad.2014.03.026>
- Calkins, A. W., McMorran, K. E., Siegle, G. J., & Otto, M. W. (2015). The effects of computerized cognitive control training on community adults with depressed mood. *Behavioural and Cognitive Psychotherapy*, 43, 578-589. <https://doi.org/10.1017/S1352465814000046>
- Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., Leucht, S., Ruhe, H. G., Turner, E. H., Higgins, J. P. T., Egger, M., Takeshima, N., Hayasaka, Y., Imai, H., Shinohara, K., Tajika, A., Ioannidis, J. P. A., & Geddes, J. R. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: A systematic review and network meta-analysis. *The Lancet*, 391, 1357-1366. [https://doi.org/10.1016/S0140-6736\(17\)32802-7](https://doi.org/10.1016/S0140-6736(17)32802-7)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cohen, N., Mor, N., & Henik, A. (2015). Linking executive control and emotional response. *Clinical Psychological Science*, 3, 15-25. <https://doi.org/10.1177/2167702614530114>
- Course-Choi, J., Saville, H., & Derakshan, N. (2017). The effects of adaptive working memory training and mindfulness meditation training on processing efficiency and worry in high worriers. *Behaviour Research and Therapy*, 89, 1-13. <https://doi.org/10.1016/j.brat.2016.11.002>
- Cuijpers, P., Sijbrandij, M., Koole, S. L., Andersson, G., Beekman, A. T., & Reynolds, C. F. (2014). Adding psychotherapy to antidepressant medication in depression and anxiety disorders: A meta-analysis. *World Psychiatry*, 13, 56-67. <https://doi.org/10.1002/wps.20089>
- Cuijpers, P., Turner, E. H., Mohr, D. C., Hofmann, S. G., Andersson, G., Berking, M., & Coyne, J. (2013). Comparison of psychotherapies for adult depression to pill placebo control groups: A meta-analysis. *Psychological Medicine*, 44, 685-695. <https://doi.org/10.1017/S0033291713000457>
- Demeyer, I., De Lissnyder, E., Koster, E. H., & De Raedt, R. (2012). Rumination mediates the relationship between impaired cognitive control for emotional information and depressive symptoms: A prospective study in remitted depressed adults. *Behaviour Research and Therapy*, 50, 292-297. <https://doi.org/10.1016/j.brat.2012.02.012>
- Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology & Community Health*, 52, 377-384. <https://doi.org/10.1136/jech.52.6.377>
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455-463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315, 629-634. <https://doi.org/10.1136/bmj.315.7109.629>

- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Perception & Psychophysics*, 16, 143-149. <https://doi.org/10.3758/bf03203267>
- Ferrari, G. R. A., Vanderhasselt, M. A., Rinck, M., Demeyer, I., De Raedt, R., Beisel, S., Lindenmeyer, J., & Becker, E. S. (2021). A cognitive control training as add-on treatment to usual care for depressed inpatients. *Cognitive Therapy and Research*. <https://doi.org/10.1007/s10608-020-10197-y>
- Gautam, M., Tripathi, A., Deshmukh, D., & Gaur, M. (2020). Cognitive behavioral therapy for depression. *Indian Journal of Psychiatry*, 62, 223. [https://doi.org/10.4103/psychiatry.indianj-psychiatry\\_772\\_19](https://doi.org/10.4103/psychiatry.indianj-psychiatry_772_19)
- Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: Current status and future directions. *Annual Review of Clinical Psychology*, 6, 285-312. <https://doi.org/10.1146/annurev.clinpsy.121208.131305>
- Gotlib, I. H., Krasnoperova, E., Yue, D. N., & Joormann, J. (2004). Attentional biases for negative interpersonal stimuli in clinical depression. *Journal of Abnormal Psychology*, 113, 127-135. <https://doi.org/10.1037/0021-843x.113.1.121>
- Gronwall, D. M. A. (1977). Paced Auditory Serial-Addition Task: A measure of recovery from concussion. *Perceptual and Motor Skills*, 44, 367-373. <https://doi.org/10.2466/pms.1977.44.2.367>
- Hoorelbeke, K., & Koster, E. H. W. (2017). Internet-delivered cognitive control training as a preventive intervention for remitted depressed patients: Evidence from a double-blind randomized controlled trial study. *Journal of Consulting and Clinical Psychology*, 85, 135-146. <https://doi.org/10.1037/ccp0000128>
- Hoorelbeke, K., Koster, E. H. W., Demeyer, I., Loeys, T., & Vanderhasselt, M. A. (2016). Effects of cognitive control training on the dynamics of (mal)-adaptive emotion regulation in daily life. *Emotion*, 16, 945-956. <https://doi.org/10.1037/em00000169>
- Hoorelbeke, K., Koster, E. H., Vanderhasselt, M. A., Callewaert, S., & Demeyer, I. (2015). The influence of cognitive control training on stress reactivity and rumination in response to a lab stressor and naturalistic stress. *Behaviour Research and Therapy*, 69, 1-10. <https://doi.org/10.1016/j.brat.2015.03.010>
- Hoorelbeke, K., Van den Bergh, N., De Raedt, R., Wichers, M., Albers, C. J., & Koster, E. H. W. (2022a). Regaining control of your emotions? Investigating the mechanisms underlying effects of cognitive control training for remitted depressed patients. *Emotion*. <https://doi.org/10.1037/em00001067>
- Hoorelbeke, K., Van den Bergh, N., De Raedt, R., Wichers, M., & Koster, E. H. W. (2021). Preventing recurrence of depression: Long-term effects of a randomized controlled trial on cognitive control training for remitted depressed patients. *Clinical Psychological Science*. <https://doi.org/10.1177/2167702620979775>
- Hoorelbeke, K., Vervaeke, J., Siegle, G. J., Baeken, C., & Koster, E. H. (2022b). Individual differences associated with treatment adherence and transfer effects following gamified web-based cognitive control training for repetitive negative thinking. *Internet Interventions*, 27, 100507. <https://doi.org/10.1016/j.invent.2022.100507>
- Iacoviello, B. M., Wu, G., Alvarez, E., Huryk, K., Collins, K. A., Murrough, J. W., Iosifescu, D. V., & Charney, D. S. (2014). Cognitive-emotional training as an intervention for major depressive disorder. *Depression and Anxiety*, 31, 699-706. <https://doi.org/10.1002/da.22266>

- Kessler, R. C. (2012). The costs of depression. *Psychiatric Clinics of North America*, 35, 1-14. <https://doi.org/10.1016/j.psc.2011.11.005>
- Khan, A., & Brown, W. A. (2015). Antidepressants versus placebo in major depression: An overview. *World Psychiatry*, 14, 294-300. <https://doi.org/10.1002/wps.20241>
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55, 352-358. <https://doi.org/10.1037/h0043688>
- Koster, E. H., De Lissnyder, E., Derakshan, N., & De Raedt, R. (2011). Understanding depressive rumination from a cognitive science perspective: The impaired disengagement hypothesis. *Clinical Psychology Review*, 31, 138-145. <https://doi.org/10.1016/j.cpr.2010.08.005>
- Koster, E. H., Hoorelbeke, K., Onraedt, T., Owens, M., & Derakshan, N. (2017). Cognitive control interventions for depression: A systematic review of findings from training studies. *Clinical Psychology Review*, 53, 79-92. <https://doi.org/10.1016/j.cpr.2017.02.002>
- Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry*, 59, 990-996. <https://doi.org/10.1016/j.biopsych.2005.09.014>
- Lass, A. N., Rokke, P. D., & Winer, E. S. (2021). Evaluating cognitive control training on symptoms of depression over time: Three potential mechanisms. *Journal of Affective Disorders Reports*, 4, 100127. <https://doi.org/10.1016/j.jadr.2021.100127>
- Lauder, N. H., Minkov, R., Davey, C. G., Finke, C., Gavelin, H. M., & Lampit, A. (2021). *Computerized cognitive training in people with depression: A systematic review and meta-analysis of randomized clinical trials*. MedRxiv. <https://doi.org/10.1101/2021.03.23.21254003>
- LeMoult, J., & Gotlib, I. H. (2019). Depression: A cognitive perspective. *Clinical Psychology Review*, 69, 51-66. <https://doi.org/10.1016/j.cpr.2018.06.008>
- Lewis, E. J., Blanco, I., Raila, H., & Joormann, J. (2019). Does repetitive negative thinking affect attention? Differential effects of worry and rumination on attention to emotional stimuli. *Emotion*, 19, 1450-1462. <https://doi.org/10.1037/em00000535>
- Lorimer, B., Delgadillo, J., Kellett, S., & Brown, G. (2019). Exploring relapse through a network analysis of residual depression and anxiety symptoms after cognitive behavioural therapy: A proof-of-concept study. *Psychotherapy Research*, 30, 650-661. <https://doi.org/10.1080/10503307.2019.1650980>
- Moshier, S. J., & Otto, M. W. (2017). Behavioral activation treatment for major depression: A randomized trial of the efficacy of augmentation with cognitive control training. *Journal of Affective Disorders*, 210, 265-268. <https://doi.org/10.1016/j.jad.2017.01.003>
- Motter, J. N., Pimontel, M. A., Rindskopf, D., Devanand, D. P., Doraiswamy, P. M., & Sneed, J. R. (2016). Computerized cognitive training and functional recovery in major depressive disorder: A meta-analysis. *Journal of Affective Disorders*, 189, 184-191. <https://doi.org/10.1016/j.jad.2015.09.022>
- Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology*, 109, 504-511. <https://doi.org/10.1037/0021-843x.109.3.504>
- Porter, R. J., Bourke, C., Carter, J. D., Douglas, K. M., McIntosh, V. V. W., Jordan, J., Joyce, P. R., & Frampton, C. M. A. (2015). No change in neuropsychological dysfunction or emo-

- tional processing during treatment of major depression with cognitive-behaviour therapy or schema therapy. *Psychological Medicine*, 46, 393-404. <https://doi.org/10.1017/S0033291715001907>
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Rose, E., & Ebmeier, K. (2006). Pattern of impaired working memory during major depression. *Journal of Affective Disorders*, 90, 149-161. <https://doi.org/10.1016/j.jad.2005.11.003>
- Sari, B. A., Koster, E. H. W., Pourtois, G., & Derakshan, N. (2016). Training working memory to improve attentional control in anxiety: A proof-of-principle study using behavioral and electrophysiological measures. *Biological Psychology*, 121, 203-212. <https://doi.org/10.1016/j.biopsycho.2015.09.008>
- Segrave, R., Arnold, S., Hoy, K., & Fitzgerald, P. (2014). Concurrent cognitive control training augments the antidepressant efficacy of tDCS: A pilot study. *Brain Stimulation*, 7, 325-331. <https://doi.org/10.1016/j.brs.2013.12.008>
- Siegle, G. J., Ghinassi, F., & Thase, M. E. (2007). Neurobehavioral therapies in the 21st century: Summary of an emerging field and an extended example of cognitive control training for depression. *Cognitive Therapy and Research*, 31, 235-262. <https://doi.org/10.1007/s10608-006-9118-6>
- Siegle, G. J., Price, R. B., Jones, N. P., Ghinassi, F., Painter, T., & Thase, M. E. (2014). You gotta work at it: Pupillary indices of task focus are prognostic for response to a neurocognitive intervention for rumination in depression. *Clinical Psychological Science*, 2, 455-471. <https://doi.org/10.1177/2167702614536160>
- Sommer, A., & Plewnia, C. (2021). Depression treatment by tDCS-enhanced cognitive control training: A test of two stimulation intensities. *Brain Stimulation*, 14, 987-989. <https://doi.org/10.1016/j.brs.2021.06.004>
- Story, T. J., Potter, G. G., Attix, D. K., Welsh-Bohmer, K. A., & Steffens, D. C. (2008). Neurocognitive correlates of response to treatment in late-life depression. *The American Journal of Geriatric Psychiatry*, 16, 752-759. <https://doi.org/10.1097/jgp.0b013e31817e739a>
- Tombaugh, T. N. (2006). A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Archives of Clinical Neuropsychology*, 21, 53-76. <https://doi.org/10.1016/j.acn.2005.07.006>
- Van den Bergh, N., Hoorelbeke, K., De Raedt, R., & Koster, E. H. (2018). Remediation of depression-related cognitive impairment: Cognitive control training as treatment augmentation. *Expert Review of Neurotherapeutics*, 18, 907-913. <https://doi.org/10.1080/1737175.2018.1537783>
- Van den Bergh, N., Vermeersch, S., Hoorelbeke, K., Vervaeke, J., De Raedt, R., & Koster, E. H. W. (2020). Cognitive control training as an augmentation strategy to CBT in the treatment of fear of failure in undergraduates. *Cognitive Therapy and Research*, 44, 1199-1212. <https://doi.org/10.1007/s10608-020-10129-w>
- Vander Zwalm, Y., Liebaert, E., Hoorelbeke, K., Nève de Mévergnies, C., Baeken, C., Verhaeghe, N., & Koster, E. H. W. (2023). Treatment response following adaptive PASAT training for depression vulnerability: A systematic review and meta-analysis. *Neuropsychology Review*. <https://doi.org/10.1007/s11065-023-09581-8>
- Vanderhasselt, M. A., De Raedt, R., Namur, V., Lotufo, P. A., Bensenor, I. M., Boggio, P. S., & Brunoni, A. R. (2015).

- Transcranial electric stimulation and neurocognitive training in clinically depressed patients: A pilot study of the effects on rumination. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 57, 93-99. <https://doi.org/10.1016/j.pnpbp.2014.09.015>
- Vanderhasselt, M. A., Demeyer, I., Van Imschoot, L., Hoorelbeke, K., & De Raedt, R. (2020) Cognitive control training in healthy older adults: A proof of concept study on the effects on cognitive functioning, emotion regulation and affect. *Cognitive Therapy and Research*. <https://doi.org/10.1007/s10608-020-10154-9>
- Vervaeke, J., Hoorelbeke, K., Baeken, C., & Koster, E. H. W. (2021). Online cognitive control training for remitted depressed individuals: A replication and extension study. *Cognitive Therapy and Research*. <https://doi.org/10.1007/s10608-021-10238-0>
- Vervaeke, J., Hoorelbeke, K., Baeken, C., Van Looy, J., & Koster, E. H. W. (2020). Transfer and motivation after cognitive control training for remitted depression in healthy sample. *Journal of Cognitive Enhancement*, 4, 49-61. <https://doi.org/10.1007/s41465-019-00135-6>
- Viswanathan, M., Patnode, C. D., Berkman, N. D., Bass, E. B., Chang, S., Hartling, L., Murad, M. H., Treadwell, J. R., & Kane, R. L. (2018). Recommendations for assessing the risk of bias in systematic reviews of health-care interventions. *Journal of Clinical Epidemiology*, 97, 26-34. <https://doi.org/10.1016/j.jclinepi.2017.12.004>

# Universele schalen voor testuitslagen

## Een pleidooi voor T-scores en percentiel rangorde scores<sup>1</sup>

EDWIN DE BEURS, JAN BOEHNKE & EIKO FRIED

### Samenvatting

De diversiteit aan meetinstrumenten die in gebruik zijn binnen de huidige ggz-praktijk bemoeilijkt de interpretatie van testuitslagen en de implementatie van Routine Outcome Monitoring (ROM). In dit artikel raden we aan bij het bespreken van testuitslagen met cliënten scores om te zetten naar universele schalen. De berekening van T-scores en percentiel rangorde scores wordt uitgelegd. We stellen voor om testresultaten uit te drukken als T-scores, met de algemene bevolking als referentiegroep. T-scores kunnen aangevuld worden met percentiel rangorde scores die gebaseerd zijn op een relevante klinische referentiegroep, om ze begrijpelijker te maken voor cliënten. We maken de voordelen van deze benadering inzichtelijk met gegevens uit drie veelgebruikte depressievragenlijsten. We bieden een Excel-file met oversteekformules voor diverse vragenlijsten om eenvoudig ruwe scores naar T-scores en percentiel rangorde scores om te zetten. Recent is voorgesteld om bij gesubsidieerd onderzoek een beperkt aantal meetinstrumenten verplicht te stellen. Dit kan de ontwikkeling en toepassing van nieuwe instrumenten belemmeren. Als alternatief raden we aan om de voorgestelde universele schalen te gebruiken, teneinde testuitslagen te harmoniseren en zo ROM in de dagelijkse praktijk te vergemakkelijken en te bevorderen.

*Trefwoorden: universele schaal, T-score, percentiel rangorde score, zelfrapportagevragenlijsten, testuitslag, depressie, ROM*

1 Dit artikel is een bewerkte vertaling van de Beurs, Boehnke en Fried (2022).

**Kernboodschappen voor de klinische praktijk**

- ▶ Het gebruik van universele schalen helpt om de uitslag van tests over de ernst van klachten aan cliënten te verduidelijken en kan hun betrokkenheid bij ROM vergroten.
- ▶ Met behulp van universele schalen kan de vooruitgang die bereikt is met de behandeling (of het ontbreken daarvan) gemakkelijker worden vastgesteld, begrepen en besproken.
- ▶ Twee universele schalen worden uitgelegd: genormaliseerde T-scores en percentiel rangorde scores, en van beide worden de voor- en nadelen besproken.

## INLEIDING

.....

Meten is de basis van wetenschappelijk onderzoek. Sinds de tijd van Wundt en Thurstone heeft meten een prominente rol gespeeld in de psychologie. Er is in de psychologische wetenschap een aparte tak gewijd aan onderzoek naar meetmethoden en meetinstrumenten: de psychometrie. De gereedschapskist van de psychometrist is goed gevuld. De afgelopen decennia laten een wildgroei aan meetinstrumenten zien voor een breed scala aan psychologische concepten.

Enerzijds weerspiegelt het grote aantal beschikbare meetinstrumenten het belang van psychometrie in de klinische psychologie. Nieuwe meetinstrumenten komen vaak voort uit nieuwe theoretische perspectieven op de fenomenologie of behandeling van psychopathologie (bijvoorbeeld uit de verschuiving van aandacht voor gedragsmatige aspecten naar cognitieve). Of ze duiden op verandering in wat we als de belangrijkste kenmerken van een stoornis beschouwen (Bohnke & Croudace, 2015). Voorbeelden zijn het onderscheid tussen internaliserende en externaliserende symptomen bij kinderen en jeugdigen, de opkomst van het paniekconcept bij angststoornissen of het dimensionele model voor persoonlijkheidspathologie. Er zijn meetinstrumenten gevalideerd voor verschillende populaties (de algemene populatie versus klinische groepen) en voor verschillende doeleinden (diagnostiek versus het volgen van de voortgang van de behandeling; McPherson & Armstrong, 2021; Patalay & Fried, 2020). Ook weerspiegelen nieuwe meetinstrumenten innovaties in de meettechnologie zelf (bijvoorbeeld *computer adapted testing*; Williams et al., 2017).

Anderzijds vormt het grote aantal meetinstrumenten een uitdaging voor psychologisch wetenschappelijk onderzoek, want de diversiteit ervan bemoeilijkt het vergelijken van onderzoeksresultaten. Zo zijn er minstens 19 instrumenten om woede te meten (Weidman et al., 2017) en zijn er meer dan 280 instrumenten voor depressie ontwikkeld, waarvan er vele nog steeds in gebruik zijn (Fried, 2017; Santor et al., 2006). Er is weinig consistentie in de meetinstrumenten die worden gebruikt in therapie-effectonderzoek, zoals opgemerkt door Ogles (2013): in 163 onderzoeken rapporteerden on-

derzoekers resultaten op 435 unieke uitkomstmaten, waarvan er 371 slechts eenmalig in een onderzoek werden gebruikt. Deze overvloed aan meetinstrumenten – in combinatie met andere twijfelachtige meetpraktijken, zoals het wisselen van uitkomstmaten (Weston et al., 2016) en een gebrek aan transparantie over hoe scores worden afgeleid (Weidman et al., 2017) – helpt niet om een consistent wetenschappelijk fundament (*knowledge base*) onder de klinische psychologie tot stand te brengen. Het bemoeilijkt de vergelijking van resultaten van verschillende onderzoeken en vertraagt voortgang in de wetenschap (Flake & Fried, 2020).

In de klinische praktijk worden meetinstrumenten steeds vaker gebruikt om bij de start van de behandeling de ernst van de klachten, het functioneren en het welbevinden van cliënten te meten, en om vervolgens periodiek de voortgang van de behandeling te monitoren. In Nederland en Vlaanderen noemen we dit Routine Outcome Monitoring (ROM). (Een mooie introductie tot ROM is te vinden op <https://kennisnet.vgct.nl/bericht/een-goede-introductie-van-rom>; Schaffrat en collega's (2022) beschrijven een casus met een ROM-systeem dat men in Trier gebruikt.) Het meten van de resultaten van zorgverlening in de klinische ggz-praktijk wordt bepleit in tal van redactionele bijdragen en overzichtsartikelen (Boehnke & Rutherford, 2021; Fortney et al., 2017; Harding et al., 2011; Lambert, 2007; Lewis et al., 2019; Snyder et al., 2012). De Nederlandse ggz was rond 2000 een voorloper in deze beweging met ROM (de Beurs et al., 2011), dat inmiddels navolging heeft gevonden met PROMs en PREMs in de medische zorg in ziekenhuizen (<http://iqprom.nl>), gespecialiseerde klinieken (zie de Nederlandse Kanker Federatie: <https://nfk.nl/promprem>) en de eerstelijnszorg ([www.nivel.nl/nl/dossiers/dossier-prems-en-proms-kwaliteit-meten-vanuit-patientenperspectief](http://www.nivel.nl/nl/dossiers/dossier-prems-en-proms-kwaliteit-meten-vanuit-patientenperspectief); zie ook de website van het Zorginstituut Nederland: [www.zorginzicht.nl/ondersteuning/prom-wijzer/over-de-prom-wijzer](http://www.zorginzicht.nl/ondersteuning/prom-wijzer/over-de-prom-wijzer)).

Paradoxaal genoeg vormt de overvloed aan meetinstrumenten een belemmering voor het gebruik ervan in de klinische praktijk (Fried, 2017; Santor et al., 2006). Ten eerste bemoeilijkt de grote verscheidenheid aan instrumenten, elk met hun eigen schaal, de communicatie tussen professionals. Zo is de ernst van de symptomen van depressie uitgedrukt in een score op de Beck Depression Inventory (BDI-II; Beck et al., 1996), met een spreiding in scores van 0 tot 63, niet compatibel met de ernst uitgedrukt in een score op de Patient Health Questionnaire (PHQ-9; Kroenke & Spitzer, 2002) met een spreiding in scores van 0 tot 27. Dit compliceert de doorverwijzing van een cliënt door een behandelaar die de BDI-II gebruikt naar een collega die meer bekend is met de PHQ-9. Ten tweede wordt de communicatie over testresultaten tussen behandelaars en cliënten bemoeilijkt door het gebruik van verschillende schalen (Snyder et al., 2019). Wanneer alleen klinici testresultaten goed kunnen interpreteren, vergroot dat ongewild de kenniskloof tussen klinici en cliënten. Tegenwoordig pleit men juist voor meer betrokkenheid van de cliënt en voor gezamenlijke besluitvorming over de koers van de behandeling (Patel et al.,

2008), ook omdat dit leidt tot betere resultaten (Lambert & Harmon, 2018). Wanneer we cliënten een gelijkwaardiger rol willen geven in het therapeutische proces op hun weg naar hun herstel, werkt een kenniskloof over de betekenis van testuitslagen belemmerend. Duidelijke informatie over de betekenis van een testuitslag en hoeveel vooruitgang er is geboekt in de richting van behandeldoelen zal cliënten helpen om hen beter bij de behandeling te betrekken, zal hun een actievere rol geven en zal hun betrokkenheid bij de verstrekte informatie versterken (Goetz, 2010). Ten slotte wordt de implementatie van meten in de dagelijkse klinische praktijk gehinderd wanneer veel ervaring met meetinstrumenten en kennis over hun specifieke psychometrische eigenschappen nodig is voor een juiste interpretatie van testuitslagen. Dit kan ertoe leiden dat professionals in de ggz hun klinische oordeel blijven stellen boven de meetresultaten van gestandaardiseerde meetinstrumenten, ondanks het feit dat dit klinische oordeel vaak vertekend wordt door wensdenken en notoir onbetrouwbaar is (Kahneman et al., 2021; Meehl, 1954). Overigens zijn we niet van mening dat een testuitslag het klinische oordeel zou moeten *vervangen*, maar vinden we dat naast de testuitslag zowel het oordeel van de clinicus als dat van de cliënt zelf van belang blijft bij het nemen van beslissingen over het continueren, opschalen, afschalen of beëindigen van de behandeling.

Een strategie om de interpretatie van testresultaten in de klinische praktijk te vergemakkelijken bestaat al tientallen jaren: het gebruik van universele schalen. Met een schaal bedoelen we hier een systeem dat de eenheden beschrijft waarin meetresultaten worden uitgedrukt en geïnterpreteerd. Een voorbeeld is het metrische systeem dat rond 1790 is ingevoerd, met schalen om afstand, volume en gewicht te beschrijven. In de medische wereld zijn universele schalen vaak gebaseerd op fysieke entiteiten, zoals de meter, kilogram, mol en Kelvin. In de psychologie meten we echter meestal niet-fysieke entiteiten. We meten abstracte concepten of latente trekken, zoals depressie, angst en zelfvertrouwen – allemaal zaken die we niet direct kunnen observeren (Flake & Fried, 2020; Kellen et al., 2021). Bijgevolg kunnen we niet terugvallen op fysieke entiteiten, maar drukken we meetresultaten uit op een schaal die geen fysieke tegenhanger heeft, zoals wanneer we de ernst van depressie uitdrukken als de afstand tot de gemiddelde score van een referentiegroep in standaardeenheden (de standaardafwijking). Bij psychologische beoordelingen zetten we zo ruwe testuitslagen om naar een algemeen toepasbare universele schaal, zoals standaardscores of percentiel rangorde scores, die gebaseerd zijn op normatieve steekproeven (Kendall et al., 1999).

Tal van leerboeken over psychologisch testen beschrijven het omzetten van ruwe scores in universele schalen (Anastasi, 1968; Cronbach, 1984; Urbina, 2014), maar het wordt nog onvoldoende benut in klinisch onderzoek en bij psychologische behandelingen. Een voorbeeld van hoe het gebruik van een universele schaal goed heeft uitgepakt, kennen we uit een ander vakgebied, namelijk dat van de psychologische beoordeling of psychodiagnostiek: het meten van intellectuele vaardigheid. De IQ-score is meer dan een eeuw geleden

ontwikkeld door Binet en Simon (1907) en heeft universele toepassing gevonden; zowel professionele psychologen als leken begrijpen (in grote lijnen) de betekenis van IQ-scores, ongeacht het specifieke meetinstrument dat wordt gebruikt om intellectuele vaardigheid te meten. De interpretatie van testresultaten in de klinische praktijk zou veel gemakkelijker worden wanneer de betekenis van testcores net zo vanzelfsprekend zou zijn als die van IQ-scores.

De rest van dit artikel is als volgt opgebouwd. Eerst beschrijven we twee universele schalen: T-scores en percentiel rangorde scores (PR-scores). Beide schalen worden vaak gebruikt in neuropsychologisch en educatief onderzoek, en worden ook steeds vaker gebruikt voor klinische beoordeling. We beschrijven T-scores en PR-scores in detail en presenteren hun belangrijkste voor- en nadelen. In een technische bijlage bij dit artikel beschrijven we hoe deze schalen op diverse manieren kunnen worden afgeleid. We bieden ook een Excel-file met formules voor diverse meetinstrumenten om ruwe schaalcores om te zetten naar T-scores en PR-scores. We illustreren het gebruik ervan met meetinstrumenten voor depressie en bespreken de keuze van de juiste referentiegroep voor deze schalen. Ten slotte stellen we verdere stappen voor die nodig zijn voor een brede implementatie van deze schalen en doen we suggesties voor toekomstig onderzoek.

KANDIDAATSCHALEN

.....

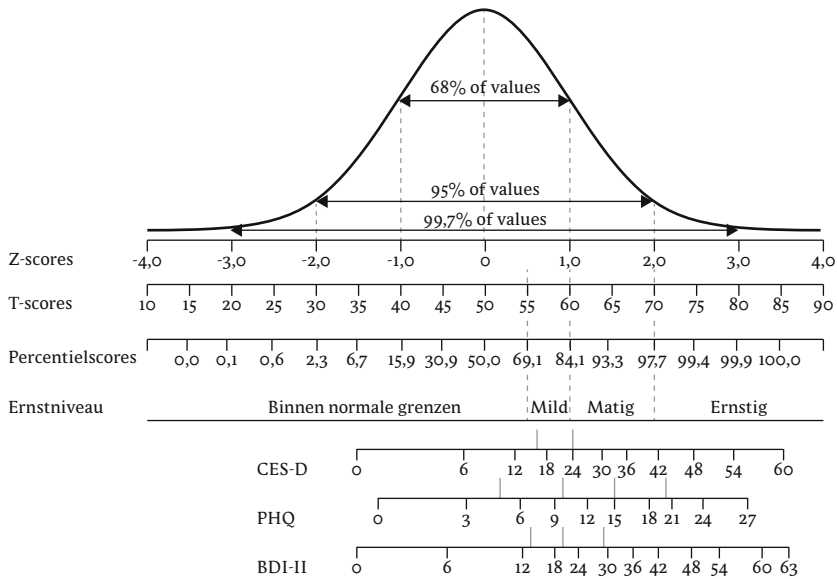
Als universele schalen voor psychologische concepten zijn twee opties voorgesteld: (1) standaardscores (Z-scores en alternatieven, zoals T-scores, stanines en stens) en (2) PR-scores (Crawford & Garthwaite, 2009; Ley, 1972). Tabel 1 geeft een overzicht van deze schalen, hoe ze berekend kunnen worden en enkele eigenschappen.

TABEL 1 *Universele schalen om testcores te standaardiseren en harmoniseren*

|          | Berekening                      | M   | SD     | Theoretisch bereik |
|----------|---------------------------------|-----|--------|--------------------|
| Z-score  | $Z = (X - M) / SD$              | 0   | 1      | -INF tot INF       |
| Stanine  | $S = Z * 2 + 5$                 | 5   | 2      | 1 tot 9            |
| Sten     | $S = Z * 2 + 5,5$               | 5,5 | 2      | 1 tot 10           |
| T-score  | $T = Z * 10 + 50$               | 50  | 10     | -INF tot INF       |
| PR-score | $PR = \frac{CF - (0,5 * F)}{N}$ | 50  | n.v.t. | 0 tot 100          |

*Noot.* M = gemiddelde; SD = standaarddeviatie; CF = cumulatieve frequentie; F = frequentie; N = totaal aantal observaties; INF = oneindig; n.v.t. = niet van toepassing

Figuur 1, geïnspireerd op Seashore (1955), toont de verschillende schalen onder de normaalverdeling en geeft labels voor de interpretatie van de scores bij het gebruik als ernstindicator of als screener. Om bijvoorbeeld betekenis te geven aan de ernst van depressie, heeft de PROMIS-groep de conventie gevolgd om de algemene bevolking in vier segmenten op te delen, van een lage naar een hoge score, respectievelijk 69,1%, 15,0%, 13,6% en 2,3%. Labels en bijbehorende T-scores zijn: 'binnen normale grenzen' ( $T < 55,0$ ), 'mild' ( $T = 55,0-59,9$ ), 'matig' ( $T = 60,0-69,9$ ) en 'ernstig' ( $T \geq 70,0$ ) ([www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points](http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points)).



FIGUUR 1 Normaalverdeling met Z-scores, T-scores en PR-scores, en overeenkomstige ruwe scores van de Amerikaanse algemene bevolking op de Center for Epidemiological Studies – Depressieschaal (CES-D), Patient Health Questionnaire (PHQ-9) en Beck Depression Inventory (BDI-II)

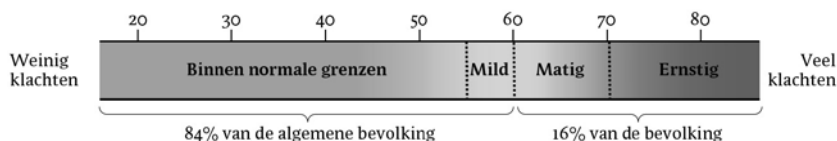
Figuur 1 bevat ook ruwe scores op de PHQ-9 (Kroenke & Spitzer, 2002), de Center for Epidemiological Studies – Depressieschaal (CES-D; Radloff, 1977) en de BDI-II (Beck & Steer, 1987) van de Amerikaanse bevolking (Choi et al., 2014) om te laten zien hoe die zich verhouden tot deze universele schalen. Voor alle drie instrumenten geldt dat de intervallen tussen ruwe scores aan de lage kant van de schaal verder uit elkaar liggen dan in het midden of aan de hoge kant van de schaal. Dit illustreert de rechts scheve frequentieverdeling van ruwe scores in de algemene bevolking: lage scores komen vaker voor dan hoge scores.

*T-scores*

153

T-scores (zo genoemd door McCall (1922) om de psychometrische pioniers Thorndike, Terman en Thurstone te eren) zijn gebaseerd op gestandaardiseerde of Z-scores. Z-scores zijn ruwe scores omgezet naar een standaardschaal met een gemiddelde van 0 en een standaarddeviatie van 1. Ze worden berekend op basis van het gemiddelde en de standaarddeviatie van een referentiegroep (zie tabel 1). Standaardisatie naar Z-scores levert echter onhandige scores op met een bereik van -3 tot 3, hetgeen negatieve scores impliceert, en meerdere decimalen zijn vereist om voldoende precisie aan te geven. Er zijn alternatieven voorgesteld met een handiger formaat, zoals stans, stanines en T-scores. Stans en stanines geven een nogal grove indeling van scoreniveaus en we laten ze hier verder buiten beschouwing. T-scores zijn Z-scores vermenigvuldigd met 10 en daaraan 50 punten toegevoegd. Ze hebben zo een gemiddelde van 50 en een standaarddeviatie van 10, en hebben in de praktijk een bereik van 20 tot 80. Het theoretisch bereik van T-scores loopt van -oneindig tot +oneindig, maar T-scores lager dan 20 of hoger dan 80 zijn zeldzaam; Figuur 1 laat zien dat een T-score van 80 drie standaarddeviaties boven het gemiddelde is, een score die volgens de cumulatieve normale verdeling door slechts 0,13% van de respondenten wordt behaald; 99,7% scoort in het bereik van 20-80.

Al geruime tijd zijn T-scores de voorkeursschaal binnen psychodiagnostiek of een assessment. Om Cronbach (1984) te citeren: ‘Confusion results from the plethora of scales. In my opinion, test developers should use the system with mean 50 and s.d. 10 unless there are strong reasons for adopting a less familiar scale’ (p. 100). (Vertaling: ‘De overdaad aan schalen kan tot verwarring leiden. Naar mijn mening zouden testontwikkelaars de schaal moeten gebruiken met een gemiddelde van 50 en een SD van 10, tenzij er heel goede redenen zijn om een minder bekende alternatieve schaal te gebruiken.’) Voor de interpretatie van T-scores zijn praktische richtlijnen opgesteld. Bij het begin van de behandeling zullen de meeste cliënten een T-score hebben tussen 65 en 75; met de behandeling kan men streven naar een score onder de 55, wat als een redelijke grenswaarde geldt voor herstel bij veel meetinstrumenten die de ernst van psychopathologie meten (Aschenbrand et al., 2005; Cella et al., 2014; de Beurs, Carlier, & van Hemert, 2019; Recklitis & Rodriguez, 2007). Uit onderzoek blijkt dat veel cliënten de voorkeur geven aan kleurcodering van scoreniveaus volgens een *heat map* van groen via geel naar rood (Brundage et al., 2015; van Muilekom et al., 2021). Figuur 2 laat zien hoe de betekenis van T-scores kan worden overgebracht naar cliënten of collega’s.



FIGUUR 2 De betekenis van T-scores voor het ernstniveau van klachten; bij weergave in kleur wordt links een groene en rechts een rode achtergrond kleur toegepast (bron: [www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points](http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points))

T-scores zijn goed ingeburgerd in de klinische psychometrie. Ze werden gekozen als schaal door het PROMIS-initiatief, dat gericht is op het ontwikkelen van een nieuwe reeks meetinstrumenten in het gezondheidszorgonderzoek (Cella et al., 2010). Op dit gebied is al veel werk verzet, bijvoorbeeld in het Prosetta Stone-initiatief (Choi et al., 2014; Schalet et al., 2015). Hier zijn ruwe scores omgezet naar de PROMIS T-scoreschaal voor veelgebruikte meetinstrumenten van psychologische concepten, waaronder depressie (Choi et al., 2014), angst (Schalet et al., 2014) en pijn (Cook et al., 2015). Ook andere onderzoekers hebben T-scores gekozen als universele schaal, bijvoorbeeld voor depressie (Wahl et al., 2014), angst (Rose & Devine, 2014), fysiek functioneren (Oude Voshaar et al., 2019), vermoeidheid (Friedrich et al., 2019) en persoonlijkheidspathologie (Zimmermann et al., 2020).

### PR-scores

De PR-score geeft voor elke ruwe score het percentage gevallen aan dat een lagere (of gelijke) score heeft dan die ruwe score. PR-scores geven zo met een waarde van 0 tot 100 de relatieve positie aan van de geteste persoon ten opzichte van zijn lotgenoten uit een referentiegroep (Kurtz & Mayo, 1979). In de klinische context laat zich dit vertalen in een boodschap aan de cliënt, zoals: 'Ten minste 75% van onze cliënten heeft een lagere score bij aanvang van de behandeling.' PR-scores geven zo de uitzonderlijkheid van de ruwe testuitslag aan met een percentage, wat helpt om de betekenis op een eenvoudige en directe manier aan collega's en cliënten over te brengen. Dit verklaart ook waarom PR-scores veel worden gebruikt in het onderwijs, bijvoorbeeld bij de Cito-toets. PR-scores worden weergegeven in figuur 1 onder de T-scoreschaal.

De literatuur over het gebruik van PR-scores in de klinische praktijk is beperkt. Crawford en Garthwaite (2009) hebben het gebruik ervan gepropageerd voor klinische (neuro)psychologische toepassingen. Crawford en collega's hebben PR-scores en betrouwbaarheidsintervallen gepubliceerd voor verschillende depressie- en angstmetingen op basis van Australische steekproeven (Crawford et al., 2011) en Britse steekproeven (Crawford et al., 2009) en een computerprogramma beschikbaar gesteld om deze scores te berekenen. Ze publiceerden ook oversteektabellen om ruwe scores om te

zetten in PR-scores (Crawford et al., 2011; Crawford & Henry, 2003). Recent zijn ruwe scores en PR-scores gepubliceerd voor de Inventory of Depression- and Anxiety Symptoms (IDAS-II; Nelson et al., 2015).

#### OVERSTEEKTABELLEN

.....

Om klinici te helpen de ruwe testscore van een individuele cliënt op een universele schaal uit te drukken, zijn voor veel meetinstrumenten oversteektabellen gepubliceerd (bijvoorbeeld: Batterham et al., 2018; Choi et al., 2014; Zimmermann et al., 2020). Ook handleidingen voor meetinstrumenten bieden soms oversteektabellen, waaronder de Minnesota Multiphasic Personality Inventory (MMPI-2; Butcher et al., 1989), de Brief Symptom Inventory (BSI; Derogatis, 1975) en de Child Behaviour Checklist (CBCL; Achenbach, 1991). Het Prosetta Stone-initiatief ([www.prosettastone.org](http://www.prosettastone.org)) heeft oversteektabellen gepubliceerd voor veel meetinstrumenten naar de PROMIS T-score, die is gebaseerd op de Amerikaanse algemene bevolking. T-scores kunnen ook worden afgeleid met behulp van de webapp [www.commonmetrics.org](http://www.commonmetrics.org) (Fischer & Rose, 2016).

T-scores of PR-scores opzoeken in tabellen is omslachtig en foutgevoelig. Daarom zijn er ook oversteekformules opgesteld voor de rekenkundige conversie van ruwe scores naar T-scores (of PR-scores). De relatie tussen de twee kolommen getallen in een oversteektabel (zoals tabel 2, van ruwe score naar T-score of van ruwe score naar PR-score) kan gemodelleerd worden met statistische software, zoals *curve-fitting* in de regressiemodule van SPSS, of de nls- en nlstools-pakketten voor *non-linear least squares* modellering in R (Baty et al., 2015). Uit een dergelijke analyse rolt een formule om de ruwe score om te zetten naar een T-score of PR-score. Een voorbeeld van conversie met een formule wordt gegeven door Roelofs en collega's (2013) en door de Beurs, Oudejans en Terluin (2022). In het laatste artikel wordt verwezen naar R software om formules te verwerven waarmee T- en PR-scores kunnen worden bepaald ([www.psycharchives.org/en/item/86e598e9-4828-4127-86ae-5f0d18e9586a](http://www.psycharchives.org/en/item/86e598e9-4828-4127-86ae-5f0d18e9586a)). Men dient dan wel te beschikken over normatieve gegevens van cliënten en de algemene Nederlandse bevolking. Recent hebben we in dit tijdschrift deze methode ook gedemonstreerd voor de Mental Health Continuum (MHC-SF) (de Beurs, Kosterman et al., 2022). Voor een aantal in Nederland gebruikte vragenlijsten hebben we een Excel-bestand samengesteld met oversteekformules (zie figuur 3 en [https://github.com/EdwindeBeurs/ScoringQuestionnaires/blob/main/Scoring\\_Vragenlijsten.xlsx](https://github.com/EdwindeBeurs/ScoringQuestionnaires/blob/main/Scoring_Vragenlijsten.xlsx)). Een gebruiker kan hier ruwe schaalscores (of itemscores) invoeren en het Excel-sheet berekent vervolgens T- en PR-scores voor cliënten, vaak apart voor mannen en vrouwen of leeftijdsgroepen. Het is de bedoeling dat dit

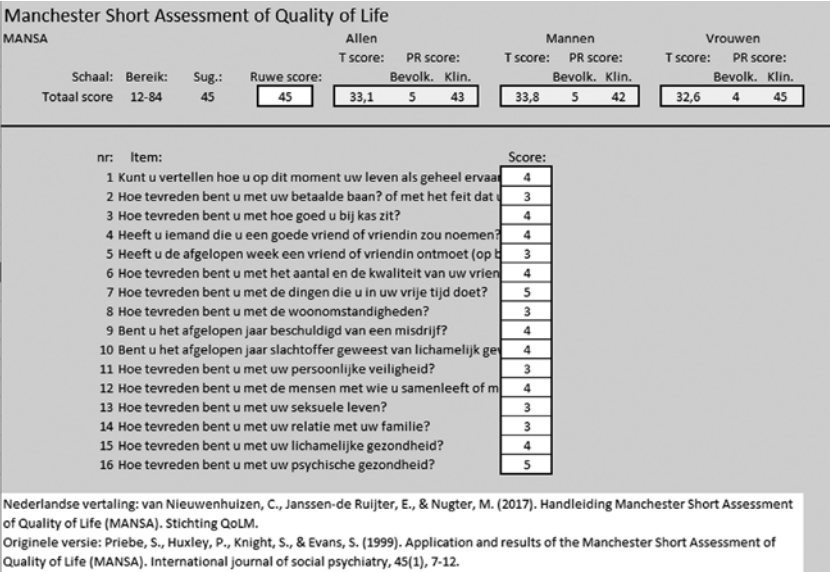


TABEL 2 Ruwe schaa scores en PROMIS T-scores op de CES-D, PHQ-9 en de BDI-II (Choi et al., 2014), en BDI-II PR-scores van de Nederlandse bevolking en van een klinische steekproef

156

| CES-D |                  |      | PHQ-9 |               |      | BDI-II |          |      |        |         |
|-------|------------------|------|-------|---------------|------|--------|----------|------|--------|---------|
| RS    |                  | T    | RS    |               | T    | RS     |          | T    | PR-pop | PR-clin |
| 0     | afwezig tot mild | 34,5 | 0     | minimaal      | 37,4 | 0      | minimaal | 34,9 | 10,1   | 0,5     |
| 2     |                  | 41,1 | 1     |               | 42,7 | 2      |          | 42,3 | 26,7   | 1,7     |
| 4     |                  | 44,7 | 2     |               | 45,9 | 4      |          | 46,2 | 38,5   | 3,4     |
| 6     |                  | 47,5 | 3     |               | 48,3 | 6      |          | 48,9 | 48,6   | 5,5     |
| 8     |                  | 49,8 | 4     |               | 50,5 | 8      |          | 51,0 | 57,3   | 7,9     |
| 10    | matig            | 51,7 | 5     | mild          | 52,5 | 10     | mild     | 52,7 | 64,1   | 10,8    |
| 12    |                  | 53,4 | 6     |               | 54,2 | 12     |          | 54,2 | 69,7   | 14,5    |
| 14    |                  | 54,8 | 7     |               | 55,8 | 14     |          | 55,6 | 74,4   | 18,4    |
| 16    |                  | 56,2 | 8     |               | 57,2 | 16     |          | 56,9 | 78,2   | 22,6    |
| 18    |                  | 57,4 | 9     |               | 58,6 | 18     |          | 58,2 | 81,3   | 27,7    |
| 20    | ernstig          | 58,6 | 10    | matig         | 59,9 | 20     | matig    | 59,3 | 84,0   | 33,3    |
| 22    |                  | 59,7 | 11    |               | 61,1 | 22     |          | 60,5 | 86,3   | 38,7    |
| 24    |                  | 60,8 | 12    |               | 62,3 | 24     |          | 61,6 | 88,5   | 44,8    |
| 26    |                  | 61,8 | 13    |               | 63,5 | 26     |          | 62,7 | 90,3   | 51,2    |
| 28    |                  | 62,9 | 14    |               | 64,7 | 28     |          | 63,8 | 91,6   | 57,5    |
| 30    |                  | 63,9 | 15    | matig ernstig | 65,8 | 30     |          | 64,8 | 92,7   | 63,6    |
| 32    |                  | 64,9 | 16    |               | 66,9 | 32     |          | 65,8 | 94,0   | 69,2    |
| 34    |                  | 66,0 | 17    |               | 68,0 | 34     |          | 65,8 | 95,2   | 74,0    |
| 36    |                  | 67,0 | 18    |               | 69,2 | 36     |          | 67,6 | 96,0   | 78,3    |
| 38    |                  | 68,1 | 19    |               | 70,3 | 38     |          | 68,9 | 96,8   | 82,5    |
| 40    |                  | 69,2 | 20    | ernstig       | 71,5 | 40     | ernstig  | 69,9 | 97,4   | 86,3    |
| 42    |                  | 70,4 | 21    |               | 72,7 | 42     |          | 70,9 | 98,0   | 89,5    |
| 44    |                  | 71,7 | 22    |               | 74,0 | 44     |          | 71,9 | 98,5   | 92,3    |
| 46    |                  | 73,0 | 23    |               | 75,3 | 46     |          | 72,9 | 98,9   | 94,6    |
| 48    |                  | 74,4 | 24    |               | 76,7 | 48     |          | 74,0 | 99,2   | 96,3    |
| 50    |                  | 76,0 | 25    |               | 78,3 | 50     |          | 75,2 | 99,5   | 97,7    |
| 52    |                  | 77,7 | 26    |               | 80,0 | 52     |          | 76,4 | 99,8   | 98,6    |
| 54    |                  | 79,7 | 27    |               | 82,3 | 54     |          | 77,7 | 99,8   | 99,2    |
| 56    |                  | 82,0 |       |               |      | 56     |          | 79,1 | 99,9   | 99,6    |
| 58    |                  | 84,3 |       |               |      | 58     |          | 80,8 | 99,9   | 99,8    |
| 60    |                  | 86,4 |       |               |      | 60     |          | 82,9 | 100,0  | 99,9    |
|       |                  |      |       |               |      | 62     |          | 85,1 | 100,0  | 100,0   |
|       |                  |      |       |               |      | 63     |          | 86,3 | 100,0  | 100,0   |

Noot. BDI-II = Beck Depression Inventory; CES-D = Center for Epidemiological Studies – Depressie-schaal; PHQ-9 = Patient Health Questionnaire; RS = ruwe score; T = PROMIS Depression T-score (gebaseerd op de algemene bevolking van de VS). T-scores kunnen berekend worden uit ruwe scores (RS) met een rationale functie:  $(T = 35,7 + (3,83 * RS - 0,0023 * RS^2)) / (1 + 0,13 * RS - 0,0012 * RS^2)$ ; PR-scores kunnen berekend worden uit T-scores met:  $PR-pop = -2,9 + 103,7 / (1 + \exp(-0,162(T - 49,6)))$ ;  $PR-cl = 1,0 + 100,7 / (1 + \exp(-0,232 * (T - 62,7)))$ .



FIGUUR 3 Voorbeeld van scoring van de MANSA: ruwe score, T-score en twee PR-scores voor allen (ongeacht sekse), en voor mannen en vrouwen apart. In de kolom 'Ruwe score' kunnen schaalscores worden ingevuld. Als alternatief kunnen achter de items in de kolom 'Score' de itemscores worden ingevuld. Bij invullen van itemscores wordt de schaalscore berekend en in de kolom 'Sug.' weergegeven. Om de T- en PR-score te krijgen, moet die schaalscore vervolgens overgebracht worden naar de kolom voor de ruwe score (knippen en plakken als waarde of overtypen).

een levend document wordt om in de toekomst uit te breiden met extra vragenlijsten zodra daarvan de benodigde normgegevens voorhanden zijn.

T-SCORES EN PR-SCORES VERGELEKEN

.....

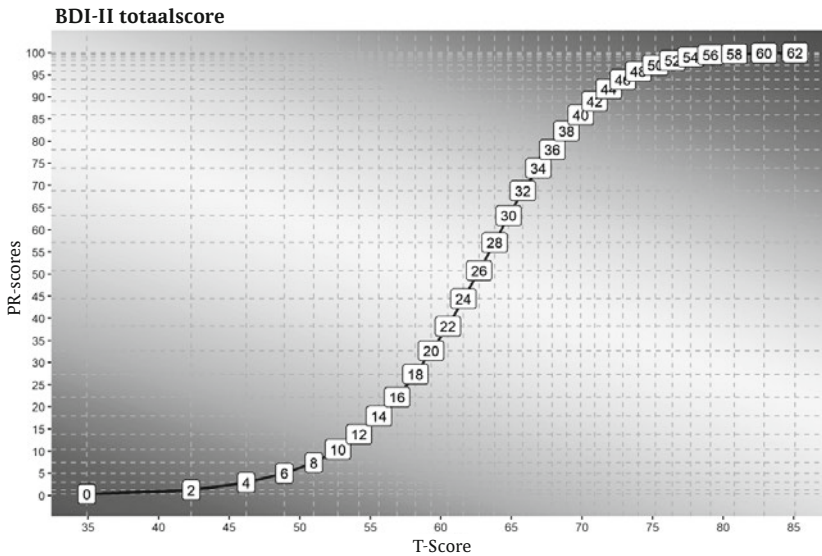
Een minimeis voor een universele schaal is dat T-scores gemakkelijk te interpreteren zijn en op een eenvoudige manier uitdrukken hoe uitzonderlijk een testuitslag is (Snyder et al., 2012). In de COSMIN-checklist wordt interpreteerbaarheid gedefinieerd als ‘the degree to which one can assign a qualitative meaning to a score’ (Prinsen et al., 2018). Veel van de problemen met meetinstrumentgebonden scoringsmethoden die in de inleiding werden genoemd kunnen zo opgelost worden. De T-score geeft de universele betekenis van een testuitslag weer met de afstand tot het gemiddelde van een referentiegroep in standardeenheden, waarbij 50 staat voor het gemiddelde en 10 punten een standaardafwijking betekenen. De interpretatie van de T-score veronderstelt bij de testgebruiker echter wel enige kennis van de normaalverdeling, zoals de ‘68-95-99,7’-regel, het ezelsbruggetje om te onthouden dat

68% van de scores binnen 1 SD van het gemiddelde valt, 95% binnen 2 SD's en 99,7% binnen 3 SD's (zie figuur 1). Dit betekent dat een T-score hoger dan 80 of lager dan 20 vrij uitzonderlijk is en alleen behaald wordt door de hoogste en laagste 0,13% van de referentiegroep. In de praktijk loopt de T-score dus van 20 tot 80. PR-scores verlenen op een meer intuïtieve manier een universele betekenis aan een testuitslag door de score uit te drukken als het percentage (0-100) respondenten met dezelfde of een lagere ruwe score.

T-scores worden gemeten op een intervalmeetschaal, ervan uitgaande dat dit een redelijke aanname is, gegeven de frequentieverdeling van de oorspronkelijke ruwe scores: de afstand van 40 naar 50 is even groot als de afstand van 50 naar 60. Dit geldt echter niet voor PR-scores. Die zitten niet op gelijke afstanden van elkaar, maar moeten als ordinale of volgordelijke scores beschouwd worden (een ordinale meetschaal). De relatie tussen T-scores en PR-scores is S-vormig. Een hoge PR-score (PR = 75) komt bijvoorbeeld overeen met een T-score van 57, wat op een relatief bescheiden afstand van het gemiddelde van 50 ligt. In het hoge bereik komen de T-scores 65, 70, 75 en 80 overeen met respectievelijk 93,3, 97,7, 99,4 en 99,9 (zie figuur 1). Tussen PR-scores zitten dus duidelijk ongelijke intervallen, vooral aan de extremen, wat bij gebruik van PR-scores leidt tot een overdrijving van verschillen tussen scores nabij het gemiddelde en onderschatting van verschillen aan de extremen.

Figuur 4 toont, voor een selectie van ruwe BDI-II-scores, hun relatie met T-scores en klinische PR-scores. (Voor de leesbaarheid van de figuur zijn alleen de even ruwe scores opgenomen.) Nederlandse gegevens zijn gebruikt om T-scores en PR-scores te verkrijgen uit een klinische steekproef (de Beurs et al., 2011). De figuur laat de S-vormige relatie tussen normaal verdeelde T-scores en PR-scores zien, wat de uitrekking van de PR-score weerspiegelt ten opzichte van de T-score aan de extremen. (Vanwege de klokvorm van de normaalverdeling zijn PR-score-intervallen smaller in het midden van de schaal en breder aan de uitersten; zie ook figuur 1.)

Als mens zijn we het meest gewend aan schalen met gelijke intervallen, zoals meters, kilogrammen, enzovoort. Als de PR-score echter ten onrechte wordt beschouwd als een schaal met gelijke intervallen, zijn conclusies over de testuitslag vaak onjuist. Deze vertekening in perceptie van PR-scores werd door Bowman (2002) aangetoond in een onderzoek met derdejaars psychologiestudenten, die ertoe neigden om PR-scores boven 80 of onder 20 als een vrij extreme testuitslag te beschouwen, terwijl onder de normaalverdeling PR = 80 overeenkomt met T = 58,4, dat wil zeggen: minder dan 1 SD verwijderd van het gemiddelde en nog niet een derde op weg naar de maximale score van T = 80 (zie figuur 1). In verband met de heersende opvatting over toelaatbare transformaties en rekenkundige en statistische bewerkingen, zijn met PR-scores eenvoudige bewerkingen niet mogelijk. Zo is een gemiddelde PR-score of een verschilscore tussen herhaalde PR-scores betekenisloos. Voor een verdere bespreking van dit onderwerp, zie: Meijer en Oos-



FIGUUR 4 Relatie tussen geselecteerde ruwe scores op de Beck Depression Inventory (BDI-II), met T-scores (x-as) en PR-scores binnen de klinische populatie (y-as); horizontale en verticale rasterlijnen corresponderen met de ruwe scores.

terloo (2008). Het grote voordeel van PR-scores, namelijk dat ze over het algemeen intuïtiever zijn, wordt door dit nadeel overschaduwd.

T-scores zijn voor de interpretatie van testresultaten en verdere verwerking van gegevens de beste keuze, maar voor een heldere communicatie met collega's en cliënten raden we aan om T-scores aan te vullen met PR-scores. T-scores zijn veelzijdiger, gezien hun psychometrische eigenschappen, maar PR-scores zijn intuïtiever te interpreteren. We wijzen er verder op dat met name extreme PR-scores vatbaar blijken voor incorrecte interpretatie. Voorzichtigheid is geboden en bij gebruik moet goede uitleg gegeven worden.

#### KEUZE VAN DE JUISTE REFERENTIEGROEP

.....

De twee gesuggereerde universele schalen, T-scores en PR-scores, hebben als doel om aan te geven hoe gewoon of hoe uitzonderlijk een testuitslag is. Hierbij is een passende referentiegroep nodig. Wat passend is, hangt af van de vraag die voorligt: willen we de ernst van de depressie van een geteste persoon weten in vergelijking met de algemene bevolking of in vergelijking met cliënten die geestelijke gezondheidszorg ontvangen? We doen hierbij uiteenlopende aanbevelingen voor T-scores en PR-scores. Voor T-scores is de algemene bevolking de beste keuze. We streven bij voorkeur naar een breed toepasbare schaal, zowel voor bijvoorbeeld epidemiologische studies

als voor toepassing in de ggz-context. Gebruik van een referentiesteekproef uit de algemene populatie maakt het mogelijk om bij diverse subgroepen binnen de ggz de ernst van het gemeten concept aan te duiden en onderling te vergelijken, omdat we voor iedereen van dezelfde referentiestandaard gebruikmaken. Een vergelijking met IQ-scores maakt deze voorkeur evident: het zou vreemd zijn om IQ-scores te normeren op personen met een verminderd intellectueel functioneren of, aan het andere uiterste, voor hoogbegaafden.

Net als voor T-scores is voor PR-scores de algemene bevolking de meest geschikte referentiegroep voor experimenteel en epidemiologisch onderzoek. Echter, voor toepassing in de klinische praktijk raden we aan om een *klinische* referentiegroep te gebruiken. Dit advies hangt samen met de eerdergenoemde vervorming van de PR-scoreschaal aan de extremen in vergelijking met de gelijke intervallen van de T-scoreschaal. Cliënten zullen, zeker aan het begin van de behandeling, vaak tot de hoogste 5 tot 10% van de algemene populatie behoren (Löwe et al., 2008; Schulte-van Maaren et al., 2013). Bijgevolg zijn PR-scores op basis van de algemene populatie vrij beperkt in bereik wanneer ze worden toegepast op cliënten die zich aanmelden voor behandeling. Hierdoor wordt het moeilijk om met PR-scores onderscheid te maken tussen cliënten met milde, matige en ernstige klachten. De S-vormige relatie tussen T-scores en PR-scores heeft ook tot gevolg dat een substantiële verandering in T-score zich aan het uiteinde van de schaal vertaalt in een kleine verandering in PR-score. Verandering in score die verbetering tijdens de behandeling weerspiegelt, wordt dus onderschat met de PR-schaal. Het uitdrukken van de score als klinische PR-score (dat wil zeggen: de PR-waarde ten opzichte van een vergelijkbare klinische groep) zal bruikbaarere informatie opleveren.

Wanneer klinische PR-scores worden gebruikt, moet wel de klinische populatie worden gespecificeerd waarvan ze zijn afgeleid (bijvoorbeeld opgenomen cliënten, poliklinische cliënten, poliklinische cliënten gezien in een privépraktijk, enzovoort). Een illustratief voorbeeld zijn de normatieve gegevens voor de Outcome Questionnaire (OQ-45; Lambert et al., 2004) die zijn gepubliceerd door Timman en collega's (2017), die normatieve gegevens leverden voor verschillende cliëntengroepen. Ook Tingey en collega's (1996) bespreken de kwestie van klinische subgroepen die variëren naar ernst. Zij stellen een oplossing voor in het kader van de benadering van Jacobson en Truax (1991) van klinisch significante verandering.

Voor een meer gedetailleerde vergelijking met subklinische cliëntengroepen (bijvoorbeeld aan het einde van een succesvolle behandeling, waarbij beide referentiegroepen gerechtvaardigd zijn), zou men zowel een algemeen populatiepercentiel als een klinisch percentiel kunnen presenteren. Men kan bijvoorbeeld rapporteren dat een cliënt een T-score van 60 heeft, wat hoog is in vergelijking met de algemene bevolking (van wie 84% een lagere score heeft), maar wel onder het gemiddelde ligt van andere cliënten die

de kliniek bezoeken (van wie 67% een hogere score heeft). De aangeboden informatie is afhankelijk van de context. Wel willen we ervoor waarschuwen de cliënt niet te overladen met te veel informatie, omdat dit verwarrend kan zijn en indruist tegen het principe van eenvoud en minimalisme: 'less is more' (Peters et al., 2007).

#### ILLUSTRATIE

.....

We illustreren het gebruik van T-scores en PR-scores met gegevens van drie verschillende meetinstrumenten die veel worden gebruikt om de ernst van depressie te meten. De oversteektabellen in tabel 2 dienen als voorbeeld om het nut van universele schalen te demonstreren. De gegevens voor tabel 2 zijn gebaseerd op de tabellen A1 tot A3 uit Choi en collega's (2014), die oversteektabellen bieden voor drie veelgebruikte depressievragenlijsten: (1) de CES-D, (2) de PHQ-9 en (3) de BDI-II. Deze oversteektabellen dienen om ruwe scores om te zetten naar de PROMIS-depressie T-scoreschaal. Tabel 2 bevat een selectie van ruwe scores, hun betekenis en T-scores, die zijn gebaseerd op de Amerikaanse algemene bevolkingssteekproef van PROMIS.

Volgens de richtlijnen voor interpretatie van ruwe schaalesscores van de CES-D (Radloff, 1977), duidt een score van 0-15 op afwezige tot milde depressie, 16-23 op matige depressie, en 24 of meer op ernstige depressie (zie tabel 2). De grens voor *caseness* (een positieve testuitslag die suggereert dat de cliënt aan depressie lijdt) is  $\geq 16$ , wat overeenkomt met een T-score van 56,2. Voor de PHQ-9 duiden scores van 0 tot 4 op minimale depressie, 5-9 op milde, 10-14 op matige, 15-19 op matig ernstige en 20-27 op ernstige depressie (Kroenke & Spitzer, 2002). De voorgestelde grens voor *caseness* is  $\geq 10$ , wat overeenkomt met een T-score van 59,9. Interpretatierichtlijnen uit de handleiding van de BDI-II duiden een totaalscore van 0-13 aan als minimale depressie, 14-19 als mild, 20-28 als matig en 29-63 als ernstig. Een recente meta-analyse suggereerde als optimale afkapwaarde voor klinische depressie de waarde  $\geq 18,2$  (von Glischinski et al., 2019). Voor deze drie maten begint *caseness* dus bij  $T = 56$  tot  $60$ , en de afkapwaarden tussen milde en matige depressie vallen over het algemeen samen met  $T = 60$ , wat door de PROMIS-groep ook werd voorgesteld voor de T-score ([www.promis.org](http://www.promis.org)). Bij het vertalen van meerdere depressiematen naar één gemeenschappelijk concept en T-scoreschaal, suggereerden Wahl en collega's (2014) voor de BDI-II  $T = 54,0$ ,  $T = 66,2$  en  $T = 72,9$  als drempels voor milde, matige en ernstige depressie. Deze afkapwaarden voor de drie vragenlijsten komen onderling overeen (maar zijn niet exact gelijk), wat mogelijk te wijten is aan verschillen in operationalisering van het concept 'depressie' of aan verschillen tussen de normatieve steekproeven. Maar de afkapwaarden in tabel 2 laten zien dat de grenswaarden 55, 60 en 70 voor mild, matig en ernstig in grote lij-

nen overeenkomen met wat auteurs voor de verschillende instrumenten als grenswaarden voor ruwe scores hebben voorgesteld.

Tabel 2 toont voor de BDI-II ook PR-scores met als referentiegroep de algemene bevolking (PR-pop) en een klinische groep (PR-clin). De gegevens zijn afkomstig uit een steekproef (N = 7500) van de Nederlandse algemene bevolking (Roelofs et al., 2013) en een steekproef van cliënten (N = 9844) die zich hadden aangemeld voor psychiatrische zorg en die deelnamen aan Routine Outcome Monitoring bij GGZ Rivierduinen (de Beurs et al., 2011). Hier wordt het verschil in bruikbaarheid duidelijk van PR-scores op basis van de algemene populatie en PR-scores van een klinische steekproef. Een persoon met een ruwe score van 20 op de BDI-II heeft PR-pop = 84,0 (en behoort daarmee tot de hoogste 16% van de algemene bevolking), maar een PR-clin van 33,3 (de laagste 33% van de klinische populatie). Een andere persoon met een ruwe score van 26 heeft een PR-pop = 90,3 (de hoogste 10% van de populatie), maar de klinische PR-score is PR-clin = 51,2, wat aangeeft dat een BDI-II-score van 26 dicht ligt bij de klinische mediane score. Bij BDI-II-scores in het hogere segment, zoals we die vinden bij cliënten, differentiëren de PR-clin-scores veel beter in ernstniveaus dan de PR-pop-scores.

## DISCUSSIE

.....

De huidige praktijk van psychologisch onderzoek, met een grote diversiteit aan instrumenten voor vergelijkbare concepten, draagt bij aan zwak gedefinieerde metingen, vergeleken met de sterk gedefinieerde metingen zoals we die kennen uit de natuurwetenschappen (Finkelstein, 2003). Dit vormt een bedreiging voor de kwaliteit van onderzoek en belemmert het tot stand komen van een robuust kennisfundament onder ons klinisch handelen. Onder professionals kan het bijdragen aan verwarring en onduidelijkheid over de betekenis van klinische testresultaten. Ten slotte – en de focus van dit artikel – belemmert dit een heldere communicatie tussen klinici en hun cliënten over de betekenis van testuitslagen.

Universele schalen zijn beschikbaar voor implementatie, zoals blijkt uit de uitvoerige lijst van referenties over hun toepassing. Ze zijn al geïmplementeerd in de Nederlandse klinische praktijk (de Beurs, 2010), waar ‘Delta T’ een veelgebruikte term is geworden om de voortgang aan te geven van cliënten van pre- tot posttest. We raden T-scores aan als universele schaal om de ernst van een score op meetinstrumenten voor klinische fenomenen uit te drukken. We herhalen hierbij ook nog eens de eerdere aanbeveling om de algemene bevolking als referentiegroep voor de T-score te gebruiken (de Beurs, Flens, & Williams, 2019). Bij cliënten die zich aanmelden voor behandeling zal de T-score hoog zijn ( $T > 50$ ) bij klachten en symptomen, of laag ( $T < 50$ ) bij functioneren, welbevinden of kwaliteit van leven. De keuze voor de algemene bevolking als referentiegroep is logisch: dezelfde keuze is ge-

maakt bij bijvoorbeeld de IQ-score. Een PR-score gebaseerd op een relevante klinische referentiegroep kan aan cliënten duidelijk maken hoe hun ernst-niveau zich verhoudt tot andere cliënten. Een PR-score spreekt intuïtief aan, maar heeft beperkingen. Het is bijvoorbeeld niet mogelijk om PR-scores van elkaar af te trekken als maat voor wat er tijdens de behandeling is bereikt.

Verder willen we twee belangrijke onderwerpen onderstrepen, elk met een aantal suggesties voor verdere ontwikkeling. Het eerste onderwerp is puur praktisch: het gebruik van universele schalen zou gemeengoed moeten zijn en de standaard voor rapportage van testuitslagen. Daar moeten verschillende stappen voor worden gezet. Ten eerste is een betere uitleg van universele schalen nodig en een demonstratie van hun nut. We hopen dat dit artikel hier een bijdrage aan levert. Een tweede belangrijke stap voor de praktische implementatie is dat voor zoveel mogelijk meetinstrumenten oversteektabellen, -formules en -figuren beschikbaar komen waarmee de testgebruiker ruwe schaalcores op een gemakkelijke manier kan omzetten naar T-scores en PR-scores, bijvoorbeeld als standaardonderdeel van de documentatie van een meetinstrument. Het Excel-bestand dat we voor dit artikel hebben samengesteld, is een eerste aanzet hiertoe. Verder moeten oversteekformules beschikbaar komen die kunnen worden ingebouwd in software die binnen de klinische praktijk wordt gebruikt om vragenlijsten af te nemen, zoals VIP-live, QuestManager, BergOp, enzovoort.

Een tweede onderwerp voor toekomstige doorontwikkeling is psychometrisch onderzoek. Een uitputtend overzicht van de resterende obstakels valt buiten het bestek van dit artikel, maar we noemen hier vier uitdagingen. Ten eerste vereist de conversie van ruwe scores naar T-scores op basis van de algemene populatie dat we kunnen beschikken over normatieve gegevens van bevolkingssteekproeven. Momenteel zijn dergelijke gegevens niet voor alle meetinstrumenten voorhanden. Een tweede uitdaging hangt samen met het feit dat beschikbare normatieve gegevens van meetinstrumenten afkomstig zijn uit specifieke landen. De PROMIS T-score is bijvoorbeeld gebaseerd op normatieve gegevens van de Amerikaanse bevolking. We zullen moeten onderzoeken of we deze normen ook internationaal kunnen toepassen en daartoe moeten gegevens in andere landen worden verzameld en scores worden vergeleken (Terwee et al., 2021). Voor Nederlandse vertalingen van de PROMIS-instrumenten is hiermee een begin gemaakt door Elsmen en collega's (Elsmen, Flens et al., 2022; Elsmen, Roorda et al., 2022). Dit zal de internationale vergelijking van uitkomstonderzoek naar behandeling in de ggz vergemakkelijken en uitkomstgegevens ook relevanter maken voor de klinische praktijk. Een derde onderzoeksuitdaging is de mogelijke invloed van de vertaling van meetinstrumenten en aanpassing aan de specifieke cultuur (Sousa & Rojjanasrirat, 2011). We moeten tevens onderzoeken of meetinstrumenten gevoelig zijn voor andere factoren, zoals geslacht, leeftijd, opleidingsniveau en/of sociaaleconomische status. Invloed van deze factoren zou impliceren dat verschillende normen (en T-scores en PR-scores) gelden

voor subgroepen van respondenten (Teresi et al., 2009; Weinfurt, 2021). Dat impliceert aanvullende conversieformules, met coëfficiënten voor geslacht, leeftijd, enzovoort, bijvoorbeeld voortvloeiend uit regressiegebaseerde normering (Timmerman et al., 2021) of alternatieve benaderingen (Lenhard & Lenhard, 2021). Voorts moet mogelijk rekening worden gehouden met andere externe variabelen, zoals de culturele achtergrond van cliënten of aspecten van hun psychopathologie (Böhnke & Croudace, 2015). In het Excel-bestand hebben we daar alvast een begin mee gemaakt door aparte formules te leveren voor mannen en vrouwen, en – waar relevant – voor leeftijdsgroepen. De vierde uitdaging ten slotte is dat conversie naar T-scores zou moeten corrigeren voor niet-normaliteit van onbewerkte testscores, om te eindigen met T-scores met echt gelijke intervallen. Om dit te bewerkstelligen, kunnen conversietabellen en -formules gebaseerd worden op de itemresponstheorie (IRT)-benadering, op de frequentieverdeling van ruwe scores (op PR-scores gebaseerde normalisatie) of op andere normalisatiebenaderingen. Welke aanpak de beste T-score oplevert, verdient nader onderzoek (Kolen & Brennan, 2014).

Het harmoniseren van het meten van psychopathologie en psychische gezondheid is een belangrijk methodologisch probleem. We hebben hier praktische oplossingen voorgesteld om meetresultaten in de klinische praktijk te rapporteren en te bespreken: twee universele schalen. Maar dat is maar een deel van de oplossing van het onderliggende probleem: de wildgroei aan meetinstrumenten in wetenschappelijk psychologisch onderzoek. Om harmonisatie te stimuleren is recent gesuggereerd een beperkte set aan meetinstrumenten verplicht te stellen om subsidie te kunnen verwerven (Wolpert, 2020). Vervolgens hebben het National Institute of Mental Health en de Wellcome Trust plannen gelanceerd (Farber et al., 2020) om een beperkte set meetinstrumenten voor onderzoek en uitkomstmeting voor te schrijven: de PHQ-9 voor depressie, de General Anxiety Disorder (GAD-7; Spitzer et al., 2006) voor angst, de Revised Child Anxiety and Depression Scale (RCADS-22; Chorpita et al., 2000) voor depressie en angst bij kinderen en adolescenten, en de World Health Organisation Disability Assessment Schedule (WHODAS; Üstün et al., 2010) voor de invloed van beperkingen op het functioneren bij volwassenen. Hoewel dit de vergelijkbaarheid en interpreteerbaarheid van onderzoeksresultaten zal vergroten, kan het ook onbedoelde negatieve gevolgen hebben (Patalay & Fried, 2020). De geselecteerde instrumenten zijn niet zonder gebreken of nadelen. De PHQ-9 en GAD-7 bijvoorbeeld zijn ontwikkeld om te screenen op depressie en angst bij de algemene bevolking, niet om de voortgang tijdens de behandeling te volgen. Beide vragenlijsten zijn vrij kort, wat hun bereik en meetprecisie beperkt, en hun betrouwbaarheid en nauwkeurigheid vermindert, waardoor ze minder geschikt zijn om (statistisch) betrouwbare verandering (Jacobson & Truax, 1991) in de klinische praktijk te volgen, wat op zijn beurt kan leiden tot de voorbarige conclusie dat er geen klinische verandering is opgetreden. Ook

kan het verplichte gebruik van slechts een paar meetinstrumenten de verdere ontwikkeling van meetinstrumenten belemmeren (zoals het PROMIS-initiatief met computer adapted testing; Cella et al., 2010) of de verbetering van bestaande meetinstrumenten of meetmethoden tegenwerken. Toch is enige harmonisatie van uitkomstmaten in de klinische praktijk op zijn plaats en heeft dit belangrijke initiatieven gestimuleerd, zoals het International Consortium for Health Outcome Measurement (ICHOM), dat standaard meetsets heeft voorgesteld voor allerlei aandoeningen, waaronder depressie en angst (Obbarius et al., 2017). Dit kan helpen om de wildgroei aan meetinstrumenten voor *measurement-based care* te beperken en bijdragen aan een verbeterslag in de kwaliteit van de geestelijke gezondheidszorg (Fortney et al., 2017; Kilbourne et al., 2018).

In een ander deel van de literatuur wordt getracht beter uit te leggen wat we nu feitelijk nastreven met universele schalen: er is een verschil tussen enerzijds het vergelijkbaar maken van testresultaten die met uiteenlopende instrumenten zijn gemeten, en anderzijds het louter uitdrukken van ruwe schaalscores op een universele schaal. Testscores uitdrukken op een universele schaal kan twee doelen dienen: (1) het vergelijkbaar maken van scores van twee tests die bedoeld zijn om hetzelfde te meten (*parallel tests*), en (2) het uitdrukken van ruwe scores van meetinstrumenten die verschillende dingen meten op dezelfde schaal. Deze twee doelstellingen moeten niet met elkaar worden verward. Bij het vergelijkbaar maken van testuitslagen van verschillende tests is een achterliggende aanname dat hetzelfde concept wordt gemeten (Kolen & Brennan, 2014). Bij het uitdrukken van scores op een universele schaal is deze aanname niet vereist en is de operatie alleen bedoeld om testuitslagen naar een gemeenschappelijke schaal te transponeren. In het laatste geval blijven we ervan uitgaan dat we concepten meten die intrinsiek verschillend zijn, hoewel ze op dezelfde schaal worden uitgedrukt. Een T-score van 60 voor depressie betekent bijvoorbeeld niet hetzelfde als een T-score van 60 voor angst, want de twee concepten verschillen van elkaar. Sterker nog, gezien de aanzienlijke verschillen in inhoud van diverse meetinstrumenten voor depressie (zoals de CES-D en de BDI), kan het zijn dat twee dezelfde T-scores op deze meetinstrumenten niet dezelfde ernst van depressie aangeven (Fried, 2017). Daarom moeten we terughoudend zijn met het onderling vergelijken van T-scores voor verschillende concepten: ze drukken uit hoe uitzonderlijk de score is in vergelijking met een referentiegroep, zoals de algemene bevolking, maar niet meer dan dat. Hoewel T-scores belangrijke informatie bevatten, waarschuwen we dus voor het vergelijken of gelijkstellen van T-scores (of PR-scores, waar feitelijk hetzelfde voor geldt) die niet dezelfde concepten representeren. Voor een persoon die  $T = 65$  scoort op depressie en  $T = 55$  op angst, zou de juiste interpretatie en boodschap zijn: uw depressiescore wijkt meer af van het gemiddelde van de algemene bevolking (namelijk 1,5 standaardafwijking) dan uw angstscore (0,5 standaardafwijking). Anders gezegd: uw depressiescore is uitzonderlijk-

ker dan de angstscores. Dit wijkt subtiel maar wezenlijk af van de foutieve interpretatie dat de cliënt meer depressief dan angstig is.

Verwarring over universele schalen en universele concepten kan ook ontstaan uit de onderzoeksliteratuur over de PROMIS-meetinstrumenten en de PROMIS T-scoreschaal. Voor diverse bestaande meetinstrumenten zijn oversteektabellen van ruwe schaalscores naar de PROMIS T-scoreschaal opgesteld. Dit onderzoek is meestal uitgevoerd per concept, zoals depressie (Choi et al., 2014), angst (Schalet et al., 2014), pijn (Cook et al., 2015) en fysiek functioneren (Schalet et al., 2015), allemaal met in de titel de formulering ‘Establishing a common metric for ...’. Doel van deze onderzoeken was om telkens *binnen een concept* (bijvoorbeeld depressie) testuitslagen van verschillende meetinstrumenten gelijk te stellen met de T-scoreschaal van de PROMIS-pendant. De PROMIS T-scores zijn allemaal gekalibreerd op de Amerikaanse algemene populatie. Er zijn meer studies die deze benadering per concept gebruiken – zie bijvoorbeeld voor depressie Wahl en collega’s (2014), voor fysiek functioneren Oude Voshaar en collega’s (2019), voor persoonlijkheidspathologie Zimmermann en collega’s (2020), voor psychische problemen Batterham en collega’s (2018) en voor vermoeidheid Lai en collega’s (2014). Maar ook hier geldt dat het gebruik van dezelfde schaal niet betekent dat identieke concepten zijn gemeten. De T-score drukt voor elk concept op een handzame manier de uitzonderlijkheid van de score uit (voor bijvoorbeeld depressie), maar dit betekent niet dat deze scores tussen concepten vergelijkbaar zijn.

Een andere standaardisatie is dat de scores op PROMIS-schalen voor het gemak allemaal in dezelfde richting gaan, dat wil zeggen: een hogere score vertegenwoordigt meer van het gemeten concept, zoals depressie, pijn, mobiliteit, functioneren of kwaliteit van leven. In visuele presentaties zijn ze allemaal op dezelfde manier kleur gecodeerd: groen is normaal of gezond, rood is disfunctioneel, ziek of ernstig (zie figuur 2).

Ten slotte is het belangrijk om erop te wijzen dat bij het omzetten van ruwe schaalscores naar T-scores de oorspronkelijke ruwe scores niet verloren gaan. De somscore van een stel items kan nog steeds worden berekend en gerapporteerd in onderzoek, bijvoorbeeld om scores te vergelijken met eerder gepubliceerde onderzoeksresultaten.

## CONCLUSIE

.....

Concluderend stellen we: (1) dat het gebruik van universele schalen, met name op de algemene bevolking gebaseerde T-scores en klinische PR-scores, nuttig zijn om de meetpraktijk in klinisch psychologisch onderzoek te harmoniseren, (2) dat ze behulpzaam zijn voor een juiste interpretatie van testuitslagen, (3) dat ze de communicatie over testuitslagen tussen professionals onderling kunnen verbeteren, en (4) dat ze het makkelijker maken

om de betekenis van testuitslagen goed uit te leggen aan cliënten in de geestelijke gezondheidszorg.

Een eerste stap om dit te stimuleren is om in handleidingen voor meetinstrumenten oversteektabellen en -formules op te nemen voor de omzetting van ruwe scores naar T-scores. Tegenwoordig worden zelfrapportagevragenlijsten meestal via de computer of smartphone afgenomen en kunnen testuitslagen eenvoudig in T-scores worden uitgedrukt. Aanvullende stappen zijn professionals opleiden en ervaring laten opdoen met de T-scoreschaal, en cliënten voorlichten over de betekenis van T-scores. Gaandeweg zal er zo 'gevoel' ontstaan bij professionals en cliënten voor de betekenis van de T-score. Het algemeen gebruik van IQ-scores en begrip door het grote publiek over wat IQ-scores overbrengen, laat zien dat deze aanpak haalbaar is. We hopen dat dit artikel het gebruik van T-scores stimuleert als de standaard om betekenis te geven aan (neuro)psychologische testuitslagen.

#### De belangrijkste aanbevelingen voor de klinische praktijk

- ▶ Bespreek testuitslagen als T-scores aan de hand van figuur 2, met de algemene bevolking als normgroep ('Uw T-score is 71, wat aangeeft dat u ernstige klachten hebt').
- ▶ Illustreer de ernst van de aandoening volgens het testresultaat mede met een klinische PR-score ('Uw percentielscore is 90, wat betekent dat slechts 10% van onze cliënten een hogere score heeft.')
- ▶ Belangrijke afkapwaarden voor de T-scoreschaal zijn 70, 60 en 55 (voor de overgangen tussen ernstig, matig, mild en normaal). Zie ook figuur 2.
- ▶ Vijf punten verandering in T-score komt overeen met een verandering van een halve standaardafwijking, wat een verandering is die door de meeste cliënten ook als zodanig ervaren wordt.  $T < 55$  kan duiden op herstel.
- ▶ T-scores en PR-scores zijn op te zoeken in de handleiding van meetinstrumenten of te benaderen met relatief eenvoudige formules. Voor een aantal meetinstrumenten leveren we ze bij dit artikel in de vorm van een Excel-bestand.

#### BIJLAGE: DIVERSE METHODEN OM T-SCORES EN PR-SCORES TE VERKRIJGEN

.....

##### *T-score via lineaire omzetting*

.....

De meest rechtstreekse benadering is de conversie van ruwe scores naar T-scores met de lineaire formule  $T = (X - M/SD) * 10 + 50$ , waarbij X de ruwe score is, en M en SD respectievelijk het gemiddelde en de standaarddeviatie van de referentiepopulatie. Dit is een goede aanpak als de ruwe scores een

normale verdeling hebben. Vaak wordt niet aan deze voorwaarde voldaan en is de verdeling rechts scheef vanwege een oververtegenwoordiging van lage scores. Bij een scheve frequentieverdeling van scores zijn het rekenkundig gemiddelde en de standaarddeviatie ongeschikte descriptoren en is de interpretatie van de testuitslag in standaardscores niet meer eenduidig. Eerst dienen de ruwe scores te worden genormaliseerd en dan omgezet naar T-scores.

Ruwe scores met een afwijkende frequentieverdeling kunnen getransformeerd worden naar een normaalverdeling of genormaliseerd worden (Box & Cox, 1964; Liu et al., 2009). Een alternatieve, grondiger benadering is om eerst genormaliseerde standaardscores vast te stellen met op regressie gebaseerde normering (Zachary & Gorsuch, 1985), waarbij ook rekening kan worden gehouden met achtergrondvariabelen, zoals sekse, leeftijd, opleidingsniveau, enzovoort (Lenhard & Lenhard, 2021). Statistische software is beschikbaar in R (GAMLSS; Stasinopoulos en collega's (2018) en Timmerman en collega's (2021) bieden een gedetailleerde handleiding). Genormaliseerde standaardscores ( $Z$ ) worden vervolgens omgezet in T-scores met  $T = 10 * Z + 50$  (de lineaire formule in tabel 1).

#### *T-scores volgens de benadering van de itemresponsstheorie*

Recent zijn op de IRT gebaseerde alternatieve benaderingen voorgesteld en beproefd om ruwe scores op bestaande meetinstrumenten om te zetten in T-scores, bijvoorbeeld met behulp van EAP-factorscores (Fischer & Rose, 2019). De IRT-benadering werd toegepast door Choi en collega's (2014) om ruwe scores op de CES-D, PHQ-9 en BDI-II om te zetten naar de PROMIS-schaal (een T-score met de Amerikaanse bevolking als referentiegroep), zoals weergegeven in figuur 1. Deze benaderingen vereisen opnieuw de aanname dat de latente trek die wordt gemeten bij benadering normaal verdeeld is, wat vaak niet het geval is wanneer klinische kenmerken, zoals depressie of angst, worden gemeten in steekproeven van de algemene bevolking. Voor toepassing van IRT bij een niet-normale verdeling van ruwe scores zijn alternatieven ontwikkeld (Reise et al., 2018). Omdat IRT-factorscores vaak worden geschaald als een standaardscore ( $M = 0$ ,  $SD = 1$ ), kunnen ze eenvoudig worden omgezet in T-scores met de lineaire formule uit tabel 1 ( $T = 10 * Z + 50$ ).

#### *Op PR-score gebaseerde omzetting naar T-scores*

Een andere methode om ruwe scores naar T-scores te converteren is op basis van PR-scores (*equipercentile linking*). Uitgangspunt is ook hier een ruwe score van een unidimensionele schaal en een frequentieverdeling die de normaalverdeling voldoende benadert (Lord & Wingersky, 1984). De conversie op basis van PR-scores en de conversie die op IRT is gebaseerd leiden tot vergelijkbare resultaten, zoals aangetoond door Choi en collega's (2014) en meer recentelijk door Schalet en collega's (2021) bij het meten van depres-

sie, en door Brady en collega's (2022) bij het meten van de ernst van burn-outklachten onder professionals in de zorg. T-scores zijn gevalideerd aan het oordeel van deskundigen voor de ernst van pijn, vermoeidheid, angst en depressie bij oncologische cliënten (Cella et al., 2014). Ernstniveaus kwamen bijna perfect overeen met de T-score:  $T = 60$  bleek een goede grenswaarde om matige ernst van gemiddelde ernst te onderscheiden, en  $T = 70$  voor de overgang van gemiddeld naar ernstig. Verder zijn T-scores op de BSI (Derogatis, 1975) vergeleken met de Clinical Global Impression verbeteringsscore (CGI; Guy, 1976): een verbetering van 5 T-scorepunten kwam goed overeen met verbetering volgens de CGI-verbeteringsscore (de Beurs, Carlier, & van Hemert, 2019).

### *Welke methode kiezen?*

Bij de keuze uit de besproken methoden om T-scores te verkrijgen (lineaire transformatie, PR-scores, IRT-modellen en regressiegebaseerde normering voor achtergrondvariabelen als geslacht en leeftijd), raden we aan om zo eenvoudig mogelijk te werk te gaan, tenzij complexiteit noodzakelijk is. Bij normaal verdeelde ruwe scores en geen significante verschillen tussen mannen en vrouwen of leeftijdsgroepen, volstaat de lineaire transformatie van ruwe schaalscores ( $T = 10 * Z + 50$ ). Wanneer de ruwe scores niet normaal verdeeld zijn (wat te zien is aan de frequentieverdeling van de scores, met een qq-plot is te visualiseren of met statistische software is te toetsen) moeten we eerst de scores normaliseren met een transformatie, met PR-scores of op basis van een IRT-model. Wanneer er geslachts- of leeftijdsverschillen spelen, leveren aparte conversies voor subgroepen of de regressiebenadering nauwkeurigere T-scores op. De IRT-benadering heeft nog andere voordelen boven eenvoudigweg schaalscores berekenen door itemscores te sommeren (de klassieke testtheoriebenadering), maar die bespreken we hier niet verder.

### *Berekening van PR-scores*

PR-scores worden berekend met de formule uit de laatste kolom van tabel 1. Wanneer normatieve gegevens beschikbaar zijn, wordt de frequentie van elke ruwe score gebruikt om de PR-score te bepalen met de formule  $PR = (CumF - (0,5 * F)) / N$  (Crocker & Algina, 1986), waarin de cumulatieve frequentie (CumF) het aantal scores kleiner of gelijk aan de ruwe score is, F de frequentie van de ruwe score zelf is en N het totaal aantal waarnemingen. De PR-score wordt dus berekend als de proportie respondenten met een score net onder de ruwe score plus de helft van de proportie respondenten die precies de ruwe score hebben behaald. Onder de aanname van een normale verdeling kunnen PR-scores ook worden afgeleid uit Z-scores (en T-scores) volgens de formule voor de cumulatieve normaalverdeling, beschikbaar als

statistische functie in veel softwarepakketten (R, SPSS, STATA, MS Excel). De formules kan ook worden benaderd door een exponentiële functie, die een S-vormige curve beschrijft:  $PR = 100 / 1 + e^{(-1,75 \cdot Z)}$ , waarbij Z de standaard-score is; of met  $PR = 100 / 1 + e^{(-0,175 \cdot T + 8,75)}$ , waarbij T de T-score is.

*Voor dit manuscript zijn voornamelijk eerder gepubliceerde gegevens gebruikt. Voormetingsgegevens van Routine Outcome Monitoring van een klinische steekproef van ggz-instelling Rivierduinen in Leiden werden gebruikt om een kolom toe te voegen aan tabel 2. De medisch-ethische toetsingscommissie van het Leids Universitair Medisch Centrum stemde in met het gebruik voor wetenschappelijk onderzoek van deze geanonimiseerde gegevens, die routinematig werden verzameld in de klinische praktijk. Cliënten gaven toestemming voor geanonimiseerd gebruik van hun gegevens.*

**Edwin de Beurs** is als hoogleraar verbonden aan de Sectie Klinische Psychologie, Universiteit Leiden, en aan Arkin GGZ, Amsterdam.

**Jan R. Boehnke** is als universitair hoofddocent (UHD) werkzaam bij de School of Health Sciences, University of Dundee, UK.

**Eiko I. Fried** is als universitair hoofddocent (UHD) verbonden aan de Sectie Klinische Psychologie, Universiteit Leiden.

*Correspondentieadres:* Edwin de Beurs, Universiteit Leiden, Sectie Klinische Psychologie, Wassenaarseweg 52, 2333 AK Leiden. E-mail: edwin.de.beurs@arkin.nl

### **Summary** *Universal metrics for test results: A plea to use T-scores and percentile rank scores*

The diversity of measurement instruments in use in current mental health care practice makes it difficult to interpret test results and hampers implementation of ROM. To discuss test results with clients, we recommend converting them to common metrics. We explain the calculation of T-scores and percentile rank scores and discuss their pros and cons. We propose to express test results as T-scores with the general population as the reference group. T-scores can be supplemented with percentile scores based on a relevant clinical reference group, to make them more understandable for clients. We demonstrate the benefits of this approach with data from three commonly used self-report questionnaires for depression. In addition, we provide an Excel file with crosswalk formulas for several self-report questionnaires to easily convert raw scores to T-scores and percentile scores. It has recently been proposed to require a limited number of measurement instruments in subsidized research. This could hinder the development and application of new instruments. As an alternative, we recommend using the proposed common metrics to harmonize test results and thus facilitate and promote ROM in daily practice.

**Keywords** *common metric, T-score, percentile rank order score, self-report questionnaires, test result, depression, ROM*

## Referenties

- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist 4-18 and 1991 profiles*. Department of Psychiatry, University of Vermont.
- Anastasi, A. (1968). *Psychological testing*. Macmillan.
- Aschenbrand, S. G., Angelosante, A. G., & Kendall, P. C. (2005). Discriminant validity and clinical utility of the CBCL with anxiety-disordered youth. *Journal of Clinical Child & Adolescent Psychology*, 34, 735-746. [https://doi.org/10.1207/s15374424jccp3404\\_15](https://doi.org/10.1207/s15374424jccp3404_15)
- Batterham, P. J., Sunderland, M., Slade, T., Calear, A. L., & Carragher, N. (2018). Assessing distress in the community: Psychometric properties and crosswalk comparison of eight measures of psychological distress. *Psychological Medicine*, 48, 1316-1324. <https://doi.org/10.1017/S0033291717002835>
- Baty, F., Ritz, C., & Baty, M. F. (2015). Package 'nlstools'. *Tools for nonlinear regression analysis*. <https://doi.org/10.18637/jss.v066.i05>
- Beck, A. T., & Steer, R. A. (1987). *Manual for the revised Beck Depression Inventory*. Psychological Corporation.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II Beck Depression Inventory Manual* (Vol. 2). The Psychological Corporation.
- Binet, A., & Simon, T. (1907). Le développement de l'intelligence chez les enfants. *L'Année Psychologique*, 14, 1-94. [https://doi.org/https://www.persee.fr/doc/psy\\_0003-5033\\_1907\\_num\\_14\\_1\\_3737](https://doi.org/https://www.persee.fr/doc/psy_0003-5033_1907_num_14_1_3737)
- Boehnke, J. R., & Rutherford, C. (2021). Using feedback tools to enhance the quality and experience of care. *Quality of Life Research*, 30, 3007-3013. <https://doi.org/10.1007/s11136-021-03008-8>
- Böhnke, J., & Croudace, T. J. (2015). Factors of psychological distress: Clinical value, measurement substance, and methodological artefacts. *Social Psychiatry and Psychiatric Epidemiology*, 50, 515-524. <https://doi.org/10.1007/s00127-015-1022-5>
- Bowman, M. L. (2002). The perfidy of percentiles. *Archives of Clinical Neuropsychology*, 17, 295-303. [https://doi.org/10.1016/S0887-6177\(01\)00116-0](https://doi.org/10.1016/S0887-6177(01)00116-0)
- Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26, 211-243. <https://doi.org/https://www.jstor.org/stable/2984418>
- Brady, K. J., Ni, P., Carlasare, L., Shanafelt, T. D., Sinsky, C. A., Linzer, M., Stillman, M., & Trockel, M. T. (2022). Establishing crosswalks between common measures of burnout in US physicians. *Journal of General Internal Medicine*, 37, 777-784. <https://doi.org/10.1007/s11606-021-06661-4>
- Brundage, M. D., Smith, K. C., Little, E. A., Bantug, E. T., Snyder, C. F., & PRODPsAB. (2015). Communicating patient-reported outcome scores using graphic formats: Results from a mixed-methods evaluation. *Quality of Life Research*, 24, 2457-2472. <https://doi.org/10.1007/s11136-015-0974-y>
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2 (Minnesota Multiphasic Personality Inventory-2): Manual for administration and scoring*. University of Minnesota Press.
- Cella, D., Choi, S., Garcia, S., Cook, K. F., Rosenbloom, S., Lai, J.-S., Tatum, D. S., & Gershon, R. (2014). Setting standards for severity of common symptoms in oncology using the PROMIS item banks and expert judgment. *Quality of Life Research*, 23, 2651-2661. <https://doi.org/10.1007/s11136-014-0732-6>
- Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., Devellis, R., DeWalt, D., Fries, J. F.,

- Gershon, R., Hahn, E. A., Lai, J. S., Pilkonis, P., Revicki, D., .... PROMIS Cooperative Group. (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *Journal of Clinical Epidemiology*, 63, 1179-1194. <https://doi.org/10.1016/j.jclinepi.2010.04.011>
- 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. <https://doi.org/10.1037/a0035768>
- Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A revised child anxiety and depression scale. *Behaviour Research and Therapy*, 38, 835-855. [https://doi.org/10.1016/S0005-7967\(99\)00130-8](https://doi.org/10.1016/S0005-7967(99)00130-8)
- 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. <https://doi.org/10.1007/s11136-014-0790-9>
- Crawford, J. R., Cayley, C., Lovibond, P. F., Wilson, P. H., & Hartley, C. (2011). Percentile norms and accompanying interval estimates from an Australian general adult population sample for self-report mood scales (BAI, BDI, CRS-D, CES-D, DASS, DASS-21, STAI-X, STAI-Y, SRDS, and SRAS). *Australian Psychologist*, 46, 3-14. <https://doi.org/10.1111/j.1742-9544.2010.00003.x>
- Crawford, J. R., & Garthwaite, P. H. (2009). Percentiles please: The case for expressing neuropsychological test scores and accompanying confidence limits as percentile ranks. *The Clinical Neuropsychologist*, 23, 193-204. <https://doi.org/10.1080/13854040801968450>
- Crawford, J. R., Garthwaite, P. H., Lawrie, C. J., Henry, J. D., MacDonald, M. A., Sutherland, J., & Sinha, P. (2009). A convenient method of obtaining percentile norms and accompanying interval estimates for self-report mood scales (DASS, DASS-21, HADS, PANAS, and sAD). *British Journal of Clinical Psychology*, 48, 163-180. <https://doi.org/10.1348/014466508X377757>
- Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology*, 42, 111-131. <https://doi.org/10.1348/014466503321903544>
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart, and Winston. <https://books.google.nl/books?id=tfgekQAAMAAJ>
- Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). Harper & Row.
- de Beurs, E. (2010). De genormaliseerde T-score, een 'euro' voor testuitslagen. *Maandblad Geestelijke Volksgezondheid*, 65, 684-695. [www.sbggz.nl](http://www.sbggz.nl)
- de Beurs, E., Boehnke, J., & Fried, E. I. (2022). Common measures or common metrics? A plea to harmonize measurement results. *Clinical Psychology & Psychotherapy*, 29(5), 1755-1767. <https://doi.org/10.1002/cpp.2742>
- de Beurs, E., Carlier, I. V., & van Hemert, A. M. (2019). Approaches to denote treatment outcome: Clinical significance and clinical global impression compared. *International Journal of Methods in Psychiatric Research*, 28. <https://doi.org/10.1002/mpr.1797>

- 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. <https://doi.org/10.1002/cpp.696>
- 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., Kosterman, S., Anten, S., Bohlmeijer, E., & Westerhof, G. J. (2022). Psychometrische evaluatie van de Mental Health Continuum – Short Form (MHC-SF): Constructvaliditeit, responsiviteit voor verandering, normen en T-scores. *Gedragstherapie*, 55, 131-155.
- de Beurs, E., Oudejans, S., & Terluin, B. (2022). 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*. Advance online publication. <https://doi.org/10.1027/1015-5759/a000740>
- Derogatis, L. R. (1975). *The Brief Symptom Inventory*. Clinical Psychometric Research.
- Elsman, E. B. M., Flens, G., de Beurs, E., Roorda, L. D., & Terwee, C. B. (2022). Towards standardization of measuring anxiety and depression: Differential item functioning for language and Dutch reference values of the PROMIS item banks v1.0 – Anxiety and Depression. *PLOS ONE*. <https://doi.org/10.1371/journal.pone.0273287>
- Elsman, E. B. M., Roorda, L. D., Smidt, N., de Vet, H. C., & Terwee, C. B. (2022). Measurement properties of the Dutch PROMIS-29 v2.1 profile in people with and without chronic conditions. *Quality of Life Research*, 31, 3447-3458. <https://doi.org/10.1007/s11136-022-03171-6>
- Farber, G., Wolpert, M., & Kemmer, D. (2020). *Common measures for mental health science: Laying the foundations*. Wellcome Trust. <https://wellcome.ac.uk/sites/default/files/CMB-and-CMA-July-2020-pdf.pdf>
- Finkelstein, L. (2003). Widely, strongly and weakly defined measurement. *Measurement*, 34, 39-48. [https://doi.org/10.1016/S0263-2241\(03\)00018-6](https://doi.org/10.1016/S0263-2241(03)00018-6)
- Fischer, H. F., & Rose, M. (2016). Wwww.common-metrics.org: A web application to estimate scores from different patient-reported outcome measures on a common scale. *BMC Medical Research Methodology*, 16, 142. <https://doi.org/10.1186/s12874-016-0241-0>
- Fischer, H. F., & Rose, M. (2019). Scoring depression on a common metric: A comparison of EAP estimation, plausible value imputation, and full Bayesian IRT modeling. *Multivariate Behavioral Research*, 54, 85-99. <https://doi.org/10.1080/00273171.2018.1491381>
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3, 456-465. <https://doi.org/10.1177/2515245920952393>
- Fortney, J. C., Unützer, J., Wrenn, G., Pyne, J. M., Smith, G. R., Schoenbaum, M., & Harbin, H. T. (2017). A tipping point for measurement-based care. *Psychiatric Services*, 68, 179-188. <https://doi.org/10.1176/appi.ps.201500439>
- Fried, E. I. (2017). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, 208, 191-197. <https://doi.org/10.1016/j.jad.2016.10.019>

- Friedrich, M., Hinz, A., Kuhnt, S., Schulte, T., Rose, M., & Fischer, F. (2019). Measuring fatigue in cancer patients: A common metric for six fatigue instruments. *Quality of Life Research*, 28, 1615-1626. <https://doi.org/10.1007/s11136-019-02147-3>
- Goetz, T. (2010). *It's time to redesign medical data* [Presentation]. TEDMED.
- Guy, W. (1976). *Clinical Global Impressions ECDEU assessment manual for psychopharmacology. Revised* (DHEW publication ADM 76-338). Government Printing Office.
- Harding, K. J., Rush, A. J., Arbuckle, M., Trivedi, M. H., & Pincus, H. A. (2011). Measurement-based care in psychiatric practice: A policy framework for implementation. *Journal of Clinical Psychiatry*, 72, 1136-1143. <https://doi.org/10.4088/JCP.10r06282whi>
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12-19. <https://doi.org/10.1037/0022-006x.59.1.12>
- Kahneman, D., Sibony, O., & Sunstein, C. (2021). *Ruis: Waarom we zo vaak verkeerde beslissingen nemen en hoe we dat kunnen voorkomen*. Nieuw Amsterdam.
- Kellen, D., Davis-Stober, C. P., Dunn, J. C., & Kalish, M. L. (2021). The problem of coordination and the pursuit of structural constraints in psychology. *Perspectives on Psychological Science*, 16, 767-778. <https://doi.org/10.1177/1745691620974771>
- Kendall, P. C., Marrs-Garcia, A., Nath, S. R., & Sheldrick, R. C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology*, 67, 285-299. <https://doi.org/10.1037/0022-006X.67.3.285>
- Kilbourne, A. M., Beck, K., Spaeth-Rublee, B., Ramanuj, P., O'Brien, R. W., Tomoyasu, N., & Pincus, H. A. (2018). Measuring and improving the quality of mental health care: A global perspective. *World Psychiatry*, 17, 30-38. <https://doi.org/10.1002/wps.20482>
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer Science & Business Media.
- Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, 32, 509-515. <https://doi.org/10.3928/0048-5713-20020901-06>
- Kurtz, A. K., & Mayo, S. T. (1979). Percentiles and percentile ranks. In *Statistical methods in education and psychology* (pp. 145-163). Springer. [https://doi.org/10.1007/978-1-4612-6129-2\\_6](https://doi.org/10.1007/978-1-4612-6129-2_6)
- Lai, J.-S., Cella, D., Yanez, B., & Stone, A. (2014). Linking fatigue measures on a common reporting metric. *Journal of Pain and Symptom Management*, 48, 639-648. <https://doi.org/10.1016/j.jpainsymman.2013.12.236>
- Lambert, M. J. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research*, 17, 1-14. <https://doi.org/10.1080/10503300601032506>
- Lambert, M. J., Gregersen, A. T., & Burlingame, G. M. (2004). The Outcome Questionnaire-45. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Volume 3: Instruments for adults* (3rd ed., pp. 191-234). Lawrence Erlbaum Associates Publishers. <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2004-14941-006&site=ehost-live>
- Lambert, M. J., & Harmon, K. L. (2018). The merits of implementing routine outcome monitoring in clinical practice. *Clinical Psychology: Science and Practice*, 25, e12268. <https://doi.org/10.1111/cpsp.12268>

- Lenhard, W., & Lenhard, A. (2021). Improvement of norm score quality via regression-based continuous norming. *Educational and Psychological Measurement*, 81, 229-261. <https://doi.org/10.1177/0013164420928457>
- Lewis, C. C., Boyd, M., Puspitasari, A., Navarro, E., Howard, J., Kassab, H., Hoffman, M., Scott, K., Lyon, A., & Douglas, S. (2019). Implementing measurement-based care in behavioral health: A review. *JAMA Psychiatry*, 76, 324-335. <https://doi.org/10.1001/jamapsychiatry.2018.3329>
- Ley, P. (1972). *Quantitative aspects of psychological assessment* (Vol. 1). Duckworth.
- Liu, H., Lafferty, J., & Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research*, 10. <https://doi.org/10.1145/1577069.1755863>
- Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score 'equatings'. *Applied Psychological Measurement*, 8, 453-461. <https://doi.org/10.1177/014662168400800409>
- Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, 45(3), 266-274. <https://doi.org/10.1097/MLR.0b013e318160d093>
- McCall, W. A. (1922). *How to measure in education*. MacMillan.
- McPherson, S., & Armstrong, D. (2021). Psychometric origins of depression. *History of the Human Sciences*, 09526951211009085. <https://doi.org/10.1177/09526951211009085>
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota. <https://doi.org/10.1037/11281-000>
- Meijer, R. R., & Oosterloo, S. J. (2008). A note on measurement scales and statistical testing. *Measurement: Interdisciplinary Research and Perspectives*, 6, 198-204. <https://doi.org/10.1080/15366360802324446>
- Nelson, E. C., Eftimovska, E., Lind, C., Hager, A., Wasson, J. H., & Lindblad, S. (2015). Patient reported outcome measures in practice. *BMJ Case Reports*, 350, g7818. <https://doi.org/10.1136/bmj.g7818>
- Obbarius, A., van Maasackers, L., Bear, L., Clark, D. M., Crocker, A. G., de Beurs, E., Emmelkamp, P. M. G., Furukawa, T. A., Hedman-Lagerlöf, E., Kangas, M., Langford, L., Legsage, A., Mwesigire, D. M., Nolte, S., Patel, V., Pilkonis, P. A., Pincus, H. A., Reis, R. A., Rojas, G., ... Rose, M. (2017). Standardization of health outcomes assessment for depression and anxiety: Recommendations from the ICHOM Depression and Anxiety Working Group. *Quality of Life Research*, 26, 3211-3225. <https://doi.org/10.1007/s11136-017-1659-5>
- Ogles, B. M. (2013). Measuring change in psychotherapy research. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (pp. 134-166). Wiley.
- Oude Voshaar, M., Vonkeman, H., Courvoisier, D., Finckh, A., Gossec, L., Leung, Y., Michaud, K., Pinheiro, G., Soriano, E., & Wulfraat, N. (2019). Towards standardized patient reported physical function outcome reporting: Linking ten commonly used questionnaires to a common metric. *Quality of Life Research*, 28, 187-197. <https://doi.org/10.1007/s11136-018-2007-0>
- Patalay, P., & Fried, E. I. (2020). Editorial perspective: Prescribing measures: unintended negative consequences of mandating standardized mental health measurement. *Journal of Child*

- Psychology and Psychiatry*, 62, 1032-1036. <https://doi.org/10.1111/jcpp.13333>
- Patel, S. R., Bakken, S., & Ruland, C. (2008). Recent advances in shared decision making for mental health. *Current Opinion in Psychiatry*, 21, 606-6012. <https://doi.org/10.1097/YCO.obo13e32830eb6b4>
- Peters, E., Dieckmann, N., Dixon, A., Hibbard, J. H., & Mertz, C. K. (2007). Less is more in presenting quality information to consumers. *Medical Care Research and Review*, 64, 169-190. <https://doi.org/10.1177/10775587070640020301>
- Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27, 1147-1157. <https://doi.org/10.1007/s11136-018-1798-3>
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401. <https://doi.org/10.1177/014662167700100306>
- Recklitis, C. J., & Rodriguez, P. (2007). Screening childhood cancer survivors with the Brief Symptom Inventory-18: Classification agreement with the Symptom Checklist-90-Revised. *Psycho-Oncology*, 16, 429-436. <https://doi.org/10.1002/pon.1069>
- Reise, S. P., Rodriguez, A., Spritzer, K. L., & Hays, R. D. (2018). Alternative approaches to addressing non-normal distributions in the application of IRT models to personality measures. *Journal of Personality Assessment*, 100, 363-374. <https://doi.org/10.1080/00223891.2017.1381969>
- Roelofs, J., van Breukelen, G., de Graaf, L. E., Beck, A. T., Arntz, A., & Huibers, M. J. H. (2013). Norms for the Beck Depression Inventory (BDI-II) in a large Dutch community sample. *Journal of Psychopathology and Behavioral Assessment*, 35, 93-98. <https://doi.org/10.1007/s10862-012-9309-2>
- Rose, M., & Devine, J. (2014). Assessment of patient-reported symptoms of anxiety. *Dialogues in Clinical Neuroscience*, 16, 197-211.
- Santor, D. A., Gregus, M., & Welch, A. (2006). Focus article: Eight decades of measurement in depression. *Measurement: Interdisciplinary Research and Perspectives*, 4, 135-155. [https://doi.org/10.1207/s15366359mea0403\\_1](https://doi.org/10.1207/s15366359mea0403_1)
- Schaffrath, J., Weinmann-Lutz, B., & Lutz, W. (2022). The Trier Treatment Navigator (TTN) in action: Clinical case study on data-informed psychological therapy. *Journal of Clinical Psychology*. <https://doi.org/10.1002/jclp.23362>
- 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. <https://doi.org/10.1016/j.janxdis.2013.11.006>
- Schalet, B. D., Lim, S., Cella, D., & Choi, S. W. (2021). Linking scores with patient-reported health outcome instruments: A validation study and comparison of three linking methods. *Psychometrika*, 86, 717-746. <https://doi.org/10.1007/s11336-021-09776-z>
- 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. <https://doi.org/10.1007/s11606-015-3360-0>
- Schulte-van Maaren, Y. W., Carlier, I. V., Zitman, F. G., van Hemert, A. M., de Waal, M. W., van der Does, A. W., van Noorden, M. S., & Giltay, E. J. (2013). Reference values for major depression questionnaires: the Leiden Routine Outcome Monitoring Study. *Journal of Affective Disorders*, 149,

- 342-349. <https://doi.org/10.1016/j.jad.2013.02.009>
- Seashore, H. G. (1955). Methods of expressing Test scores. *Test Service Bulletin*, 48, 7-10. <https://eric.ed.gov/?id=EDo79347>
- Snyder, C. F., Aaronson, N. K., Choucair, A. K., Elliott, T. E., Greenhalgh, J., Halyard, M. Y., Hess, R., Miller, D. M., Reeve, B. B., & Santana, M. (2012). Implementing patient-reported outcomes assessment in clinical practice: A review of the options and considerations. *Quality of Life Research*, 21, 1305-1314. <https://doi.org/10.1007/s11136-011-0054-x>
- Snyder, C., Smith, K., Holzner, B., Rivera, Y. M., Bantug, E., Brundage, M., & PRO Data Presentation Delphi Panel. (2019). Making a picture worth a thousand numbers: Recommendations for graphically displaying patient-reported outcomes data. *Quality of Life Research*, 28, 345-356. <https://doi.org/10.1007/s11136-018-2020-3>
- Sousa, V. D., & Rojjanasirrat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, 17, 268-274. <https://doi.org/10.1111/j.1365-2753.2010.01434.x>
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, 166, 1092-1097. <https://doi.org/10.1001/archinte.166.10.1092>
- Stasinopoulos, M. D., Rigby, R. A., & Bastiani, F. D. (2018). GAMLSS: A distributional regression approach. *Statistical Modelling*, 18, 248-273. <https://doi.org/10.1177/1471082X18759144>
- Teresi, J. A., Ocepek-Welikson, K., Kleinman, M., Eimicke, J. P., Crane, P. K., Jones, R. N., Lai, J.-S., Choi, S. W., Hays, R. D., Reeve, B. B., Reise, S. P., Pilkonis, P. A., & Cella, D. (2009). Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): An item response theory approach. *Psychology Science*, 51, 148-180.
- Terwee, C. B., Crins, M. H. P., Roorda, L. D., Cook, K. F., Cella, D., Smits, N., & Schalet, B. D. (2021). International application of PROMIS computerized adaptive tests: US versus country-specific item parameters can be consequential for individual patient scores. *Journal of Clinical Epidemiology*, 134, 1-13. <https://doi.org/10.1016/j.jclinepi.2021.01.011>
- 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. <https://doi.org/10.1002/cpp.1979>
- Timmerman, M. E., Voncken, L., & Albers, C. J. (2021). A tutorial on regression-based norming of psychological tests with GAMLSS. *Psychological Methods*, 26(3), 357-373. <https://doi.org/10.1037/met0000348>
- Tingey, R., Lambert, M., Burlingame, G., & Hansen, N. (1996). Assessing clinical significance: Proposed extensions to method. *Psychotherapy Research*, 6, 109-123. <https://doi.org/10.1080/10503309612331331638>
- Urbina, S. (2014). *Essentials of psychological testing*. John Wiley & Sons.
- Üstün, T. B., Kosstanjsek, N., Chatterji, S., & Rehm, J. (2010). *Measuring health and disability: Manual for WHO Disability Assessment Schedule WHODAS 2.0*.
- van Muilekom, M. M., Luijten, M. A., van Oers, H. A., Terwee, C. B., van Litsenburg, R. R., Roorda, L. D., Grootenhuis, M. A., & Haverman, L. (2021). From statistics to clinics: The visual feedback of PROMIS® CATs. *Journal of Patient-Reported Outcomes*, 5, 1-14. <https://doi.org/10.1186/s41687-021-00324-y>

- von Glischinski, M., von Brachel, R., & Hirschfeld, G. (2019). How depressed is 'depressed'? A systematic review and diagnostic meta-analysis of optimal cut points for the Beck Depression Inventory revised (BDI-II). *Quality of Life Research*, 28, 1111-1118. <https://doi.org/10.1007/s11136-018-2050-x>
- 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. <https://doi.org/10.1016/j.jclinepi.2013.04.019>
- Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion*, 17, 267. <https://doi.org/10.1037/em00000226>
- Weinfurt, K. P. (2021). Constructing arguments for the interpretation and use of patient-reported outcome measures in research: An application of modern validity theory. *Quality of Life Research*, 30, 1715-1722. <https://doi.org/10.1007/s11136-021-02776-7>
- Weston, J., Dwan, K., Altman, D., Clarke, M., Gamble, C., Schroter, S., Williamson, P., & Kirkham, J. (2016). Feasibility study to examine discrepancy rates in prespecified and reported outcomes in articles submitted to The BMJ. *BMJ Open*, 6, e010075. <https://doi.org/10.1136/bmjopen-2015-010075>
- Williams, G., Flens, G., & de Beurs, E. (2017). Computer adaptief testen: Moderne meettechniek voor uitkomstmaten. *PsyXpert*, 3, 14-21. <https://www.psyxpert.nl/tijdschrift/editie/artikel/t/computergestuurd-adaptief-testen-cat>
- Wolpert, M. (2020). *Funders agree first common metrics for mental health science*. [www.linkedin.com/pulse/funders-agree-first-common-metrics-mentalhealth-science-wolpert](http://www.linkedin.com/pulse/funders-agree-first-common-metrics-mentalhealth-science-wolpert)
- Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, 41, 86-94. [https://doi.org/10.1002/1097-4679\(198501\)41:1<86::AID-JCLP2270410115>3.0.CO;2-W](https://doi.org/10.1002/1097-4679(198501)41:1<86::AID-JCLP2270410115>3.0.CO;2-W)
- Zimmermann, J., Müller, S., Bach, B., Hutsebaut, J., Hummelen, B., & Fischer, F. (2020). A common metric for self-reported severity of personality disorder. *Psychopathology*, 53, 168-178. <https://doi.org/10.1159/000507377>

# *Virtual reality voorafgaand aan exposure bij patiënten met een paniekstoornis en/of sociale-angststoornis*

Een pilotstudie

ANNE SCHIMMEL & JULIE KRANS

## *Samenvatting*

Hoewel virtual reality (VR) een effectief medium voor exposure bij angstbehandelingen lijkt, is er weinig bekend over de inzet van VR als *voorbereiding* op exposure in vivo. In deze pilotstudie werd onderzocht of er aanwijzingen zijn dat VR: (1) de bereidheid tot exposure in vivo vergroot, en (2) de (anticipatie)angst verlaagt. Zes patiënten met een paniekstoornis en/of sociale-angststoornis volgden een kortdurende VR-module voorafgaand aan een behandeling met exposure in vivo. VR lijkt de bereidheid te vergroten en angst te verlagen wanneer een patiënt met een lage bereidheid en/of hoge angst zich blootstelt aan VR-situaties. Omdat er voorafgaand aan de module al een relatief hoge bereidheid bleek te zijn, lijkt het niet nodig deze premodule standaard op te nemen in het behandelaanbod. Deze pilotstudie kan mogelijk vervolgd worden met onderzoek naar de noodzaak van een premodule binnen andere populaties, met mogelijk een enkele VR-sessie die qua techniek en uitvoering van de module verbeterd is.

*Trefwoorden:* VR-exposure, premodule, exposure in vivo, angststoornis, pilotstudie

## **Kernboodschappen voor de klinische praktijk**

- Het lijkt niet nodig om standaard een VR-premodule op te nemen in het behandelaanbod bij angststoornissen, aangezien de bereidheid tot exposure in vivo over het algemeen hoog lijkt in een specialistische ggz-setting.

- ▶ Wanneer de bereidheid tot exposure in vivo laag en/of de angst hoog is, zijn er voorzichtige eerste aanwijzingen dat een VR-premodule de bereidheid kan vergroten en de (anticipatie)angst kan verlagen voor exposure in vivo.
- ▶ Een enkele VR-sessie is mogelijk al voldoende.

## INLEIDING

.....

Exposure met behulp van virtual reality (VR) is de laatste jaren in opkomst als veelbelovend alternatief voor exposure in vivo bij angststoornissen (Maples-Keller et al., 2017). VR wordt gedefinieerd als technologie die een gesimuleerde ervaring creëert, waardoor een gebruiker zich daadwerkelijk aanwezig voelt in een virtuele omgeving, in plaats van in de fysieke ‘echte’ wereld (Riva et al., 2016). VR biedt behandelaars de mogelijkheid om patiënten virtueel bloot te stellen aan angstinducerende situaties. VR-exposure bij angst- en stemmingsklachten blijkt grotere effecten op de klachten te hebben dan een wachtlijstconditie (Fodor et al., 2018). Daarnaast laat een recente meta-analyse op basis van 16 *randomized controlled trials* (RCT's) zien dat de sterkte van de effecten van VR-exposure op angstsymptomen niet verschilt van die van exposure in vivo (van Loenen et al., 2022).

VR-exposure biedt een aantal praktische voordelen ten opzichte van exposure in vivo (Maples-Keller et al., 2017). Ten eerste kunnen patiënten via VR-exposure worden blootgesteld aan omgevingen die in vivo moeilijk beschikbaar zijn. Onderzoek van Pittig en collega's (2019) geeft aanwijzingen dat behandelaars in de ambulante setting praktische barrières ervaren bij het succesvol aanbieden van een behandeling met exposure in vivo. Ten tweede vindt VR-exposure plaats in aanwezigheid van een behandelaar die de patiënt kan coachen voor een optimale leerervaring. Ten derde hebben behandelaars bij VR-exposure meer controle over de stimuli waar de patiënt aan wordt blootgesteld dan bij exposure in vivo. Virtuele omgevingen kunnen worden aangepast aan de angstverwachting en behoeften van de individuele patiënt (Kampmann et al., 2016).

Ondanks de voordelen van VR-exposure vindt er in de klinische behandelpraktijk nog geen massale implementatie plaats (Lindner, 2021). In de klinische behandelpraktijk wordt VR-exposure bij de behandeling van angststoornissen met name ingezet als aanvullend aanbod naast exposure in vivo. Hoewel er geen empirische data over bekend zijn, is het een bekende claim op websites van VR-aanbieders dat VR-exposure in de klinische praktijk kan worden ingezet als een meer laagdrempelige opstap naar exposure in vivo (Moovd, z.d.; Vrendle, z.d.). VR-exposure zou dan van meerwaarde zijn voor patiënten met een angststoornis voor wie de drempel tot exposure in vivo te hoog is. Uit de eerste onderzoeken blijkt inderdaad dat patiënten met een angststoornis VR-exposure beleven als voorspelbaarder, gecontroleerder en veiliger dan exposure in vivo (Bouchard et al., 2017). In een studie van Garcia-

Palacios en collega's (2001) prefereerde 81% tot 89% van de studenten met angstklachten VR-exposure boven exposure in vivo. Ook in een vervolgonderzoek prefereerden 76% van de deelnemers VR-exposure boven exposure in vivo. Het percentage deelnemers dat VR-exposure weigerde (3%) was lager dan het percentage deelnemers dat exposure in vivo weigerde (27%). Van de deelnemers met een voorkeur voor VR-exposure rapporteerde 90% dat er sprake was van te veel angst voor daadwerkelijke blootstelling aan de echte situatie of het echte object (Garcia-Palacios et al., 2007).

Hoewel er geen cijfers bekend zijn over het aantal patiënten dat te angstig is voor exposure in vivo, herkent de cognitief gedragstherapeut in de klinische praktijk dat het niet eenvoudig is om een deel van de patiënten zover te krijgen dat ze zich daadwerkelijk blootstellen aan situaties die zij het meest vrezen. Om: (1) het behandelaanbod voor patiënten die te angstig zijn voor exposure in vivo te verbeteren, en (2) mogelijke aanbevelingen te doen voor de implementatie van VR, is het van meerwaarde om te onderzoeken of de claim van VR-aanbieders klopt: zijn er aanwijzingen dat VR *voorafgaand* aan exposure in vivo de bereidheid tot exposure in vivo kan vergroten en de angst kan verlagen?

Om deze vraag te beantwoorden worden in dit artikel de ervaringen van zes patiënten met een VR-premodule voorafgaand aan een behandeling met exposure in vivo beschreven. In deze pilotstudie ligt de focus specifiek op een VR-module van vier sessies met drukke sociale situaties (busreis, winkelstraat, supermarkt en tuinfeest). Deze situaties zijn bij uitstek geschikt, omdat behandelaren niet in staat zijn ze in een spreekkamer te creëren. Dit vraagt van patiënten dus een hoge mate van bereidheid om deze situaties zelfstandig in vivo aan te gaan. Daarnaast bevatten deze situaties meerdere angstige stimuli die relevant zijn voor verschillende klachten. Onze aanname is dat drukke sociale situaties angstinducerend zullen zijn voor patiënten met een paniekstoornis (PS) en/of sociale-angststoornis (SAS) (APA, 2014). Patiënten met een paniekstoornis en/of sociale-angststoornis in de ambulante setting behoorden dan ook tot de doelgroep van deze pilotstudie.

Onderzocht is of er aanwijzingen zijn dat VR voorafgaand aan een behandeling met exposure in vivo: (1) de *bereidheid* tot exposure in vivo lijkt te vergroten, en (2) de *angst* voor exposure in vivo lijkt te verlagen. Daarnaast is de haalbaarheid van de VR-module onderzocht. Hiertoe werd geëvalueerd of de afzonderlijke VR-sessies in staat zijn om: (3) de geloofwaardigheid van de angstige verwachtingen te beïnvloeden, en (4) angst te activeren. Ook werd onderzocht of: (5) de VR overeenkomt met de situatie die patiënten in het dagelijks leven het meest vermijden. Tot slot werd: (6) het klachtenverloop voorafgaand aan en na afloop van de VR-module in kaart gebracht, en wel om eventuele (ernstige) bijwerkingen in de vorm van een klachtentoeename te monitoren. Hierbij is de toename van zowel angst- als stemmingsklachten gemonitord, omdat onderzoek aantoont dat er sprake is van een hoge mate van comorbiditeit (ter Meulen et al., 2021). Om de onderzoeksvragen

te beantwoorden werd gebruikgemaakt van kwantitatieve data. Vanwege het exploratieve karakter van de pilotstudie en de kleine steekproefomvang ( $N = 6$ ) worden in dit artikel de resultaten per deelnemer visueel weergegeven en beschreven. Middels een semigestructureerd interview werd de beleving van de deelnemers met de VR-module uitgevraagd. Deze kwalitatieve data worden samengevat weergegeven. Er vindt een voorzichtige integratie en interpretatie van de bevindingen plaats in de discussie.

## METHODE

.....

Het onderzoeksprotocol werd goedgekeurd door de medisch-ethische toetsingscommissie CMO regio Arnhem-Nijmegen en is geregistreerd onder referentiecode NL72392.091.19.

### *Design*

.....

Deze pilotstudie betreft een haalbaarheidsonderzoek in de vorm van een casestudie met herhaalde metingen, waarbij zes deelnemers binnen twee weken vier VR-sessies aangeboden krijgen voorafgaand aan een reguliere CGT-behandeling.

### *Deelnemers*

.....

Deelnemers werden geworven op de polikliniek van een Nederlandse instelling voor gespecialiseerde geestelijke gezondheidszorg. Inclusiecriteria voor deelname aan het onderzoek waren: (1) een primaire diagnose van paniekstoornis of sociale-angststoornis volgens de DSM-5, en (2) een leeftijd tussen 18 en 65 jaar. Exclusiecriteria voor deelname aan het onderzoek waren: (1) de aanwezigheid van een ernstige depressieve stoornis die onmiddellijke behandeling behoefde, (2) een psychotische stoornis, (3) middelenmisbruik waarvoor specialistische behandeling vereist was, (4) beperkingen in de verstandelijke vermogens ( $IQ < 70$ ) en/of cognitieve disfuncties op basis van de anamnestiche informatie, en (5) binnen acht weken voorafgaand aan de start van de VR-module ingesteld zijn op (een gewijzigde dosering van) psychofarmaca. De aanwezigheid van andere comorbide stoornissen vormde geen exclusie criterium. Inclusie- en exclusiecriteria werden tijdens de intake vastgesteld met behulp van de Nederlandstalige versie van het Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998), in combinatie met het klinisch oordeel van de intaker of er voldaan werd aan de criteria van een DSM-5-classificatie. Patiënten die binnen de looptijd van het onderzoek voldeden aan de inclusiecriteria en vanaf de wachtlijst na intake konden starten met hun reguliere angstbehandeling werden benaderd om voor aanvang van hun geplande behandeling deel te nemen aan het onderzoek. Deelname was

vrijwillig en had geen gevolgen voor het reguliere behandel aanbod. Gedurende het onderzoek kregen deelnemers geen andere behandeling.

In totaal werden 11 patiënten benaderd voor deelname. Zij ontvingen mondelinge en schriftelijke informatie over het onderzoek. Vijf patiënten zagen af van deelname. Redenen om niet deel te nemen waren: (1) te gespannen zijn voor de reguliere behandeling, waardoor deelname aan de studie als te belastend werd ervaren, (2) te weinig tijd ervaren voor deelname, en (3) dreigende bijwerkingen van VR (migraine en misselijkheid). Twee patiënten gaven geen reden op.

Zes deelnemers werden succesvol geïncludeerd. Er waren geen drop-outs. In totaal namen vier mannen en twee vrouwen deel met een gemiddelde leeftijd van 39,16 jaar (range 28-48). Drie van de deelnemers hadden als primaire diagnose een sociale-angststoornis en de drie andere een paniekstoornis. Vier deelnemers voldeden aan de criteria van ten minste één comorbide stoornis (depressieve stoornis, angststoornis en/of een autismespectrumstoornis). Vier deelnemers hadden eerder een CGT-behandeling gevolgd. Hierbij was niet duidelijk of er binnen die behandeling ook daadwerkelijk exposure in vivo had plaatsgevonden. Een van de vier deelnemers was daarnaast bekend met EMDR-behandeling en schematherapie. Twee deelnemers gebruikten langere tijd (gedurende meer dan 8 weken voorafgaand aan de start van het onderzoek) psychofarmaca (paroxetine, nortriptyline en quetiapine).

### *Materialen*

---

**VR-software** — Voor de VR-module werd het rollenspel uit de VR-applicatie Social Worlds van softwareleverancier CleVR gebruikt. Deze VR-software maakt gebruik van een headset met microfoon (inclusief stemvormer) en twee computers voor de behandelaar. Met een touchscreen bestuurt de behandelaar de virtuele wereld en met een gamingcomputer observeert de behandelaar wat de deelnemer door de VR-bril ziet. Voor de deelnemer bestaat het materiaal uit een VR-bril, hoofdtelefoon en een Microsoft Xbox gamecontroller waarmee de deelnemer zich door de virtuele omgeving begeeft. De VR-apparatuur stond opgesteld in een hiervoor bestemde behandelkamer binnen de onderzoeksinstelling. Specifiek voor dit onderzoek werden vier sociale situaties van de Social Worlds-applicatie gebruikt: een busreis, een winkelstraat, een supermarkt en een tuinfeest.

**Vragenlijsten** — *Mobility Inventory* (MI; Chambless et al., 1985). De MI is een zelfrapportagevragenlijst waarmee agorafobisch vermijdingsgedrag in kaart wordt gebracht. Vermijdingsgedrag wordt op 27 situaties gescoord. Daarnaast kan de respondent eigen situaties toevoegen en scoren. Per situatie wordt onderscheid gemaakt tussen de aanwezigheid of ernst van het

vermijdingsgedrag wanneer de deelnemer alleen is en wanneer hij of zij vergezeld wordt door iemand anders. Vermijdingsgedrag wordt gescoord op een zespunts likertschaal (0 = 'ik vermijd het nooit', 5 = 'ik vermijd het altijd'). De interne consistentie van de twee subschalen (vermijding alleen en vermijding samen) is goed bevonden in de literatuur (Cronbachs  $\alpha$  van respectievelijk 0,95 en 0,96). De test-hertestbetrouwbaarheid van vermijding alleen is  $r = 0,76$  en van vermijding samen is  $r = 0,90$ . Ook de convergente en discriminerende validiteit van de schalen zijn goed bevonden (convergente  $r = 0,63$  met andere meetinstrumenten van agorafobische vermijding; discriminerende  $r = 0,10$ - $0,29$  voor andere angststoornissen (Chambless et al., 2011).

Specifiek voor dit onderzoek werd aan de MI een vraag toegevoegd. Deelnemers dienden vijf situaties uit de MI te selecteren die door de angst en vermijding die ze opriepen hun leven het meest in negatieve zin beïnvloedden. Uit deze top 5 selecteerden de onderzoekers voor iedere deelnemer de persoonlijk meest gevreesde situatie die niet voorkwam in de VR-module (busreis, winkelstraat, supermarkt en tuinfeest). De persoonlijk meest gevreesde situatie werd opgenomen in de door de onderzoekers zelf opgestelde vragenlijsten die gedurende het onderzoek werden afgenomen (zie hierna). Zo werd gemeten in hoeverre de VR-situaties overeenkwamen met de persoonlijk meest gevreesde situatie van de individuele deelnemer. Ook werd hiermee gekeken naar aanwijzingen of er generalisatie optrad van de VR-module naar de persoonlijk meest gevreesde situatie.

*Vragenlijst Bereidheid tot Exposure in Vivo.* Om de bereidheid tot exposure in vivo te meten hebben de onderzoekers zelf een vragenlijst opgesteld. Daarop gaven deelnemers op een 11-punts likertschaal (0 = 'helemaal niet bereid', 10 = 'volledig bereid') aan in hoeverre zij bereid waren zich in vivo bloot te stellen aan de vier VR-situaties (busreis, winkelstraat, supermarkt en tuinfeest) en aan de persoonlijk meest gevreesde situatie.

*5-Situatie Angstlijst (5-SA; Watson & Marks, 1971).* De 5-SA is een zelfrapportagevragenlijst waarbij op vijf zelf in te vullen situaties het angstniveau voor exposure in vivo aan de betreffende situaties in kaart wordt gebracht. Het angstniveau wordt gescoord op een schaal van 0 ('helemaal op mijn gemak') tot 8 ('helemaal in paniek'). Specifiek voor dit onderzoek werden de vijf te scoren situaties vooraf bepaald: busreis, winkelstraat, supermarkt, tuinfeest en de persoonlijk meest gevreesde situatie.

*Vragenlijst Geloofwaardigheid Angstige Verwachting.* Om de geschiktheid van de VR-sessies te meten hebben de onderzoekers zelf een vragenlijst opgesteld. Aan de deelnemers werd gevraagd om voorafgaand aan de VR-sessie een angstige verwachting over de betreffende VR-sessie te rapporteren. De geloofwaardigheid van de angstige verwachting werd door de deelnemer gescoord op een 11-punts likertschaal (0 = 'helemaal niet geloofwaardig', 10 = 'helemaal geloofwaardig').

*Vragenlijst Angstactivatie (SUD-piek):* Aan deelnemers werd gevraagd om het hoogste angstniveau tijdens de VR-sessie uit te drukken op een 11-punts

likertschaal (0 = 'helemaal geen angst', 10 = 'zoveel angst als maar kan'), wat een Subjective Units of Distress (SUD)-score oplevert.

*Vragenlijst Overeenkomst:* Aan deelnemers werd gevraagd om op een 11-punts likertschaal (0 = 'benadert helemaal niet', 10 = 'benadert helemaal') aan te geven in hoeverre de VR-situatie overeenkwam met de persoonlijk meest gevreesde situatie.

*Depression Anxiety Stress Scale* (DASS-21; Lovibond & Lovibond, 1995). De DASS-21 is een zelfrapportagevragenlijst bestaande uit 21 items die de ernst van depressieve symptomen (7 items), angst (7 items) en stressklachten (7 items) over de afgelopen 7 dagen in kaart brengt. Klachten worden op een vierpunts likertschaal (0 = 'helemaal niet of nooit van toepassing', 3 = 'zeer zeker of meestal van toepassing') gescoord. De scores van de items worden bij elkaar opgeteld en verdubbeld voor een totaalscore. Deze score kan variëren van 0 tot 126. De samenhang met andere meetinstrumenten voor angst en depressie is goed. De vragenlijst heeft een goede interne consistentie voor alle subschalen (Cronbachs  $\alpha$  tussen 0,94 en 0,97) en heeft een test-hertestbetrouwbaarheid van respectievelijk  $r = 0,75$ ,  $r = 0,89$  en  $r = 0,79$  voor de subschalen Depressie, Angst en Stress (de Beurs et al., 2001).

*Inventory of Depressive Symptomatology* (IDS-SR; Rush et al., 1996). De IDS-SR is een zelfbeoordelingsvragenlijst waarmee de ernst van depressieve klachten over de afgelopen 7 dagen wordt gemeten. Uitgebreider dan bij de DASS-21 worden de symptoomdomeinen van een depressie inclusief melancholische en atypische symptomen en angst en pijn gemeten. Deze totaalscore kan variëren van 0 tot 84. Zowel de interne consistentie (Cronbachs  $\alpha = 0,85$ ) als de construct-, inhouds- en convergente validiteit zijn goed bevonden in de literatuur (Meesters et al., 2016).

*Semigestructureerd interview.* Om na afloop van de VR-module de beleving van de deelnemers van de VR-module kwalitatief te exploreren, is door de onderzoekers zelf een semigestructureerd interview opgesteld (zie bijlage 1). In het interview worden uitgevraagd: de acceptatie van de VR-module (gebruiksvriendelijkheid, duur van de VR-sessies en totale module), voordelen, verbeterpunten en eventuele negatieve bijwerkingen van de VR-module. De afname van het semigestructureerd interview en het tekstueel uitwerken van de antwoorden werden uitgevoerd door een onafhankelijk onderzoeksassistent.

### *Procedure*

---

**VR-module** — Direct voorafgaand aan hun reguliere behandeling bij de ggz-instelling volgden de deelnemers in twee weken tijd vier geprotocolleerde behandel sessies van 60 minuten (twee sessies per week). Er was geen minimale of maximale tijdspreiding tussen behandel sessies binnen een week. De sessies werden uitgevoerd door behandelaren die VR-getraind waren en

waren opgeleid tot ten minste basispsycholoog. Per sessie stond één virtuele situatie in een vaste volgorde centraal: sessie 1 – busreis; sessie 2 – winkelstraat; sessie 3 – supermarkt; en sessie 4 – tuinfeest. Bij aanvang van de eerste sessie werd de rationale van exposure door de behandelaar uitgelegd. In de volgende sessies werd de rationale bij aanvang nagevraagd en indien nodig herhaald door de behandelaar. Na de rationale werd door de deelnemers de vragenlijst Geloofwaardigheid Angstige Verwachting ingevuld. Op basis van de angstverwachting werden de VR-situaties door de behandelaar op maat gemaakt, zodat de angstverwachting zo goed mogelijk virtueel getoetst kon worden. De behandelaar stelde het aantal aanwezige karakters in, beïnvloedde het uiterlijk en gedrag van de karakters (geslacht, etniciteit, kleding, emoties en bewegingen), en simuleerde via een stemvormer dialogen tussen de deelnemers en fictieve karakters. Deelnemers werden per sessie gemiddeld 45 minuten (drie keer 15 minuten) virtueel blootgesteld aan de op maat gemaakte VR-situatie. Om de kans op *cyber sickness* te verkleinen werd er na iedere 15 minuten blootstelling een pauze ingelast. Tijdens de blootstelling werden deelnemers via een microfoonverbinding gecoacht door de behandelaar in het uitvoeren van de exposure. Meteen nadat de VR-bril was afgezet werden de volgende vragenlijsten in vaste volgorde afgenomen: (1) Angstactivatie (SUD-piek), (2) Geloofwaardigheid Angstige Verwachting, (3) Overeenkomst, en (4) Bereidheid tot Exposure in Vivo.

**Algemeen** — Een week voor de start van de VR-module (week 1) vond een online voormeting plaats via een beeldbelafsprake met een onafhankelijke onderzoeksassistent. In de beeldbelafsprake ontvingen de deelnemers per e-mail een weblink naar online vragenlijsten. Met de eerste vragenlijst werden uitgevraagd: demografische gegevens (geslacht en leeftijd), de primaire diagnose, comorbiditeit, behandelgeschiedenis en psychofarmaca. Vervolgens vulden de deelnemers de MI in. De persoonlijk meest gevreesde situatie uit de MI werd in de beeldbelafsprake met de onderzoeksassistent gedeeld, zodat deze opgenomen kon worden in de vragenlijsten die tijdens het onderzoek werden afgenomen. Vervolgens werden in vaste volgorde de vragenlijsten DASS-21, IDS-SR, 5-SA en de Bereidheid tot Exposure in Vivo door de deelnemers ingevuld. In de daaropvolgende twee weken (week 2 en 3) vond de VR-module plaats. Een week na afronding van de VR-module (week 4) vond de online nameting plaats op dezelfde wijze en met dezelfde vragenlijsten als de voormeting (met uitzondering van de persoonsgegevens en MI). Tot slot nam de onderzoeksassistent het semigestructureerde interview af.

### *Data-analyse*

---

Vanwege het exploratieve karakter van de pilotstudie en de kleine steekproefomvang ( $N = 6$ ) worden de resultaten per deelnemer visueel weergegeven en beschreven. Om de onderzoeksvragen te kunnen beantwoorden

vindt in de discussie een voorzichtige integratie en interpretatie van de bevindingen plaats.

*Bereidheid tot en angst voor exposure in vivo.* Per deelnemer worden: (1) de bereidheid tot, en (2) de angst voor exposure in vivo over de tijd in een lijngrafiek visueel gemaakt. Hierbij worden per deelnemer mogelijke trends van toename of afname over de tijd beschreven.

*Geschiktheid VR-sessies.* Per deelnemer wordt: (1) de geloofwaardigheid van de angstige verwachting voorafgaand en na afloop van de VR-sessie in een lijngrafiek visueel gemaakt. Ook hier wordt per deelnemer een mogelijke trend van toename of afname beschreven. (2) De angstactivatie (SUD-piek) en (3) de overeenkomstsscores staan per deelnemer in een tabel en worden beoordeeld op hoogte van de score (0-3 = laag, 4-6 = gemiddeld en 7-10 = hoog).

*Klachtenverloop (monitoren klachtentoe name).* Per deelnemer worden de totaalscores op de DASS-21 en IDS-SR in een lijngrafiek visueel gemaakt. Hierbij wordt beoordeeld of de totaalscores op de nameting hoger zijn dan op de voormeting.

*Beleving deelnemers.* De beleving van alle deelnemers (N = 6) uit het semi-structureerd interview worden samengevat weergegeven. Overeenkomsten en verschillen in de beleving tussen deelnemers worden uitgewerkt. De antwoorden tijdens de interviews zijn door een onafhankelijk onderzoeks-assistent gerapporteerd en worden door de eerste auteur samengevat.

## RESULTATEN

.....

### *Deelnemer 1*

.....

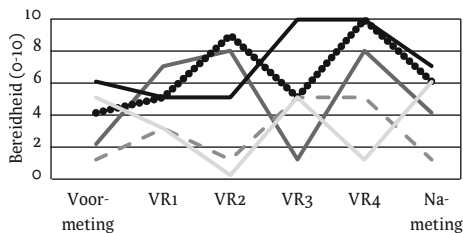
**Bereidheid tot exposure in vivo** — Zoals figuur 1 laat zien, scoorde deelnemer 1 de bereidheid tot exposure in vivo sterk wisselend over de tijd. Er was geen duidelijke trend van toename of afname zichtbaar.

**Angst voor exposure in vivo** — Zoals figuur 2 laat zien, scoorde deelnemer 1 op de angst voor exposure in vivo bij vier van de vijf situaties bij de nameting lager dan bij de voormeting.

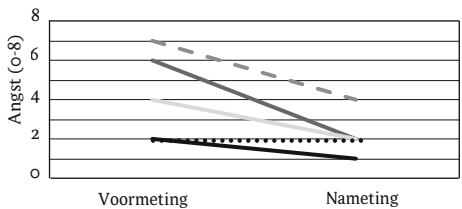
**Geschiktheid VR-sessies** — *Geloofwaardigheid angstige verwachting.* Zoals figuur 3 laat zien, scoorde deelnemer 1 op de geloofwaardigheid van de angstige verwachting bij twee van de vier VR-situaties na afloop van de VR-sessie lager dan voorafgaand aan de VR-sessie. Bij de andere twee situaties was geen verandering.

Deelnemer 1

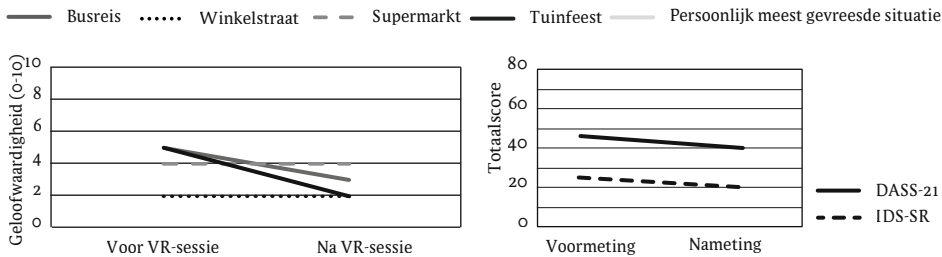
188



FIGUUR 1 BEREIDHEID tot exposure in vivo van deelnemer 1 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting, na de VR-sessies en nameting) en de Y-as de bereidheidsscores (0 tot 10).



FIGUUR 2 Angst voor exposure in vivo van deelnemer 1 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de angstscores (0 tot 8).



FIGUUR 3 Geloofwaardigheid van de angstige verwachting van deelnemer 1 voor en na een VR-sessie per VR-situatie. De X-as geeft de meetmomenten weer (voor en na de VR-sessie) en de Y-as de geloofwaardigheidsscores (0 tot 10) per VR-situatie.

FIGUUR 4 Klachtenverloop deelnemer 1. Totaalscore op de DASS-21 en IDS-SR van deelnemer 1 voor en na de VR-module. De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de totaalscore op de DASS-21 (0-126) en de IDS-SR (0-84).

TABEL 1 Angstactivatie en overeenkomst deelnemer 1

| VR-sessies         | SUD-piek (0-10) | Overeenkomst (0-10) |
|--------------------|-----------------|---------------------|
| VR 1: Busreis      | 10              | 5                   |
| VR 2: Winkelstraat | 9               | 2                   |
| VR 3: Supermarkt   | 10              | 9                   |
| VR 4: Tuinfeest    | 10              | 8                   |

Noot. SUD = Subjective Units of Distress

*Angstactivatie (SUD-piek) en overeenkomst.* Zoals tabel 1 laat zien, scoorde deelnemer 1 op de SUD-piek voor alle situaties hoog. De overeenkomst tussen de VR-situaties en de persoonlijk meest gevreesde situatie was vooral zichtbaar voor supermarkt en tuinkeest.

**Klachtenverloop (monitoren klachtentoe name)** — Zoals figuur 4 laat zien, waren de totaalscores op de voormeting 46 (DASS-21) en 25 (IDS-SR), en op de nameting 40 (DASS-21) en 21 (IDS-SR). Er was geen duidelijke trend van klachtentoe name zichtbaar.

#### *Deelnemer 2*

---

**Bereidheid tot exposure in vivo** — Zoals figuur 5 laat zien, scoorde deelnemer 2 op de bereidheid tot exposure in vivo na de eerste VR-sessie bij één situatie hoger (busreis) en bij drie situaties lager (tuinkeest, winkelstraat, persoonlijk meest gevreesde situatie) dan bij de voormeting. De score op supermarkt bleef onveranderd hoog over de tijd. De score op busreis bleef over de tijd hoger en de persoonlijk meest gevreesde situatie lager dan bij de voormeting. De score op tuinkeest en winkelstraat wisselden over de tijd.

**Angst voor exposure in vivo** — Zoals figuur 6 laat zien, scoorde deelnemer 2 op de angst voor exposure in vivo bij twee situaties bij de nameting hoger dan bij de voormeting. De score op één situatie was bij de nameting lager dan bij de voormeting. Bij twee situaties was geen verandering zichtbaar. Er was geen duidelijke trend van toename of afname zichtbaar.

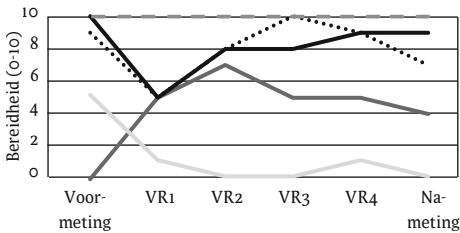
**Geschiktheid VR-sessies** — *Geloofwaardigheid angstige verwachting.* Zoals figuur 7 laat zien, scoorde deelnemer 2 op de geloofwaardigheid van de angstige verwachting bij drie van de vier VR-situaties na afloop van de VR-sessie lager dan voorafgaand aan de VR-sessie. Bij één situatie was geen verandering.

*Angstactivatie (SUD-piek) en overeenkomst.* Zoals tabel 2 laat zien, scoorde deelnemer 2 op de SUD-piek voor twee situaties gemiddeld en twee situaties laag. De overeenkomst tussen de VR-situaties en de persoonlijk meest gevreesde situatie was voor geen enkele situatie hoog.

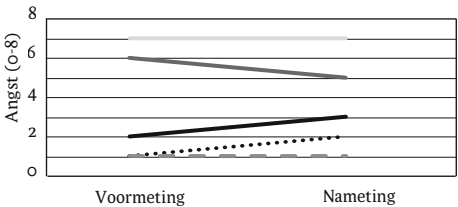
**Klachtenverloop (monitoren klachtentoe name)** — Zoals figuur 8 laat zien, waren de totaalscores op de voormeting 74 (DASS-21) en 33 (IDS-SR), en op de nameting 28 (DASS-21) en 13 (IDS-SR). Er was geen duidelijke trend van klachtentoe name zichtbaar.

190

Deelnemer 2

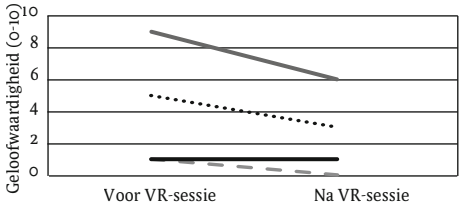


FIGUUR 5 *Bereidheid tot exposure in vivo van deelnemer 2 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting, na de VR-sessies en nameting) en de Y-as de bereidheidsscores (0 tot 10).*

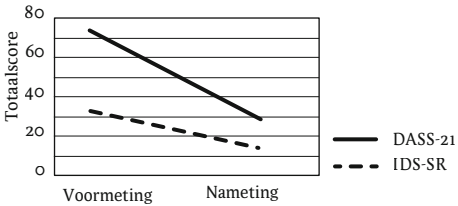


FIGUUR 6 *Angst voor exposure in vivo van deelnemer 2 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de angstscores (0 tot 8).*

— Busreis ..... Winkelstraat - - Supermarkt — Tuinfeest — Persoonlijk meest gevreesde situatie



FIGUUR 7 *Geloofwaardigheid van de angstige verwachting van deelnemer 2 voor en na een VR-sessie per VR-situatie. De X-as geeft de meetmomenten weer (voor en na de VR-sessie) en de Y-as de geloofwaardigheidsscores (0 tot 10) per VR-situatie.*



FIGUUR 8 *Klachtenverloop deelnemer 2. Totaalscore op de DASS-21 en IDS-SR van deelnemer 2 voor en na de VR-module. De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de totaalscore op de DASS-21 (0-126) en de IDS-SR (0-84).*

TABEL2 *Angstactivatie en overeenkomst deelnemer 2*

| VR-sessies         | SUD-piek (0-10) | Overeenkomst (0-10) |
|--------------------|-----------------|---------------------|
| VR 1: Busreis      | 6               | 5                   |
| VR 2: Winkelstraat | 6               | 5                   |
| VR 3: Supermarkt   | 2               | 4                   |
| VR 4: Tuinfeest    | 2               | 4                   |

Noot. SUD = Subjective Units of Distress

*Deelnemer 3*

191

**Bereidheid tot exposure in vivo** — Zoals figuur 9 laat zien, scoorde deelnemer 3 op de bereidheid tot exposure in vivo na de eerste VR-sessie bij drie situaties (busreis, tuinfeest, persoonlijk meest gevreesde situatie) hoger dan bij de voormeting. Er was een trend van toename zichtbaar. Bij twee situaties (winkelstraat en supermarkt) bleven de scores over de tijd onveranderd hoog.

**Angst voor exposure in vivo** — Zoals figuur 10 laat zien, scoorde deelnemer 3 op de angst voor exposure in vivo bij alle situaties bij de nameting lager dan bij de voormeting. Er was een trend van afname zichtbaar.

**Geschiktheid VR-sessies** — *Geloofwaardigheid angstige verwachting.* Zoals figuur 11 laat zien, scoorde deelnemer 3 op de geloofwaardigheid van de angstige verwachting bij alle VR-situaties na afloop van de VR-sessie lager dan voorafgaand aan de VR-sessie. Er was een trend van afname zichtbaar.

*Angstactivatie (SUD-piek) en overeenkomst.* Zoals tabel 3 laat zien, scoorde deelnemer 3 op de SUD-piek voor alle situaties hoog. De overeenkomst tussen de VR-situaties en de persoonlijk meest gevreesde situatie ontbreekt vanwege een fout in de dataverzameling.

**Klachtenverloop (monitoren klachtentoe name)** — Zoals figuur 12 laat zien, waren de totaalscores op de voormeting 76 (DASS-21) en 32 (IDS-SR), en op de nameting 50 (DASS-21) en 26 (IDS-SR). Er was geen duidelijke trend van klachtentoe name zichtbaar.

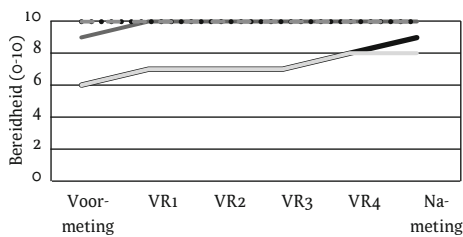
*Deelnemer 4*

**Bereidheid tot exposure in vivo** — Zoals figuur 13 laat zien, scoorde deelnemer 4 op de bereidheid tot exposure in vivo na de eerste VR-sessie bij één situatie (persoonlijk meest gevreesde situatie) hoger dan bij de voormeting. Deze score bleef over de tijd hoger dan bij de voormeting. Er was een trend van toename zichtbaar. Bij twee situaties (tuinfeest en busreis) was geen duidelijke trend van toename of afname zichtbaar. Bij twee situaties (winkelstraat en supermarkt) bleven de scores over de tijd onveranderd hoog.

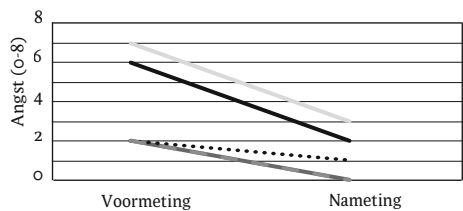
**Angst voor exposure in vivo** — Zoals figuur 14 laat zien, scoorde deelnemer 4 op de angst voor exposure in vivo bij vier van de vijf situaties bij de nameting hoger dan bij de voormeting. Bij één situatie was geen verandering.

Deelnemer 3

192

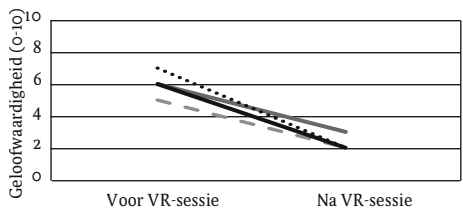


FIGUUR 9 *Bereidheid tot exposure in vivo van deelnemer 3 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting, na de VR-sessies en nameting) en de Y-as de bereidheidsscores (0 tot 10).*

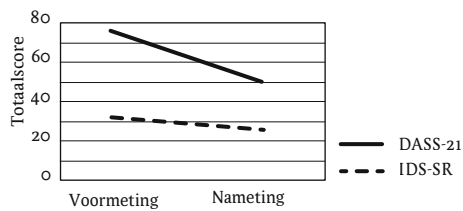


FIGUUR 10 *Angst voor exposure in vivo van deelnemer 3 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de angstscores (0 tot 8).*

— Busreis ..... Winkelstraat - - Supermarkt — Tuinfeest — Persoonlijk meest gevreesde situatie



FIGUUR 11 *Geloofwaardigheid van de angstige verwachting van deelnemer 3 voor en na een VR-sessie per VR-situatie. De X-as geeft de meetmomenten weer (voor en na de VR-sessie) en de Y-as de geloofwaardigheidsscores (0 tot 10) per VR-situatie.*



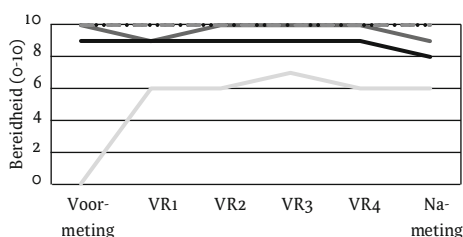
FIGUUR 12 *Klachtenverloop deelnemer 3. Totaalscore op de DASS-21 en IDS-SR van deelnemer 3 voor en na de VR-module. De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de totaalscore op de DASS-21 (0-126) en de IDS-SR (0-84).*

TABEL 3 *Angstactivatie en overeenkomst deelnemer 3*

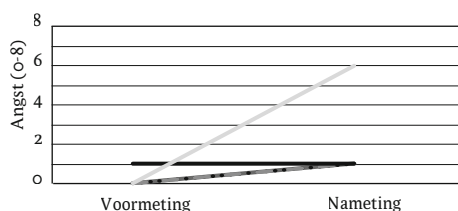
| VR-sessies         | SUD-piek (0-10) | Overeenkomst (0-10) |
|--------------------|-----------------|---------------------|
| VR 1: Busreis      | 9               | Onbekend            |
| VR 2: Winkelstraat | 9               | Onbekend            |
| VR 3: Supermarkt   | 9               | Onbekend            |
| VR 4: Tuinfeest    | 9               | Onbekend            |

Noot. SUD = Subjective Units of Distress

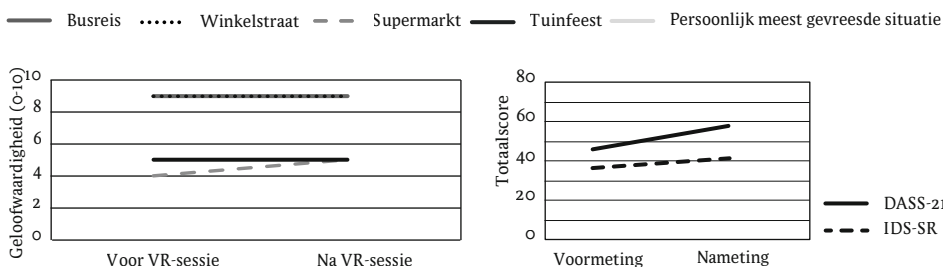
# Deelnemer 4



FIGUUR 13 Bereidheid tot exposure in vivo van deelnemer 4 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting, na de VR-sessies en nameting) en de Y-as de bereidheidsscores (0 tot 10).



FIGUUR 14 Angst voor exposure in vivo van deelnemer 4 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de angstscores (0 tot 8).



FIGUUR 15 Geloofwaardigheid van de angstige verwachting van deelnemer 4 voor en na een VR-sessie per VR-situatie. De X-as geeft de meetmomenten weer (voor en na de VR-sessie) en de Y-as de geloofwaardigheidsscores (0 tot 10) per VR-situatie.

FIGUUR 16 Klachtenverloop deelnemer 4. Totaalscore op de DASS-21 en IDS-SR van deelnemer 4 voor en na de VR-module. De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de totaalscore op de DASS-21 (0-126) en de IDS-SR (0-84).

TABEL 4 Angstactivatie en overeenkomst deelnemer 4

| VR-sessies         | SUD-piek (0-10) | Overeenkomst (0-10) |
|--------------------|-----------------|---------------------|
| VR 1: Busreis      | 1               | 1                   |
| VR 2: Winkelstraat | 2               | 2                   |
| VR 3: Supermarkt   | 5               | 4                   |
| VR 4: Tuinfeest    | 6               | 7                   |

Noot. SUD = Subjective Units of Distress

**Geschiktheid VR-sessies** — *Geloofwaardigheid angstige verwachting.* Zoals figuur 15 laat zien, scoorde deelnemer 4 op de geloofwaardigheid van de angstige verwachting bij één situatie (supermarkt) na afloop van de VR-sessie hoger dan voorafgaand aan de VR-sessie. Bij de andere drie situaties was geen verandering.

*Angstactivatie (SUD-piek) en overeenkomst.* Zoals tabel 4 laat zien, scoorde deelnemer 4 op de SUD-piek voor twee situaties gemiddeld en twee situaties laag. De overeenkomst tussen de VR-situaties en de persoonlijk meest gevreesde situatie was vooral zichtbaar voor tuinkeest.

**Klachtenverloop (monitoren klachtentoe name)** — Zoals figuur 16 laat zien, waren de totaalscores op de voormeting 46 (DASS-21) en 36 (IDS-SR), en op de nameting 58 (DASS-21) en 41 (IDS-SR). Er was een trend van klachtentoe name zichtbaar.

#### *Deelnemer 5*

---

**Bereidheid tot exposure in vivo** — Zoals figuur 17 laat zien, is er een trend van toename en vervolgens afname van bereidheid tot exposure in vivo zichtbaar bij deelnemer 5 over de tijd. De bereidheid is bij alle vijf situaties bij de nameting lager dan bij VR-sessie 4.

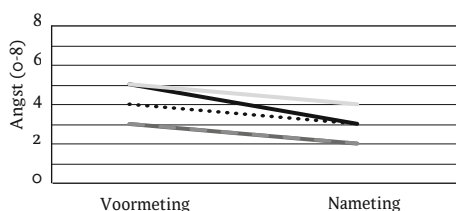
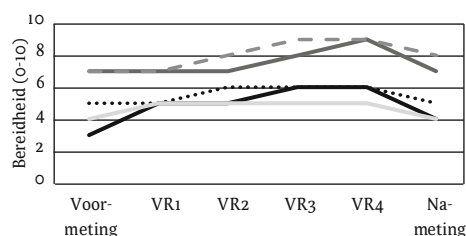
**Angst voor exposure in vivo** — Zoals figuur 18 laat zien, scoorde deelnemer 5 op de angst voor exposure in vivo bij alle situaties bij de nameting lager dan bij de voormeting. Er was een trend van afname zichtbaar.

**Geschiktheid VR-sessies** — *Geloofwaardigheid angstige verwachting.* Zoals figuur 19 laat zien, scoorde deelnemer 5 op de geloofwaardigheid van de angstige verwachting bij drie situaties na afloop van de VR-sessie lager dan voorafgaand aan de VR-sessie. Er was een trend van toename zichtbaar. Bij één situatie (busreis) was geen verandering.

*Angstactivatie (SUD-piek) en overeenkomst.* Zoals tabel 5 laat zien, scoorde deelnemer 5 op de SUD-piek hoger voor de eerste twee VR-situaties. De overeenkomst tussen de VR-situaties en de persoonlijk meest gevreesde situatie was voor geen enkele situatie hoog.

**Klachtenverloop (monitoren klachtentoe name)** — Zoals figuur 20 laat zien, waren de totaalscores op de voormeting 50 (DASS-21) en 33 (IDS-SR), en op de nameting 44 (DASS-21) en 35 (IDS-SR). Er was geen duidelijk trend van klachtentoe name zichtbaar.

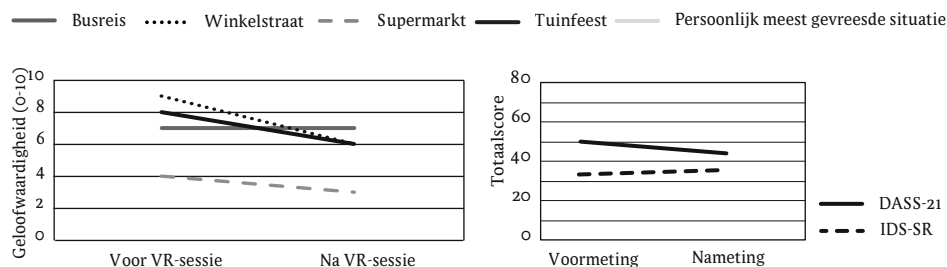
# Deelnemer 5



195

FIGUUR 17 Bereidheid tot exposure in vivo van deelnemer 5 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting, na de VR-sessies en nameting) en de Y-as de bereidheidsscores (0 tot 10).

FIGUUR 18 Angst voor exposure in vivo van deelnemer 5 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de angstscores (0 tot 8).



FIGUUR 19 Geloofwaardigheid van de angstige verwachting van deelnemer 5 voor en na een VR-sessie per VR-situatie. De X-as geeft de meetmomenten weer (voor en na de VR-sessie) en de Y-as de geloofwaardigheidsscores (0 tot 10) per VR-situatie.

FIGUUR 20 Klachtenverloop deelnemer 5. Totaalscore op de DASS-21 en IDS-SR van deelnemer 5 voor en na de VR-module. De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de totaalscore op de DASS-21 (0-126) en de IDS-SR (0-84).

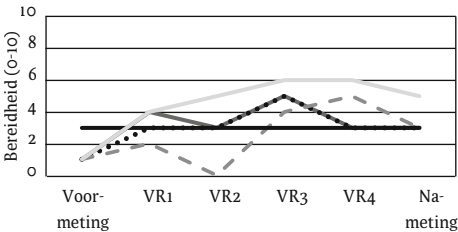
TABEL 5 Angstactivatie en overeenkomst deelnemer 5

| VR-sessies         | SUD-piek (0-10) | Overeenkomst (0-10) |
|--------------------|-----------------|---------------------|
| VR 1: Busreis      | 7               | 4                   |
| VR 2: Winkelstraat | 7               | 4                   |
| VR 3: Supermarkt   | 4               | 4                   |
| VR 4: Tuinfeest    | 5               | 3                   |

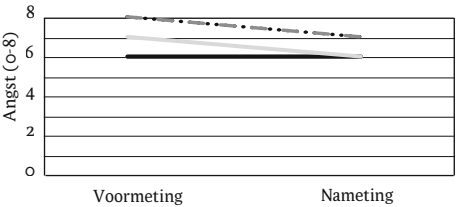
Noot. SUD = Subjective Units of Distress

Deelnemer 6

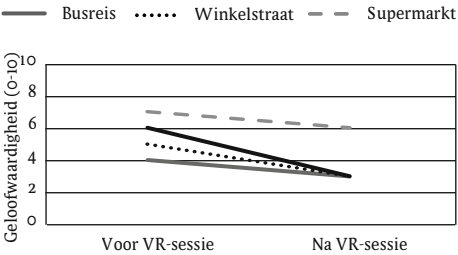
196



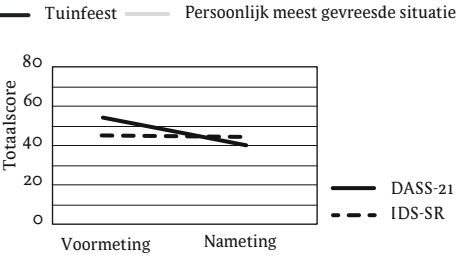
FIGUUR 21 *Bereidheid tot exposure in vivo van deelnemer 6 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting, na de VR-sessies en nameting) en de Y-as de bereidheidsscores (0 tot 10).*



FIGUUR 22 *Angst voor exposure in vivo van deelnemer 6 aan de vijf situaties (vier VR-situaties en de persoonlijk meest gevreesde situatie). De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de angstscores (0 tot 8).*



FIGUUR 23 *Geloofwaardigheid van de angstige verwachting van deelnemer 6 voor en na een VR-sessie per VR-situatie. De X-as geeft de meetmomenten weer (voor en na de VR-sessie) en de Y-as de geloofwaardigheidsscores (0 tot 10) per VR-situatie.*



FIGUUR 24 *Klachtenverloop deelnemer 6. Totaalscore op de DASS-21 en IDS-SR van deelnemer 6 voor en na de VR-module. De X-as geeft de meetmomenten weer (voormeting en nameting) en de Y-as de totaalscore op de DASS-21 (0-126) en de IDS-SR (0-84).*

TABEL 6 *Angstactivatie en overeenkomst deelnemer 6*

| VR-sessies         | SUD-piek (0-10) | Overeenkomst (0-10) |
|--------------------|-----------------|---------------------|
| VR 1: Busreis      | 7               | 4                   |
| VR 2: Winkelstraat | 8,5             | 7                   |
| VR 3: Supermarkt   | 8               | 6                   |
| VR 4: Tuinfeest    | 7               | 4                   |

Noot. SUD = Subjective Units of Distress

*Deelnemer 6*

197

**Bereidheid tot exposure in vivo** — Zoals figuur 21 laat zien, scoorde deelnemer 6 de bereidheid tot exposure in vivo na de eerste VR-sessie bij vier situaties hoger dan bij de voormeting. De bereidheid van deze vier VR-situaties wisselde vervolgens over de tijd. Wel bleef de bereidheid bij de vier situaties hoger dan bij de voormeting. Bij één situatie (tuinfeest) was geen verandering over de tijd.

**Angst voor exposure in vivo** — Zoals figuur 22 laat zien, scoorde deelnemer 6 op de angst voor exposure in vivo bij drie situaties bij de nameting lager dan bij de voormeting. Er was een trend van afname zichtbaar. Bij twee situaties was geen verandering.

**Geschiktheid VR-sessies** — *Geloofwaardigheid angstige verwachting.* Zoals figuur 23 laat zien, scoorde deelnemer 6 op de geloofwaardigheid van de angstige verwachting bij alle situaties na afloop van de VR-sessie lager dan voorafgaand aan de VR-sessie. Er was een trend van afname zichtbaar.

*Angstactivatie (SUD-piek) en overeenkomst.* Zoals tabel 6 laat zien, scoorde deelnemer 6 op de SUD-piek het hoogst op winkelstraat en supermarkt. De overeenkomst tussen deze twee VR-situaties en de persoonlijk meest gevreesde situatie was ook het meest zichtbaar.

**Klachtenverloop (monitoren klachtentoeename)** — Zoals figuur 24 laat zien, waren de totaalscores op de voormeting 54 (DASS-21) en 45 (IDS-SR), en op de nameting 40 (DASS-21) en 44 (IDS-SR). Er was geen duidelijk trend van klachtentoeename zichtbaar.

*Kwalitatieve evaluatie van de VR-module door de deelnemers  
(semigestructureerd interview)*

Vier van de zes deelnemers rapporteerden dat de blootstelling aan de onderzoekafpraak en het dragen van de VR-bril al angstinducerend waren. Drie deelnemers noemden als voordeel van de VR-module dat zij in een angst-inducerende situatie terechtkwamen in aanwezigheid van de behandelaar. Door de feedback van de behandelaar werden zij zich meer bewust van hun gevoelens en gedrag (deelnemers 4, 5 en 6). Alle deelnemers rapporteerden dat een sessieduur van een uur passend was. Twee deelnemers gaven aan dat meerdere pauzes (waarbij de VR-bril werd afgezet) binnen een uur nodig waren (deelnemers 5 en 6). Volgens alle deelnemers was de drempel tot exposure in vivo voor enkele VR-situaties enigszins verlaagd na de VR. Vier van de zes deelnemers rapporteerden dat blootstelling aan VR-situaties die overeenkwamen met gevreesde situaties in het dagelijks leven het meest drempelverlagend waren (deelnemers 1, 2, 3 en 5).

Als verbeterpunten rapporteerden alle deelnemers dat: (1) niet alle VR-situaties voor hen voldoende angst opriepen, en (2) zij niet overeenkwamen met de persoonlijk meest gevreesde situatie. Hierdoor voelde de VR-module niet als volledig passend. Vijf deelnemers vonden de VR-situaties te rustig (deelnemers 1, 2, 3, 4 en 5). Vier van de zes deelnemers vonden het uiterlijk en de stem van de karakters in de VR-situaties onrealistisch (deelnemers 2, 4, 5 en 6). Twee deelnemers misten een VR-situatie waarin zij werden blootgesteld aan autorijden (deelnemers 1 en 2). Twee deelnemers rapporteerden dat er verbetering nodig was in hoe de behandelaar de VR-module uitvoerde. Als voorbeeld noemde een deelnemer dat de behandelaar fouten maakte met de stemvormer, waardoor een karakter gedurende één VR-sessie verschillende stemmen had (deelnemer 5). Deelnemer 3 noemde als voorbeeld dat zijn virtuele lengte tijdens een VR-sessie verkeerd was ingesteld, waardoor de verhoudingen niet klopte. Deze voorbeelden maakten de virtuele ervaring minder realistisch en minder angstinducerend. Als bijwerkingen tijdens de VR-sessies werden evenwichtsproblemen (deelnemers 3, 5 en 6) en misselijkheid (deelnemers 3 en 4) gerapporteerd. Na de VR-sessies werd hoofdpijn gerapporteerd (deelnemers 3, 4, en 6).

#### DISCUSSIE

.....

In deze pilotstudie werden de ervaringen van zes patiënten met een paniekstoornis en/of sociale-angststoornis met een kortdurende VR-module voorafgaand aan een behandeling met exposure in vivo onderzocht. Het primaire doel van deze pilotstudie was om te evalueren of er aanwijzingen zijn dat VR voorafgaand aan een behandeling met exposure in vivo: (1) de bereidheid tot exposure in vivo vergroot, en (2) de angst voor exposure in vivo verlaagt.

Bij vier van de zes deelnemers werd na één VR-sessie hogere bereidheid tot exposure in vivo gezien bij het merendeel van de situaties waartoe de bereidheid bij de voormeting relatief laag was. Deze resultaten geven dus voorzichtige aanwijzingen dat het aanbieden van VR-exposure voorafgaand aan exposure in vivo de *bereidheid* tot exposure in vivo kan vergroten voor situaties waartoe de bereidheid relatief laag is. Er bleek na één VR-sessie geen duidelijke trend van toename in bereidheid tijdens het verdere verloop van de VR-module. Het uitblijven van een duidelijke trend van toename in bereidheid na de eerste sessie is een mogelijke aanwijzing dat één VR-sessie voorafgaand aan een behandeling met exposure in vivo afdoende zou kunnen zijn om de bereidheid tot exposure in vivo te vergroten.

Overeenkomstige bevindingen werden zichtbaar met betrekking tot de *angst* voor exposure in vivo. Ook werd er bij vier van de zes deelnemers lagere angst voor exposure in vivo gezien bij het merendeel van de situaties waartoe de angst bij de voormeting relatief hoog was. De resultaten zouden er dus op kunnen wijzen dat het aanbieden van VR-exposure voorafgaand

aan exposure in vivo van meerwaarde is bij situaties die in vivo een relatief hoge mate van angst oproepen. De resultaten geven een eerste voorlopige indruk dat VR-exposure deze relatief hoge angstniveaus kan verlagen. Bovenstaande bevindingen lijken in lijn met eerdere meta-analyses die aantonen dat VR-exposure positieve veranderingen op angstsymptomen kan bewerkstelligen (van Loenen et al., 2022). Deze pilotstudie geeft voorzichtige aanwijzingen dat VR de bereidheid vergroot en de angst verlaagt wanneer per individu de juiste VR-situatie geselecteerd wordt: een situatie met een lage bereidheid en/of hoge angst.

Daarnaast werd in deze pilotstudie de geschiktheid van de VR-sessies geëvalueerd. Onderzocht werd in hoeverre er aanwijzingen zijn dat de VR-sessies: (1) de geloofwaardigheid van de angstige verwachtingen verlagen, (2) angst activeren en (3) overeenkomen met de persoonlijk meest gevreesde situatie. Bij vier van de zes deelnemers was de geloofwaardigheid van de angstige verwachting bij het merendeel van de VR-sessies na afloop lager dan voorafgaand aan de sessie.

Deze bevinding geeft voorzichtige aanwijzingen dat de sessies in de VR-module geschikt lijken te zijn om de betekenis van angstinducerende situaties te veranderen. Net als bij exposure in vivo zou disconfirmatie tijdens de VR-module dus samen kunnen gaan met angstreductie (Vervliet et al., 2014). Daarnaast scoorden drie van de zes deelnemers bij alle VR-sessies relatief hoge angstscores (score 6 of hoger). Daarentegen rapporteerden alle deelnemers bij de kwalitatieve evaluatie van de VR-module dat niet alle VR-sessies voldoende angst opriepen. Deze bevinding sluit aan bij de scores van de deelnemers met betrekking tot de mate waarin de VR-sessies overeenkwamen met de persoonlijk meest gevreesde situatie. Alle deelnemers gaven aan dat het merendeel van de VR-sessies in relatief laag tot gemiddelde mate (score 5 of lager) overeenkwam met de persoonlijk meest gevreesde situatie. Het overgrote deel van de deelnemers (vijf van de zes) vond de VR-sessies te rustig. Vier van de zes deelnemers vond het uiterlijk en de stem van de karakters onrealistisch. Bovendien leek volgens twee deelnemers verbetering nodig in de uitvoering van de VR-sessies door de behandelaar om de VR-sessies realistisch en angstinducerend te maken. Net als bij exposure in vivo (Pittig et al., 2019) lijken er voor de behandelaar dus ook bij VR praktische barrières te zijn om een succesvolle exposurebehandeling aan te kunnen bieden. Al met al geven deze resultaten aan dat er verbetering nodig lijkt te zijn in de VR-sessies om voldoende angst te activeren en ze overeen te laten komen met de persoonlijk meest gevreesde situatie.

Tot slot werd het klachtenverloop (angst- en stemmingsklachten) geëvalueerd om eventuele (ernstige) bijwerkingen in de vorm van een klachten-toename te monitoren. Angst- en stemmingsklachten werden gemeten met de DASS-21 en IDS-SR. Bij vijf van de zes deelnemers was er geen duidelijke trend van klachtentoe name op de DASS-21 en IDS-SR zichtbaar. Bij één deelnemer (deelnemer 4) was wel een trend van klachtentoe name zichtbaar.

Het is onduidelijk welke patiëntkenmerken hierbij een rol hebben kunnen spelen. Daarnaast werden door vier van de zes deelnemers relatief milde bijwerkingen gedurende de VR-module gerapporteerd. Deze bestonden uit: evenwichtsproblemen (drie deelnemers), misselijkheid (twee deelnemers) en hoofdpijn (drie deelnemers). Deze relatief milde bijwerkingen zijn herkenbaar bij het gebruik van VR (Tian et al., 2022). De resultaten geven dus eerste aanwijzingen dat de VR-module voor het merendeel van de deelnemers relatief veilig lijkt te zijn. Hierbij dient opgemerkt te worden dat er wel individuele verschillen lijken te bestaan. Het lijkt belangrijk om gebruikers van de VR-module te informeren over mogelijke relatief milde bijwerkingen, omdat deze bij het merendeel van de deelnemers voor leken te komen.

Een opvallende bevinding was dat bij bijna alle deelnemers (vijf van de zes) de bereidheid tot exposure in vivo aan het begin van de VR-module al relatief hoog was voor het merendeel van de VR-situaties. Bij meerdere deelnemers was zelfs de bereidheid voor de persoonlijk meest gevreesde situatie al relatief hoog. Een mogelijke verklaring is dat patiënten die op een wachtlijst staan voor een behandeling met exposure in vivo al bereid zijn om deze behandeling aan te gaan. Hierop voortbordurend viel op dat vijf van de elf benaderde patiënten afzagen van deelname aan de studie. Verkregen redenen voor niet deelnemen waren: (1) te gespannen zijn voor de reguliere behandeling, waardoor deelname aan de studie als te belastend werd ervaren; (2) te weinig tijd ervaren voor deelname; en (3) de dreigende bijwerkingen van VR (migraine en misselijkheid). Dit was een opvallende bevinding en lijkt niet in lijn met het onderzoek van Garcia-Palacios en collega's (2007), waarbij deelnemers VR prefereerden boven exposure in vivo. Onze bevinding zou kunnen betekenen dat patiënten de VR-module mogelijk niet specifiek beoordelen als laagdrempeliger dan exposure in vivo. Daarnaast scoorde de helft van de deelnemers de angst voor exposure in vivo al relatief laag voor het merendeel van de VR-situaties. Een mogelijke verklaring is dat de situaties uit de VR-module onvoldoende angst induceren bij de doelgroep van deze pilotstudie (en mogelijk dus onvoldoende passend zijn). Het was opvallend dat de helft van de deelnemers aangaf dat de angst voor exposure in vivo voor de persoonlijk meest gevreesde situatie lager was dan een van de VR-situaties. Het is verleidelijk om over mogelijke verklaringen te speculeren, maar een degelijke onderbouwing daarvoor ontbreekt.

Gezien de bevindingen van dit onderzoek rijst de vraag of een VR-premodule een zinnige aanvulling is op de standaardzorg voor patiënten met een angststoornis die op de wachtlijst staan voor een behandeling met exposure in vivo in de specialistische ggz. De deelnemers aan dit onderzoek hadden voorafgaand aan de VR-module al grotendeels een relatief hoge bereidheid tot exposure in vivo. De VR-situaties induceerden ook geen hoge angst (SUD-piek). Mogelijk is de VR-premodule minder geschikt voor de doelgroep van deze pilotstudie. Wanneer de bereidheid vooraf echter relatief laag was en/of de anticipatieangst voor exposure in vivo relatief hoog was, leek een en-

kele VR-sessie (maar niet meer) wel behulpzaam te kunnen zijn. Verder onderzoek zou zich kunnen richten op de effecten van een enkele VR-sessie op de bereidheid en anticipatieangst specifiek bij patiënten die op de wachtlijst staan en die zich minder bereid voelen tot exposure in vivo en/of zeer veel anticipatieangst ervaren. Dit onderzoek zou bij voorkeur een meer gecontroleerd design moeten hebben (bijvoorbeeld random toewijzing aan de VR-sessies versus een controlesessie, of een *multiple baseline design*).

De bevindingen van dit onderzoek dienen echter met voorzichtigheid geïnterpreteerd te worden. Het doel was om een eerste haalbaarheidsstudie uit te voeren via een pilotstudie. Passend bij dit doel is gekozen voor een *single case design*. De keuze voor dit design maakt het echter niet mogelijk om causale conclusies te trekken of bevindingen te generaliseren naar de totale populatie van de pilotstudie. De resultaten geven dan ook slechts een indicatie voor de haalbaarheid van de module en de vruchtbaarheid voor verder onderzoek. Het onderzoekskader maakte dat de behandelaren deels minder vrijheid hadden in het vormgeven van de VR-exposuresessies dan in de klinische praktijk het geval zou zijn. Het is mogelijk dat een meer gepersonaliseerde en technisch correcte uitvoering grotere effecten heeft op de bereidheid en (anticipatie)angst voor exposure in vivo. Een verdere beperking van het onderzoek is dat maar zes van de elf benaderde patiënten hebben deelgenomen aan de pilotstudie. Hierdoor is er mogelijk sprake geweest van selectiebias, waarbij patiënten die deelnamen aan de studie de VR-module eventueel als minder belastend zouden kunnen ervaren, een hogere of juist lagere lijdensdruk hebben, of meer bereidheid voor exposure in het algemeen ervaren dan patiënten die afzagen van deelname. De uitvoering van de VR was vervolgens niet geheel foutloos. Niet alle VR-sessies bleken achteraf door de behandelaren volledig volgens het draaiboek te zijn uitgevoerd. Zo bleek dat niet bij elke VR-sessie de juiste VR-situatie uit het draaiboek was ingeladen. Ook waren niet alle vragenlijsten goed afgenomen. Bij één deelnemer is een vragenlijst niet afgenomen (deelnemer 3) en bij een andere deelnemer is onduidelijk of de behandelaar bij alle meetmomenten dezelfde persoonlijk meest gevreesde situatie in de vragenlijsten heeft genoteerd (deelnemer 2). Enkele behandelaren ervoeren, ondanks adequate training in VR, moeilijkheden in de uitvoering van de VR-sessies. De door de deelnemers gerapporteerde problemen met de stemvervorming en virtuele lengte maakten dat zij de virtuele ervaring als minder realistisch en angstinducerend beoordeelden. Deze laatste twee beperkingen onderschrijven het belang van goede training van behandelaren in onderzoeksvaardigheden en in het geven van VR, en van constante monitoring daarvan gedurende praktijkonderzoek. Hoewel het niet ging om onbegrijpelijke of geheel onnavolgbare fouten, kunnen zij bij een kleine steekproef een relatief groot effect op de resultaten hebben. De bevindingen geven hierdoor mogelijk een conservatiever beeld dan het maximale potentieel van een dergelijke module.

Vervolgonderzoek zou zich kunnen richten op de vraag welke doelgroep mogelijk meer baat heeft bij een premodule voorafgaand aan exposure in vivo. Gezien de relatief hoge bereidheid en lage angst voor exposure in vivo bij de doelgroep van deze pilotstudie, zou de VR-module mogelijk relevanter zijn voor patiënten met een angststoornis die minder bereid zijn tot en banger voor exposure in vivo. Hierbij kan gedacht worden aan patiënten met een angststoornis die nog niet op een wachtlijst durven te staan voor een behandeling met exposure in vivo.

De resultaten van onze pilotstudie zouden erop kunnen wijzen dat in vervolgonderzoek één VR-sessie voorafgaand aan de behandeling met exposure in vivo mogelijk voldoende is. Tot slot lijkt het nodig om de VR-techniek te verbeteren en behandelaren te monitoren in het consciëntieus uitvoeren van onderzoeksgerelateerde handelingen.

### *Klinische implicaties*

---

Concluderend geeft deze pilotstudie voorzichtige eerste aanwijzingen dat VR voorafgaand aan een behandeling met exposure in vivo de bereidheid tot exposure in vivo zou kunnen vergroten en de angst zou kunnen verlagen in individuele gevallen, mits een situatie gebruikt wordt met een relatief lage bereidheid tot en/of hoge angst voor exposure in vivo. De VR-module lijkt relatief veilig aangeboden te kunnen worden. Een enkele VR-sessie is mogelijk voldoende. Daarnaast is het de vraag in hoeverre patiënten als groep – in dit geval patiënten met een angststoornis die op de wachtlijst staan voor exposure in vivo in de specialistische ggz – behoefte hebben aan een dergelijke premodule. Er leek relatief hogere bereidheid te zijn al voor de premodule werd afgenomen. Het lijkt op het eerste oog niet evident om deze VR-premodule standaard op te nemen in het aanbod. Deze pilot onderschrijft daarnaast het belang van goede training van behandelaren in onderzoeksvaardigheden en het geven van VR, en constante monitoring daarvan. Onderzoek naar de noodzaak van een dergelijke module binnen andere populaties, met mogelijk één enkele VR-sessie met verbeteringen aan de techniek en uitvoering van de module, zijn mogelijke wegen voor vervolgonderzoek.

BIJLAGE 1 SEMIGESTRUCTUREERD INTERVIEW

.....

**Cliëntnummer:**

203

**Datum:**

*Vragen*

.....

**Module**

- ▶ Hoe hebt u VR als module ervaren?
- ▶ Hoeveel sessies bent u geweest?
- ▶ Hoe ervoer u het aantal sessies? Was het aantal sessies te lang of juist te kort?
- ▶ Hoe ervoer u de duur van de sessies?
- ▶ Hoe vond u de omgevingen en de virtuele personen?
- ▶ Wat ziet u als sterke punten van de VR?
- ▶ Wat zijn verbeterpunten?
- ▶ In hoeverre zou u de module VR aanraden aan een ander met soortgelijke angstklachten?
- ▶ Hebt u een bepaalde VR-omgeving gemist? Of had u een VR-omgeving anders gewild?

**Proces**

- ▶ Bent u van tevoren goed ingelicht over de module?
- ▶ Hebt u negatieve bijwerkingen ervaren (zoals misselijkheid, hoofdpijn, duizeligheid) tijdens de VR-module?
- ▶ Voelt de VR-module als passend bij uw klachten en hulpvraag?
- ▶ Heeft VR voor u de drempel tot het doorbreken van vermijdingsgedrag verlaagd? Of juist niet?

**Anne Schimmel** is werkzaam bij Kairos Forensische Zorg, Pompestichting, Arnhem.

**Julie Krans** is verbonden aan Pro Persona Overwaal, Expertisecentrum voor angst, dwang en PTSS, te Nijmegen; bij het Behavioural Science Institute, Radboud Universiteit, Nijmegen; en bij de afdeling Klinische Psychologie, KU Leuven, Leuven.

*Correspondentieadres:* Anne Schimmel, Kairos Arnhem, Velperweg 26, 6824 BJ Arnhem. E-mail: a.schimmel@pompestichting.nl

**Summary** *VR prior to exposure in vivo for patients with panic and/or social anxiety disorder: A pilot study*

Although VR appears to be an effective medium for exposure in anxiety treatments, little is known about the use of VR as preparation for exposure in vivo. In this pilot study it was investigated whether there are indications that VR:

(1) increases the willingness for exposure in vivo, and (2) decreases the (anticipatory) fear of exposure in vivo. Six patients with panic disorder and/or social anxiety disorder followed a brief VR module prior to an exposure in vivo treatment. VR appears to increase the willingness and decrease fear when patients expose themselves to VR situations of low willingness and/or high fear. There appeared to be a relatively high willingness prior to the VR module. It seems unnecessary to include this VR module in the standard treatment plan. Investigating the need for a premodule in other clinical populations, possibly involving a single VR session with technical and procedural improvements, are suggestions for further research.

**Keywords** VR exposure, premodule, exposure in vivo, anxiety disorder, pilot study

## Referenties

- American Psychiatric Association (APA). (2014). *Handboek voor de classificatie van psychische stoornissen (DSM-5). Nederlandse vertaling van Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*. American Psychiatric Association.
- Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, E., Forget, H., Loranger, C., & Roucaut, F. X. (2017). Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: A three-arm randomised controlled trial. *British Journal of Psychiatry*, 210, 276-283. <https://doi.org/10.1192/bjp.bp.116.184234>
- Chambless, D. L., Caputo, G. C., Jasin, S. E., Gracely, E. J., & Williams, C. (1985). The Mobility Inventory for Agoraphobia. *Behaviour Research and Therapy*, 23, 35-44. [https://doi.org/10.1016/0005-7967\(85\)90140-8](https://doi.org/10.1016/0005-7967(85)90140-8)
- Chambless, D. L., Sharpless, B. A., Rodriguez, D., McCarthy, K. S., Milrod, B. L., Khalsa, S. R., & Barber, J. P. (2011). Psychometric properties of the Mobility Inventory for Agoraphobia: Convergent, discriminant, and criterion-related validity. *Behavior Therapy*, 42, 689-699. <https://doi.org/10.1016/j.beth.2011.03.001>
- de Beurs, E., van Dyck, R., Marquenie, L. A., Lange, A., & Blonk, R. W. (2001). De DASS: Een vragenlijst voor het meten van depressie, angst en stress. *Gedragstherapie*, 34, 35-54.
- Fodor, L. A., Coteș, C. D., Cuijpers, P., Szamosközi, S., David, D., & Cristea, I. A. (2018). The effectiveness of virtual reality based interventions for symptoms of anxiety and depression: A meta-analysis. *Scientific Reports*, 8. <https://doi.org/10.1038/s41598-018-28113-6>
- Garcia-Palacios, A., Botella, C., Hoffman, H., & Fabregat, S. (2007). Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by patients with specific phobias. *Cyberpsychology & Behavior*, 10, 722-724. <https://doi.org/10.1089/cpb.2007.9962>
- Garcia-Palacios, A., Hoffman, H. G., See, S. K., Tsai, A., & Botella, C. (2001). Redefining therapeutic success with virtual reality exposure therapy. *Cyberpsychology & Behavior*, 4, 341-348. <https://doi.org/10.1089/109493101300210231>
- Kampmann, I. L., Emmelkamp, P. M. G., Hartanto, D., Brinkman, W. P., Zijlstra, B. J. H., & Morina, N. (2016). Exposure to virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial. *Behaviour Research and Therapy*, 77, 147-156. <https://doi.org/10.1016/j.brat.2015.12.016>
- Lindner, P. (2021). Better, virtually: The past, present, and future of virtual reality cognitive behavior therapy.

- International Journal of Cognitive Therapy*, 14, 23-46. <https://doi.org/10.1007/s41811-020-00090-7>
- Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research & Therapy*, 33, 335-343. [https://doi.org/10.1016/0005-7967\(94\)00075-u](https://doi.org/10.1016/0005-7967(94)00075-u)
- Maples-Keller, J. L., Bunnell, B. E., Kim, S. J., & Rothbaum, B. O. (2017). The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. *Harvard Review of Psychiatry*, 25, 103. <https://doi.org/10.1097/hrp.0000000000000138>
- Meesters, Y., Duijzer, W. B., Nolen, W. A., Schoevers, R. A., & Ruhé, H. G. (2016). Inventory of Depressive Symptomatology en verkorte versie in routine outcome monitoring van Stichting Benchmark GGZ. *Tijdschrift voor Psychiatrie*, 58, 48-54.
- Moovd. (z.d.). *Virtual reality exposure therapie*. <https://moovd.nl/methodes/virtual-reality-exposure-therapie>
- Pittig, A., Kotter, R., & Hoyer, J. (2019). The struggle of behavioral therapists with exposure: Self-reported practicability, negative beliefs, and therapist distress about exposure-based interventions. *Behavior Therapy*, 50, 353-366. <https://doi.org/10.1016/j.beth.2018.07.003>
- Riva, G., Banos, R. M., Botella, C., Mantovani, F., & Gaggioli, A. (2016). Transforming experience: The potential of augmented reality and virtual reality for enhancing personal and clinical change. *Frontiers in Psychiatry*, 7, 164. <https://doi.org/10.3389/fpsy.2016.00164>
- Rush, A. J., Gullion, C. M., Baco, M. R., Jarrett R. B., & Trivedi, M. H. (1996). The Inventory of Depressive Symptomatology (IDS): Psychometric properties. *Psychological Medicine*, 26, 477-86. <https://doi.org/10.1017/s0033291700035558>
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, 59, 22-33.
- ter Meulen, W. G., Draisma, S., van Hemert, A. M., Schoevers, R. A., Kupka, R. W., Beekman, A. T., & Penninx, B. W. (2021). Depressive and anxiety disorders in concert – A synthesis of findings on comorbidity in the NESDA study. *Journal of Affective Disorders*, 284, 85-97. <https://doi.org/10.1016/j.jad.2021.02.004>
- Tian, N., Lopes, P., & Boulic, R. (2022). A review of cybersickness in head-mounted displays: Raising attention to individual susceptibility. *Virtual Reality*, 1-33. <https://doi.org/10.1007/s10055-022-00638-2>
- van Loenen, I., Scholten, W., Muntingh, A., Smit, J., & Batelaan, N. (2022). The effectiveness of virtual reality exposure-based cognitive behavioral therapy for severe anxiety disorders, obsessive compulsive disorder, and posttraumatic stress disorder: Meta-analysis. *Journal of Medical Internet Research*, 24, e26736. <https://doi.org/10.2196/26736>
- Vervliet, B., Depreeuw, B., Trensou, C., Conway, T., Zbozinek, M., & Craske, G. (2014). Exposuretherapie maximaliseren: Een benadering volgens het inhibitorisch-leermodel. *Gedrags-therapie*, 47, 296-339.
- Vrendle. (z.d.). *Virtual reality (VR) in de zorg met Vrendle*. <https://vrendle.nl/vr-in-de-zorg>
- Watson, J. P., & Marks, I. M. (1971). Relevant and irrelevant fear in flooding – A crossover study of phobic patients. *Behavior Therapy*, 2, 275-293. [https://doi.org/10.1016/s0005-7894\(71\)80062-x](https://doi.org/10.1016/s0005-7894(71)80062-x)

.....

## *Voor welke problemen bieden regelgeleid gedrag en de functionele analyse 2.0 eigenlijk oplossingen?*

---

Een reactie op De Houwer & Boddez  
(2023)

---

ERIK TEN BROEKE & KEES KORRELBOOM

### *Samenvatting*

.....

In december 2023 verscheen in *Gedragstherapie* een themanummer over het werk van B.F. Skinner. Deze reactie betreft het artikel van De Houwer en Boddez (2023) in dat nummer. Meer in het bijzonder gaat het daarbij om wat deze auteurs schrijven over de door hen gesignaleerde praktijk bij het gebruik van functi(onel)e analyses in de gedragstherapie. Na een selectieve bloemlezing uit het artikel van De Houwer en Boddez geven wij onze kritische reflecties op hun beweringen en stellingen. De conclusie luidt dat er sprake is van een oplossing voor een in onze ogen niet bestaand probleem. *Trefwoorden: Skinner, functionele analyse 2.0, functieanalyse, relational frame theory, regelgeleid gedrag, betekenisanalyse*

### Kernboodschappen voor de klinische praktijk

- ▶ (Operant) gedrag staat niet alleen onder consequente controle van *feitelijke* bekrachtigers (Sr) maar vooral van *mentale representaties* van die bekrachtigers (Sr).
- ▶ In de analyse van zulke mentale representaties moet de idiosyncratische betekenisverlening van de patiënt aan die bekrachtigers worden meegewogen.
- ▶ In de theorie en praktijk van de Nederlandstalige gedragstherapie worden voor zulke analyses al geruime tijd sjablonen gebruikt die zijn gebaseerd op het paradigma van de klassieke conditionering.
- ▶ Het is niet zonder meer helder of het begrip ‘regelgeleid gedrag’ wezenlijk andere of zelfs betere handvatten biedt voor dergelijke analyses dan deze ‘klassieke sjablonen’.

### INLEIDING

.....

In december vorig jaar verscheen in *Gedragstherapie* onder redactie van Arnold van Emmerik en Pier Prins een interessant themanummer over de betekenis van B.F. Skinner voor de samenleving in het algemeen en de wetenschappelijk psychologie in het bijzonder. In het verlengde daarvan wordt hier en daar gewag gemaakt van zijn relevantie voor de (ontwikkeling van) gedragstherapie. Ondanks het begrijpelijke enthousiasme voor het themanummer, spreekt de hoofdredacteur in het redactioneel haar teleurstelling uit dat er geen vrouwelijke auteurs zijn betrokken en moeten Van Emmerik en Prins – zo te merken deze keer tot hún teleurstelling – erkennen dat het hun niet is gelukt ‘empirische bijdragen’ op te nemen. Voor klinici als wij zijn, is het daarop aansluitend jammer dat ook niet meer plaats kon worden toebedeeld aan de betekenis van het skinneriaanse gedachtegoed voor de behandelpraktijk. Die behandelpraktijk komt nog het meest aan bod in de bijdrage van De Houwer en Boddez (2023). Juist om die reden en omdat beide auteurs gerenommeerde wetenschappers zijn, trok hun artikel extra onze aandacht. Al aandachtig lezend riep het artikel echter een aantal vragen bij ons op. Wij denken dat het de moeite waard is om deze vragen en bijbehorende overwegingen te delen met de lezers van *Gedragstherapie*. In aanloop daarnaartoe, eerst een selectieve bloemlezing uit De Houwer en Boddez.

### OVER SKINNER, REGELGELEID GEDRAG

.....

Na een theoretische inleiding over bepaalde aspecten van het denken van Skinner, gaan De Houwer en Boddez onder andere in op hun zorgen over de ‘traditionele functionele analyse’, dat wil zeggen op hoe volgens hen de praktijk is van gedragstherapeuten die functionele (of wat ons betreft functie)-

analyses (FA's) opstellen van de problematiek van hun patiënten. Volgens de auteurs zou die praktijk zich ten onrechte beperken tot het beschrijven van 'hoe gedrag een functie is van contingenties in de omgeving' (p. 402). Van daaruit – en bij wijze van boodschap voor de klinische praktijk – wordt betoogd dat het beter is om zich bij het opstellen van functi(onel)e analyses niet te beperken tot die 'contingenties in de omgeving, maar om ook rekening te houden met regelgeleid gedrag' (p. 403). En passant wordt beweerd dat daarmee ook wordt voorkomen dat 'gedrag gezien wordt als het resultaat van mentale representaties en informatieverwerking' (p. 403). Die zin bevreedt ons enigszins. Kennelijk vinden de auteurs het belangrijk om dat te voorkómen en achten zij het beter om de nadruk te leggen op een 'functi-onele benadering', die zowel de inhoud als het proces van cognitief functioneren opvat als gedrag: 'Het cruciale verschil tussen cognitieve therapie en een functionele benadering die rekening houdt met regelgeleid gedrag zit echter in hoe men kijkt naar overtuigingen en denkprocessen [...]. Binnen een functionele benadering worden die beschouwd als vormen van gedrag, terwijl ze vanuit een cognitief perspectief verwijzen naar mentale representaties en informatieverwerking' (p. 407). Anders dan de gedragstherapeut die zich bedient van de 'traditionele functi(onel)e analyse', zo wordt gesteld, maakt de 'functionele therapeut' een FA die 'wordt uitgevoerd op basis van de subjectieve in plaats van de feitelijke leefwereld van de patiënt' (p. 413). Hoewel wij niet zomaar weten wat een 'functionele therapeut' precies is en of wij ons daar zelf toe moeten rekenen, voelen wij ons – zoals verderop wordt uitgewerkt – aangesproken en gesteund door deze stellingname.

### *Relational frame theory*

---

De Houwer en Boddez betrekken in hun betoog ook een pleidooi voor de algemene relevantie van de *relational frame theory* en de mogelijkheid die deze theorie biedt om aspecten van Skinners ideeën die nog onvoldoende zijn uitgewerkt van een nader kader te voorzien. Daarbij ligt er betrekkelijk veel nadruk op de zorgen van de auteurs: (1) dat gedragstherapeuten Skinners werk vaak verkeerd interpreteren, (2) dat in de klinische praktijk daardoor verkeerde of in ieder geval te beperkte functieanalyses worden opgesteld, om vervolgens vast te stellen dat (3) het daarmee afgelopen moet zijn: 'Dit ondermijnt de basis van traditionele functionele analyses van menselijk gedrag, inclusief gedrag dat we observeren in de klinische praktijk. Om het op een niet-technische, cognitieve manier te zeggen: het gedrag van mensen is meestal niet gebaseerd op wat er feitelijk *gebeurt* in hun omgeving, maar op wat ze *geloven* dat er gebeurt in hun omgeving. Het is vanuit dit perspectief niet meer aangewezen om traditionele functionele analyses op te stellen waarin gedrag enkel begrepen wordt vanuit feitelijke contingenties in de omgeving' (p. 406). De auteurs laten het niet bij deze vaststelling en geven even verderop aan hoe 'vanaf nu' een FA zou moeten worden opgesteld: 'In

een moderne (2.0) functionele analyse probeert de therapeut niet enkel de feitelijke leefwereld van de patiënt in kaart te brengen, maar vooral de subjectieve leefwereld zoals de patiënt die construeert. De therapeut gaat op zoek naar wat de gepercipieerde gevolgen zijn van gedrag in een bepaalde context (subjectieve Sd:R-Sr-contingenties), hoe de patiënt betekenis verleent aan gebeurtenissen (hoe situaties, prikkels, acties, gedachten, gevoelens, waarden, enzovoort, aan elkaar gerelateerd worden) en hoe dit alles gedrag bepaalt' (p. 412). Of De Houwer en Boddez onze werkwijze bij het opstellen van FA's zien als '1.0' weten we niet helemaal zeker, maar kennelijk kijken zij in ieder geval totaal anders aan tegen de ontwikkeling die wij zien, waarin verreweg het grootste deel van de praktiserende gedragstherapeuten (in ieder geval die binnen het Nederlands taalgebied) al vele jaren FA's opstellen vanuit het subjectieve persepectief. De moderne (2.0) FA die zij lijken voor te staan is dus geen toekomst, maar een reeds lang bestaand heden. Aanleiding voor nadere reflectie.

#### EEN AANTAL OPVALLENDE OPMERKINGEN

.....

Min of meer terzijde worden in de loop van het artikel nog een aantal min of meer aan het voorgaande gerelateerde beweringen gedaan die ons intrigeerden: 'De kernvraag voor een klinisch psycholoog die problemen wil oplossen is dan ook: "Wat werkt er?" en niet zozeer: "Waarom werkt het?"' (p. 411) en 'Als fysisch wezen kan een klinisch psycholoog echter enkel rechtstreeks ingrijpen in de fysische wereld. Mentale processen (dat wil zeggen: informatieverwerking) vallen hier buiten' (p. 411). Dit zijn beweringen die raken aan de (beroeps)identiteit van 'klinisch psychologen' in het algemeen en (cognitieve) gedragstherapeuten in het bijzonder. Nog een reden voor reflectie.

#### REFLECTIES BIJ DE BEWERINGEN VAN DE HOUWER EN BODDEZ

.....

De wetenschappelijke psychologie beoogt inzicht te geven in hoe mensen functioneren. Daartoe zijn legio theorieën en modellen naar voren gebracht, onderzocht, verworpen, bijgesteld, bevestigd en met elkaar verbonden. Soms is zelfs sprake van wat Borsboom (2023) noemt een 'groot verhaal', zoals in het geval van het werk van Skinner. Het is – zoals het themanummer beoogt – alleszins de moeite waard dat verhaal opnieuw te vertellen, van enige exegese te voorzien en – tenminste in de bijdrage van De Houwer en Boddez – te verbinden met de (klinische) praktijk van de gedragstherapie. Hun 'Kernboodschappen voor de klinische praktijk' (p. 402) passen expliciet bij het laatste. Daarin wordt immers met zoveel woorden gezegd dat 'traditionele functionele analyses' zich ten onrechte beperken tot 'contingenties in de omgeving'; daadwerkelijk aanwezige, concreet waarneembare stimuli,

derhalve. Met name die laatste kernboodschap voor de klinische praktijk is onderwerp van onze reflecties. Naar ons idee wordt gedrag in laatste instantie altijd aangestuurd door de subjectieve leefwereld van de patiënt. Dit is bijvoorbeeld ook het uitgangspunt in twee veelgebruikte leerboeken (Korrelboom & ten Broeke, 2014, hoofdstuk 9; ten Broeke et al., 2022, hoofdstuk 1).

Wij menen dat het primaat van die ‘subjectieve leefwereld’ als ultieme bepaler van menselijk gedrag een ruime empirische ondersteuning kent (zie onder meer de literatuur over US-inflatie en deflatie en US-herevaluatie: Davey, 2006; Hosoba et al., 2001). Bovendien zijn wij bepaald niet de enigen die in ‘onze’ analyseschema’s in belangrijke mate uitgaan van de ‘subjectieve leefwereld’ van de patiënt. Dat geldt bijvoorbeeld ook voor een ander veelgebruikt Nederlandstalig leerboek (Hermans et al., 2018). Tezamen maken deze leerboeken al vele jaren deel uit van veel opleidingen in de gedragstherapie in Nederland. Dat heeft er mede toe geleid dat het overgrote deel van de praktiserende gedragstherapeuten binnen het Nederlands taalgebied al sinds jaar en dag in hun functieanalyses uitgaan van de mentale representatie van de Sr en niet van de feitelijke Sr. Ook in de internationale literatuur is de subjectieve leefwereld van de patiënt al lang een belangrijk onderdeel van de casusconceptualisatie. Soms wordt daarbij de functieanalyse zelf deels ‘cognitief ingevuld’, in andere gevallen wordt een ‘cognitief-gedragsmatige aanvulling’ op die functieanalyse noodzakelijk geacht om tot een adequate casusconceptualisatie te kunnen komen (Bond, 1998; Bruch, 1998).

De bewering van De Houwer en Boddez dat in de huidige gedragstherapeutische functieanalyses ten onrechte te zeer wordt uitgegaan van de feitelijke leefwereld van de patiënt en onvoldoende van de subjectieve beleving daarvan, staat wat ons betreft kortom op flink gespannen voet met de ter zake relevante literatuur en met de feitelijke behandelpraktijk zoals wij die denken te kennen. Om met het eerste te beginnen: wij komen dat niet tegen (Burger, 1994; Hermans et al., 2018; Korrelboom & ten Broeke, 2014; Orlemans, 1988). Al deze auteurs zoeken in hun analyses naar en interveniëren op geleide van de persoonlijke *betekenissen* die patiënten aan specifieke concrete (antecedente en consequente) stimuli en gedragingen geven. Dat gaat in ieder geval al terug op de tweefactorentheorie, waarin klassieke en operante conditionering op een geïntegreerde en onderling samenhangende wijze werden toegepast (Mowrer, 1947). Sindsdien is die invalshoek ook in de behandelpraktijk alleen maar prominenter geworden; geen (cognitieve) gedragstherapeut kan zich volgens ons nog voorstellen dat het zinvol zou kunnen zijn om nadere ‘analyses van betekenisverlening’ uit de casusconceptualisatie en de analyseschema’s die daaraan ten grondslag liggen opzettelijk achterwege te laten. De bewering dat traditionele FA’s ten onrechte worden beperkt tot contingenties tussen feitelijke, concrete, externe stimuli en gedrag (zie p. 412, box 3) is naar onze mening dan ook een even boude als onterechte bewering.

Om de (vaak ondergeschikte) rol van *daadwerkelijke, fysieke Sr's* in relatie tot gedrag van nog wat meer context te voorzien is het goed zich te realiseren dat er bij het ontstaan en voortduren van probleemgedrag (lang) niet altijd sprake is van een directe 'harde' conditionering. Zo is bij het dwanggedrag (R) van een patiënt met OCS die meent dat met dat dwanggedrag het sterven van zijn kinderen wordt afgewend (~S<sup>1</sup>) hopelijk nimmer concreet sprake geweest van het sterven van zijn kinderen. De Sr wordt in de FA dan ook uitdrukkelijk beschreven als een *representatie* van de vermeende consequenties van gedrag.

Onze conclusie is dan ook dat de 2.o functi(onel)e analyse die door De Houwer en Boddez naar voren wordt geschoven al sinds jaar dag bestaat. Het voorstel om bij een FA op zoek te gaan 'naar wat de gepercipieerde gevolgen zijn van gedrag in een bepaalde context (subjectieve Sd:R-Sr-contingenties), hoe de patiënt betekenis verleent aan gebeurtenissen [...] en hoe dit alles gedrag bepaalt' (p. 412, box 3) had dan ook zomaar een citaat uit ons eigen handboek *Geïntegreerde cognitieve gedragstherapie* kunnen zijn geweest. Het even verderop aangegeven 'cruciale' verschil met de 'traditionele analyse' is naar onze mening dan ook juist een mooi voorbeeld van een onderscheid zonder een verschil. Hetgeen volgens ons ook geldt voor de stevige bewering van De Houwer en Boddez dat vaak ten onrechte of zelfs onwenselijk gebruik wordt gemaakt van 'cognitieve' begrippen als 'mentale representaties' (Arntz, 2020) en beter gesproken kan worden van 'regelgeleid' gedrag. We nemen dat in dit kader maar even voor kennisgeving aan en vragen ons (opnieuw) af welke problemen worden opgelost door het begrip 'cognitie' te vervangen door 'regel'. Ter illustratie: in zijn hoofdstuk over *case formulation* bij de behandeling van een patiënt met een gegeneraliseerde angststoornis stelt Bond (1998, p. 84): 'Cognitive-behavioural therapists who do not subscribe to radical behaviourism can, throughout this chapter, substitute the word "belief" for the word "rule"; because for the purpose of this chapter, both terms are interchangeable.'

#### BEWERINGEN OVER DE KLINISCHE PRAKTIJK

.....

De Houwer en Boddez doen in hun artikel een poging het inhoudelijke be-  
toog te illustreren met wat ons betreft nogal 'stroeve' klinische voorbeelden.  
Dat weerhoudt de auteurs er echter niet van om zonder omhaal te beweren  
dat het voor een behandelaar niet uitmaakt of zijn analyse klopt of niet, als  
het maar werkt. Wij onderschrijven deze stellingname niet. Natuurlijk, als  
het werkt kan worden overgegaan tot de orde van de dag; maar als dat niet zo

1 ~S- is het symbool waarmee Korrelboom & ten Broeke (2014) aangeven dat de  
persoon meent dat een negatieve uitkomst (S-) die normaal gesproken wel zou  
worden verwacht door het uitvoeren van het (dwang)gedrag, uitblijft.

is, dan dient de therapeut de vraag te stellen en te beantwoorden wat al dan niet klopt aan de opgestelde analyses. Wij plaatsen ook serieuze vraagtekens bij de bewering dat door therapeuten enkel rechtstreeks kan worden ingegrepen in de 'fysische wereld' en (dus) niet in mentale processen of informatieverwerking. Als zij daarin gelijk zouden hebben zou dat een treurige boodschap zijn voor al die therapeuten (en hun patiënten) die imaginaire rescripting, EMDR-therapie, imaginaire exposure, COMET enzovoort toepassen, in de verwachting daarmee de veronderstelde mentale representaties van (de effecten van) leerervaringen te corrigeren (Arntz, 2020).

#### RESUMÉ

.....

Door het themanummer worden misverstanden met betrekking tot het werk van Skinner signaleerd, uit de weg geruimd en wordt benadrukt dat Skinner het destijds allemaal behoorlijk goed zag. De ambitie van De Houwer en Boddez om in het kader daarvan het gedachtegoed van Skinner te verhelderen en te relateren aan de hedendaagse therapeutische praktijk valt zonder meer te waarden; iedere gedragstherapeut die gebruikmaakt van FA's is immers schatplichtig aan diens 'grote verhaal'. Daar waar het werk van Skinner toch tekortschiet wordt een beroep gedaan op de RFT en – in het verlengde daarvan – de *acceptance and commitment therapy* (ACT). Dit alles creëert een kader waarbinnen de auteurs kritische noten kraken met betrekking tot de huidige praktijk van 'traditionele FA's', zoals zij die zien. Op basis van de inzichten van Skinner en de RFT lijken zij een verbeterde werkwijze naar voren te willen brengen. Hoewel niet in detail uitgewerkt lijkt die werkwijze procedureel niet op gespannen voet te staan met wat veel gedragstherapeuten, onder wie wijzelf, al sinds jaar en dag gebruikten bij het opstellen van een FA. Op zichzelf is dat goed nieuws, want er is in zeker opzicht op dat punt consensus, maar het roept wel de vraag op voor welk probleem het op deze manier opnieuw onder de aandacht brengen van het gedachtegoed van Skinner en de daarop gebaseerde RFT een oplossing is. Het eerst signaleren van een niet-bestaand probleem (de huidige cognitieve gedragstherapie hanteert onvoldoende een subjectief perspectief), en daar dan een oplossing voor aandragen (we moeten het als regelgeleid gedrag gaan beschouwen) is immers niet de meest nuttige besteding van tijd en moeite.

**Erik ten Broeke** is gedeeltelijk gepensioneerd klinisch psycholoog en gedragstherapeut te Deventer/Barcelona.

**Kees Korrelboom** is gepensioneerd klinisch psycholoog en gedragstherapeut.

*Correspondentieadres:* Erik ten Broeke. E-mail: [ebroeke@planet.nl](mailto:ebroeke@planet.nl)

**Summary** *For which problems do rule-guided behavior and functional analysis 2.0 actually offer solutions?*

In December 2023, a special issue on the work of B.F. Skinner was published in *Gedragstherapie*. This response concerns the article by De Houwer and Boddez (2023) in that issue. More specifically, it addresses the assertions made by these authors regarding the practice they observed in the use of functional analyses in behavior therapy. After a selective excerpt from the article by De Houwer and Boddez, we provide our critical reflections on their claims and statements. The conclusion is that there is a proposed solution for what we consider a non-existent problem.

**Keywords** *Skinner, functional analysis 2.0, functional analysis, relational frame theory, rule governed behavior, analysis of meaning*

## Referenties

- Arntz, A. (2020). A plea for more attention to mental representations. *Journal of Behavior Therapy and Experimental Psychiatry*, 67, 1-11. <https://doi.org/10.1016/j.jbtep.2019.101510>
- Bond, F. W. (1998). Using a case formulation to understand and treat a person with generalized anxiety disorder. In: M. Bruch & F.W. Bond (eds.), *Beyond diagnosis: Case formulation approaches in CBT* (pp. 81-103). Wiley.
- Borsboom, D. (2023). B.F. Skinner en het laatste grote verhaal. *Gedragstherapie*, 56, 460-463.
- Bruch, M. (1998). The UCL case formulation model: clinical applications and procedures. In M. Bruch & F.W. Bond (eds.), *Beyond diagnosis: Case formulation approaches in CBT* (pp. 19-43). Wiley.
- Burger, A. (1994). *Functie-analyse van neurotisch gedrag: Een handleiding voor gedragstherapeuten* (3de druk). Van Rossum.
- Davey, G. C. L. (2006). Cognitive mechanisms in fear acquisition and maintenance. In: M. G. Craske, D. Hermans, & D. Vansteenwegen (eds.), *Fear and learning: From basic processes to clinical implications* (pp. 99-117). American Psychological Association. <https://doi.org/10.1037/11474-000>
- De Houwer, J., & Boddez, Y. (2023). Over Skinner, regelgeleid gedrag en relational frame theory: Denken als gedrag. *Gedragstherapie*, 56, 402-415.
- Hermans, D., Raes, F., & Orlemans, H. (2018). *Inleiding tot de gedragstherapie*. Bohn Stafleu van Loghum.
- Hosoba, T., Iwanaga, M., & Seiwa, H. (2001). The effect of UCS-inflation and deflation procedures on 'fear' conditioning. *Behaviour Research and Therapy*, 39, 465-475. [https://psycnet.apa.org/doi/10.1016/S0005-7967\(00\)00025-5](https://psycnet.apa.org/doi/10.1016/S0005-7967(00)00025-5)
- Korrelboom, K., & ten Broeke, E. (2014). *Geïntegreerde cognitieve gedragstherapie: Handboek voor theorie en praktijk*. Uitgeverij Coutinho.
- Mowrer, O. H. (1947). On the dual nature of learning: A reinterpretation of 'conditioning' and 'problem-solving'. *Harvard Educational Review*, 17, 102-148.
- Orlemans, J. W. G. (1988). *Inleiding tot de gedragstherapie*. Van Loghum Slaterus.
- ten Broeke, E., van der Heiden, C., Meijer, S., Driessen, A., & Korrelboom, K. (2022). *Cognitieve gedrags therapie: de basisvaardigheden*. Boom.

.....

## *Verduidelijkingen en bijkomende overwegingen bij het belang van functionele theorieën*

---

Een antwoord op de reactie van ten Broeke  
& Korrelboom (2024)

---

JAN DE HOUWER & YANNICK BODDEZ

### *Samenvatting*

.....

In het recent verschenen themanummer van *Gedragstherapie* rond Skinner, schreven we een bijdrage over de relatie tussen het werk van Skinner enerzijds, en *relational frame theory* (RFT) anderzijds. Ten Broeke en Korrelboom (2024) reageren op onze bespreking, waarbij de kern van hun betoog is dat wij zouden beweren dat in de huidige gedragstherapeutische functieanalyses te zeer wordt uitgegaan van de feitelijke leefwereld van de patiënt en onvoldoende van de subjectieve beleving daarvan. Deze bezorgdheid berust naar onze mening op een misverstand. Wel zien we aanknopingspunten voor een alternatief conceptueel en theoretisch kader voor het maken van functionele analyses en de daarbij horende betekenisanalyses, waarbij een cognitieve visie op begrippen als CS, US en CR plaatsmaakt voor strikt functionele begrippen, zoals regelgeleid gedrag en AARR.

*Trefwoorden: Skinner, functieanalyse, relational frame theory, regelgeleid gedrag, betekenisanalyse*

**Kernboodschappen voor de klinische praktijk**

- ▶ Functionele analyses zoals die werden uitgevoerd voor de komst van de cognitieve gedragstherapie, beperkten zich ertoe te beschrijven hoe gedrag een functie is van feitelijke contingenties in de omgeving.
- ▶ Zowel de cognitieve benadering als Skinner stelt dat het beter is om ook de subjectieve beleving van de patiënt te betrekken in analyses.
- ▶ Om dit te doen worden de oorspronkelijke functionele analyses in de cognitieve benadering wat vervormd om ze toe te kunnen passen. Wij pleiten voor een nieuw begrippenkader op basis van de theorie van Skinner en de *relational frame theory*.

Op vraag van de redacteurs van het themanummer rond Skinner, schreven we recent een bijdrage over de relatie tussen het werk van Skinner enerzijds, en *relational frame theory* (RFT) anderzijds (De Houwer & Boddez, 2023). Hierbij gingen we ook in op de klinische implicaties van de ideeën van Skinner en RFT, meer bepaald rond het opstellen van functionele analyses. In een reactie op onze bijdrage focussen ten Broeke en Korrelboom (2024) op onze bespreking van de klinische implicaties. De kern van hun betoog is dat wij zouden beweren dat ‘in de huidige gedragstherapeutische functieanalyses ten onrechte te zeer wordt uitgegaan van de feitelijke leefwereld van de patiënt en onvoldoende van de subjectieve beleving daarvan’, wat volgens hen ‘op flink gespannen voet [staat] met de ter zake relevante literatuur en met de feitelijke behandelpraktijk’ (p. 210).

Deze bezorgdheid berust naar onze mening op een misverstand. We beargumenteerden dat de ideeën van Skinner rond regelgeleid gedrag en de uitwerking ervan binnen de RFT een aanpassing vereisen van *traditionele* functionele analyses. Hiermee doelden we op analyses die het gedrag van mensen uitsluitend beschrijven ‘met behulp van functionele begrippen en principes die ontwikkeld zijn op basis van dieronderzoek’ en ‘waarin gedrag enkel begrepen wordt vanuit feitelijke contingenties in de omgeving’ (De Houwer & Boddez, 2023, p. 406). In de inleiding verduidelijkten we dat het begrip ‘functioneel’ voor ons verwijst naar hoe gedrag een functie is van elementen in de (vroegere en huidige) omgeving zonder dat er verwezen wordt naar de cognitieve mechanismen waardoor omgeving een impact heeft op gedrag. Het gaat dus om functionele analyses zoals die oorspronkelijk werden uitgevoerd in de gedragstherapie, voor de komst van de cognitieve gedragstherapie en zoals die nu nog steeds gebruikelijk zijn in de toegepaste gedragsanalyse (*applied behavior analysis*; zie bijvoorbeeld Roane et al., 2015).

Een van de centrale punten in ons oorspronkelijk artikel was dat het werk van Skinner rond regelgeleid gedrag en de verderzetting hiervan binnen de RFT-literatuur impliceren dat dergelijke traditionele functionele analyses tekortschieten voor de analyse van het gedrag van talige personen. Onze re-

den om deze boodschap onder de aandacht te brengen was niet om de lezer eraan te herinneren dat traditionele functionele analyses tekortschieten – wij zijn ons weldegelijk bewust van het feit dat net deze conclusie mee aan de basis lag van het ontstaan van de cognitieve benadering binnen de gedragstherapie en dat die conclusie algemeen aanvaard wordt door cognitieve gedragstherapeuten. Onze doelstelling was om duidelijk te maken dat deze conclusie óók volgt uit het werk van Skinner zelf en uit meer recent (RFT-gerelateerd) werk binnen de functionele (gedragsanalytische) benadering in de psychologie (zie Dixon & Hayes, 2022). Gezien de misverstanden over het werk van Skinner en de frequente discussies over de relatie tussen de cognitieve therapie en de (traditionele, strikt functionele) gedragstherapie, alsook de recente aandacht voor *acceptance and commitment therapy* (ACT), leek het ons waardevol om, in het kader van een themanummer rond Skinner, de ideeën van Skinner rond regelgeleid gedrag in herinnering te brengen en te verduidelijken hoe deze gerelateerd zijn aan RFT, ACT en functionele analyses zoals deze in de huidige praktijk van de cognitieve gedragstherapie courant zijn.

Met betrekking tot dat laatste punt benadrukten we zowel raakvlakken als verschillen. We waren enigszins verrast dat ten Broeke en Korrelboom onze bespreking van de raakvlakken blijkbaar over het hoofd hebben gezien. Zo schreven we op pagina 413 dat een functionele analyse in termen van regelgeleid gedrag of *arbitrarily applicable relational responding* (AARR) veel weg heeft van ‘hoe een cognitieve gedragstherapeut een functionele analyse uitvoert (zie bijvoorbeeld: Vanden Bogaerde et al., 2022). Ook cognitieve gedragstherapeuten zullen niet enkel kijken naar de feitelijke leefwereld van de patiënt, maar tevens naar diens subjectieve leefwereld, zowel wat betreft de gepercipieerde gevolgen van gedrag als de subjectieve betekenis van gebeurtenissen in de omgeving. Door rekening te houden met de ideeën van regelgeleid gedrag en AARR wordt de kloof tussen een strikt functionele en een meer cognitief geïnspireerde klinische praktijk dus duidelijk kleiner.’

Blijft nog de vraag die ten Broeke en Korrelboom stellen in de titel van hun reactie: ‘Voor welke problemen bieden regelgeleid gedrag en de functionele analyse 2.0 eigenlijk oplossingen?’ Deze vraag lag niet aan de basis van ons oorspronkelijk artikel – zoals eerder gezegd bestond onze opdracht er voornamelijk in om de relatie tussen het werk van Skinner en RFT te duiden – maar het is wel een interessante vraag. Een aanzet tot het beantwoorden van deze vraag kan gevonden worden in ons oorspronkelijk artikel, meer bepaald in de bespreking van de verschillen tussen een strikt functionele en een meer cognitief geïnspireerde klinische praktijk. Om dit te stofferen kunnen we opnieuw citeren uit het oorspronkelijk artikel: ‘Toch blijven er belangrijke verschillen. Voor een functionele therapeut is de subjectieve leefwereld van de patiënt geen mentale toestand binnen in de patiënt, maar een geheel van gedragingen. [...] De functioneel geïnspireerde therapeut blijft zich dus beperken tot het perspectief van een buitenstaander die de

patiënt zich ziet gedragen in een historische context, maar die nu gebruikmaakt van een breder conceptueel arsenaal, waaronder de begrippen regelgeleid gedrag en AARR' (p. 413). Het verschil tussen een strikt functionele en meer cognitief geïnspireerde klinische praktijk is dus vooral een verschil in perspectief. Zoals we aangaven in ons oorspronkelijk artikel, zien we de functionele en cognitieve benadering in de (klinische) psychologie niet als incompatibel en is het voor ons dus niet nodig om te kiezen voor ofwel de ene benadering ofwel de andere (zie p. 405). We zien de verschillende benaderingen eerder als verschillende perspectieven die elk kunnen bijdragen aan probleemoplossend (klinisch) handelen (zie p. 411). Terugverwijzend naar de vraag in de titel van ten Broeke en Korrelboom biedt deze visie misschien geen oplossing voor een specifiek praktisch probleem in de huidige klinische praktijk, maar wel een potentiële verrijking voor de klinische praktijk.

Het tweede deel van ons antwoord op de vraag van ten Broeke en Korrelboom is eerder impliciet aanwezig in de argumenten die we in ons oorspronkelijk artikel naar voor brachten. Traditionele functionele begrippen zoals klassieke en operante conditionering verwijzen uitsluitend naar de feitelijke (vroegere en huidige) omgeving en naar hoe gedrag daar een functie van is (zie De Houwer et al., 2024, voor een bespreking). Als dergelijke begrippen volgens Skinner en RFT problematisch zijn voor de functionele analyse van het gedrag van talige mensen, dan zijn die begrippen ook problematisch wanneer ze in hun oorspronkelijke, strikt functionele betekenis gebruikt worden door cognitieve gedragstherapeuten.

Een cognitieve gedragstherapeut zal wellicht antwoorden dat zij of hij traditioneel functionele begrippen niet meer gebruikt in de oorspronkelijke, strikt functionele betekenis (zoals klassieke en operante conditionering als veranderingen in gedrag als gevolg van contingenties in de vroegere en huidige omgeving) maar eerder in termen van mentale processen (zoals klassieke en operante conditionering als het zoeken naar relaties in de omgeving en het vormen van subjectieve verwachtingen over de toekomstige gebeurtenissen). Wij worstelen echter met de conceptuele, theoretische en praktische implicaties van deze aanpak, vooral omdat hierdoor het onderscheid tussen het functionele en het cognitieve niveau van verklaring wordt weggeveegd (zie Hughes et al., 2016). We kunnen, bijvoorbeeld, begrijpen dat het ervaren van een paniekaanval in een lift gezien kan worden als conditioneringsepisode waarbij de lift de rol speelt van de CS en de paniekaanval de rol van een US, en de angst voor de lift als een CR ten gevolge van het samengaan van de lift en de paniekaanval. Deze analyse is strikt functioneel omdat er enkel verwezen wordt naar gebeurtenissen in de vroegere en huidige omgeving en hun potentiële impact op gedrag. Wat is echter nog de betekenis van de begrippen zoals CS, US en CR als men ze, bijvoorbeeld, gebruikt om te verwijzen naar spanning (CR) die optreedt op het werk (CS) omwille van de overtuiging 'ik ben een teleurstelling als ik niet zorg voor stabiliteit en veiligheid' (US; zie Vanden Bogaerde et al., 2022, p. 159)? Der-

gelijke analyses kunnen een therapeutische meerwaarde hebben (bijvoorbeeld door aanleiding te geven tot interventies voor het wijzigen van specifieke overtuigingen), maar we stellen ons wel vragen bij de noodzaak om oorspronkelijk functionele begrippen als CS, US en CR zo te vervormen dat ze op een dergelijke manier kunnen worden ingezet. Zijn deze vervormde begrippen nog meer dan een ongemakkelijke herbeschrijving van het idee dat er een prikkel is (CS) die klachten ontlokt (CR) omwille van een bepaalde gedachte (US)?

Als men toch functionele concepten wil gebruiken om dergelijke cognitieve analyses aan te vullen, is het dan niet beter om concepten te gebruiken zoals regelgeleid gedrag of AARR (bijvoorbeeld: de patiënt gedraagt zich volgens de regel dat hij of zij in de werksituatie tekortschiet in het zorgen voor stabiliteit en veiligheid)? Dit is niet louter een kwestie van woorden (zoals 'overtuiging' vervangen door 'regel' of 'relateren aan') want het kan ook implicaties hebben voor handelen (bijvoorbeeld een zoektocht naar de antecedenten en consequenten van het regelgeleid gedrag of de contextuele cues die het relationeel gedrag controleren). We blijven dus geloven in een meerwaarde van een écht functioneel perspectief in de cognitieve gedragstherapie. Dit is echter niet noodzakelijk het traditionele functionele perspectief dat in wezen moet vervormd worden tot een cognitief perspectief om zinvol te zijn, maar een strikt functioneel perspectief dat geënt is op moderne functionele analyses van talig gedrag en net daardoor compatibel is met een cognitief perspectief op psychisch lijden.

We beseffen dat dit fundamentele thema's zijn die niet zomaar in het kader van een kort artikel, laat staan een antwoord op een reactie op een artikel, kunnen behandeld worden. We hopen in de komende maanden en jaren vooruitgang te boeken in het bieden van een alternatief conceptueel en theoretisch kader voor het maken van functionele analyses en de daarbij horende betekenisanalyses. We zullen onderzoeken of begrippen uit de RFT (zoals AARR, contextuele relationele cue en contextuele functionele cue) een strikt functioneel alternatief kunnen bieden voor of aanvulling kunnen zijn op begrippen uit de traditionele conditioneringsliteratuur (zoals CS, US of bekrachtiging). Hierbij blijven we streven naar een duidelijk onderscheid tussen analyses op het functionele niveau en op het cognitieve niveau omdat op die manier de meerwaarde van elk perspectief gemaximaliseerd wordt (Hughes et al., 2016). We ambiëren in de toekomst dus een vollediger antwoord te kunnen bieden op de vraag die ten Broeke en Korrelboom stellen in de titel van hun reactie, maar hopelijk hebben we bij deze voldoende verduidelijkt wat de boodschap was van ons oorspronkelijk artikel en waarin de mogelijke meerwaarde van dat artikel is gelegen.

*De voorbereiding van dit artikel werd mogelijk gemaakt door de steun van het Methusalem project BOF22/MET\_V/002 van de Universiteit Gent aan Jan De Houwer.*

**Jan De Houwer** en **Yannick Boddez** zijn beiden verbonden aan de Universiteit Gent.  
*Correspondentieadres:* Jan De Houwer, Universiteit Gent, Henri Dunantlaan 2,  
 B-9000 Gent, België. E-mail: jan.dehouwer@ugent.be

**Summary** *Clarifications and additional considerations on the importance of functional theories of linguistic behavior for cognitive behavioral therapy: An answer to the response from ten Broeke & Korrelboom (2024)*

In the recently published special issue on Skinner of *Gedragstherapie*, we wrote a contribution about the relationship between Skinner's work on the one hand, and relational frame theory (RFT) on the other. Ten Broeke and Korrelboom (2024) respond to our discussion, where the core of their argument is that we would argue that current behavioral therapeutic functional analyses rely too much on the patient's actual living environment and insufficiently on the subjective experience of it. In our opinion, this concern is based on a misunderstanding. We do see starting points for an alternative conceptual and theoretical framework for making functional analyses and the associated meaning analyses, in which a cognitive view on concepts such as CS, US and CR makes room for strictly functional concepts such as rule-governed behavior and AARR.

**Keywords** *Skinner, function analysis, relational frame theory, rule-guided behavior, meaning analysis*

## Referenties

- De Houwer, J., & Boddez, Y. (2023). Over Skinner, regelgeleid gedrag, en relationale frame theorie: Denken als gedrag. *Gedragstherapie*, 56, 402-415.
- De Houwer, J., Finn, M., Boddez, Y., Hughes, S., & Cummins, J. (2024). Relating different perspectives on how outcomes of behavior influence behavior. *Journal of the Experimental Analysis of Behavior*, 121, 123-133. <https://doi.org/10.1002/jeab.887>
- Dixon, M. R., & Hayes, S. C. (2022). On the disruptive effects of behavior analysis on behavior analysis: The high cost of keeping out acceptance and commitment therapy and training. *Behavior Analysis in Practice*. <https://doi.org/10.1007/s40617-022-00742-4>
- Hughes, S., De Houwer, J., & Perugini, M. (2016). The functional-cognitive framework for psychological research: Controversies and resolutions. *International Journal of Psychology*, 51, 4-14. <https://doi.org/10.1002/ijop.12239>
- Roane, H. S., Ringdahl, J. E., & Falcomata, T. S. (Eds.). (2015). *Clinical and organizational applications of applied behavior analysis*. Elsevier Academic Press.
- ten Broeke, E., & Korrelboom, K. (2024). Voor welke problemen bieden regelgeleid gedrag en de functionele analyse 2.0 eigenlijk oplossingen? Een reactie op De Houwer & Boddez (2023). *Gedragstherapie*, 57, 206-213.
- Vanden Bogaerde, A., Stas, P., Hoorelbeke, K., & Koster, E. (2022). *Cognitieve gedragstherapie in de praktijk: Verandering op maat*. Acco.

.....

## *Is er interactie tussen psychotherapie en farmacotherapie?*

.....

Cuijpers, P., Miguel, C., Harrer, M., Ciharova, M., & Karyotaki, E. (2023). Does the use of pharmacotherapy interact with the effects of psychotherapy? A meta-analytic review. *European Psychiatry*, 66, e63, 1-9. <https://doi.org/10.1192/j.eurpsy.2023.2437>

---

Grote meta-analyses met honderden onderzoeken hebben aangetoond dat psychotherapie en farmacotherapie effectief zijn bij de behandeling van depressie. Het is ook bekend dat de combinatie van de twee effectiever is dan elk afzonderlijk. Psychotherapie blijkt op de langere termijn effectiever dan farmacotherapie, hoewel de combinatie van beide nog steeds effectiever is. Het is alleen niet duidelijk of er sprake is van een interactie tussen psychotherapie en farmacotherapie. In dit artikel wordt onderzocht of (1) er helemaal geen interactie is (de effecten zijn onafhankelijk van elkaar), (2) antidepressiva de effecten van psychotherapie verminderen, en (3) antidepressiva de effecten van psychotherapie versterken.

De auteurs hebben meta-analyses uitgevoerd van gerandomiseerde onderzoeken (RCT's) waarin psychotherapieën voor depressie bij volwassenen werden vergeleken met inactieve controlegroepen (zoals placebo en wachtlijst). De onderzoeken die zijn opgenomen in de meta-analyse, zijn gevonden via een al bestaande database van gerandomiseerde onderzoeken naar de psychologische behandeling van depressie. Het aandeel gebruikers van antidepressiva werd gebruikt als voorspeller van de effectgrootte van de psychotherapeutische behandeling.

De effectgroottes zijn op verschillende manieren berekend om te onderzoeken of verschillende methoden tot verschillende uitkomsten resulteerden. Zo voerden de auteurs een univariate meta-regressieanalyse uit om het verband te onderzoeken tussen de effectgrootte en het aandeel deelnemers dat bij aanvang een antidepressivum gebruikte. Vervolgens voerden ze een multivariabele meta-regressieanalyse uit waarin ze ook andere belangrijke

kenmerken van de deelnemers (waaronder de ernst bij aanvang), de interventies en de onderzoeken toevoegden. Op die manier werd nagegaan of deze kenmerken het resultaat beïnvloedden.

In totaal voldeden 300 gerandomiseerde gecontroleerde onderzoeken (met 353 vergelijkingen tussen een behandelings- en een controlegroep) aan de inclusiecriteria voor deze meta-analyse. Aan de onderzoeken namen in totaal 32.852 patiënten deel, 17.405 aan de interventie en 15.447 aan de controlecondities. De belangrijkste effectgrootte van psychotherapie was  $g = 0,71$  (95% BI: I2 0,64; 0,79) met een hoge heterogeniteit ( $= 82$ ; 95% BI: 80; 84). Dit betekent dat psychotherapie voor depressie tot betere resultaten leidt dan de controlegroepen. De auteurs vonden geen significant verband tussen het aandeel antidepressivagebruikers en de effectgrootte ( $p = .07$ ). Ze vonden wel een significante associatie met enkele andere voorspellers, waaronder het type controlegroep en het risico op bias. Het gebruik van antidepressiva werd geassocieerd met hogere responspercentages binnen de controlegroepen (zonder actieve behandeling), maar niet met de relatieve effecten van de behandelingen, vergeleken met de controlegroepen.

De resultaten van dit onderzoek suggereren dat de eerste mogelijkheid – er is geen interactie tussen psychotherapie en farmacotherapie – waarschijnlijk juist is en dat het gebruik van antidepressiva de effecten van psychologische behandelingen niet verstoort. Dit is goed nieuws vanuit klinisch perspectief. Blijkbaar kunnen patiënten starten met psychotherapie en hoeven ze niet bang te zijn dat de effecten van de therapie kleiner zullen zijn als ze antidepressiva gebruiken. Het is algemeen bekend dat een gecombineerde behandeling effectiever is dan alleen psychotherapie of farmacotherapie, maar patiënten kunnen er de voorkeur aan geven om met één behandeling te beginnen in plaats van meteen met de combinatie. In dergelijke gevallen is het goed om te weten dat de bijkomende behandeling even effectief zal zijn als deze later wordt gestart en in combinatie met de andere.

#### COMMENTAAR VAN DE REDACTIE (SARA SCHEVENEELS)

.....

In lijn met bestaande behandelrichtlijnen wordt psychotherapie bij depressie vaak in combinatie met een antidepressivum aangeboden. De vraag (en in sommige gevallen bezorgdheid) werpt zich dan op of patiënten die een antidepressivum nemen anders reageren op psychotherapie. Deze meta-analyse vindt geen associatie tussen het percentage deelnemers dat een antidepressivum nam en de uitkomst van psychotherapie voor depressie in klinische trials. Deelnemers die medicatie namen bleken het dus niet slechter maar ook niet beter te doen na psychotherapie dan deelnemers die geen medicatie namen. Deze bevinding wijkt enigszins af van voorgaand onderzoek dat betere effecten vond van een combinatie van psychotherapie en medicatie dan van de afzonderlijke behandelingen. De auteurs concluderen dat het

gebruik van een antidepressivum in ieder geval niet op een nadelige manier lijkt te interfereren met psychotherapie.

Hoewel dit een belangrijke bevinding is, kunnen er enkele kanttekeningen gemaakt worden bij de resultaten. Ten eerste is een beperking van de meta-analyse dat het een correlatieel onderzoek betreft, waardoor enkel conclusies over een verband getrokken kunnen worden, maar niet over de daadwerkelijke impact van medicatie op de uitkomsten van psychotherapie. Hiervoor zouden klinische trials met vier armen ofwel condities moeten worden geanalyseerd: (1) combinatie psychotherapie en medicatie, (2) enkel psychotherapie, (3) enkel medicatie, (4) placebo- of wachtlijstcontrolecondities. Er kan dan worden nagegaan of de effecten van de combinatie (1) gelijk of verschillend zijn aan die van (de som van) de twee aparte condities (2 en 3). Ten tweede worden in de meta-analyse verschillende vormen van psychotherapie geanalyseerd, waaronder ook minder effectieve therapieën. De vraag is nog maar of conclusies hetzelfde blijven wanneer enkel naar evidence-based psychotherapie gekeken wordt. Niettegenstaande kan in geval van psychotherapie die al erg effectief is net voorspeld worden dat er minder ruimte is voor verdere verbetering door medicatie. Ten derde werd in de meta-analyse gefocust op de uitkomsten op korte termijn. Op lange termijn worden mogelijks andere resultaten gevonden. Wanneer de effecten van psychotherapie onterecht worden toegeschreven aan de medicatie, is er het risico op een hogere terugval wanneer de medicatie gestopt wordt. Dit kan een belangrijke richting zijn voor verder onderzoek, waarbij naar de effecten op langere termijn gekeken wordt en specifiek na het stoppen van medicatie. Ten vierde is het denkbaar dat medicatie in sommige gevallen op een gunstige manier met psychotherapie interageert, bijvoorbeeld bij patiënten die dermate geïnactiveerd zijn dat ze weinig ontvankelijk zijn voor psychotherapie en bij wie medicatie een eerste stap is om de patiënt te activeren. Ten slotte is deze vraag naar de interactie tussen medicatie en psychotherapie ook erg relevant binnen de behandeling van angststoornissen, waar zich een gelijkaardige vraag stelt of medicatie de effecten van psychotherapie kan versterken dan wel er afbreuk aan doet.

De redactie van *Gedragstherapie* ontving de volgende boeken. U bent van harte uitgenodigd deze te recenseren. Bent u geïnteresseerd, stuurt u dan een berichtje met een korte motivatie en een eveneens korte beschrijving van uw achtergrond naar: [redactie@tijdschriftgedragstherapie.nl](mailto:redactie@tijdschriftgedragstherapie.nl).

*Gedragstherapie* is een peer reviewed tijdschrift. Plaatsing van een recensie is daarom niet gegarandeerd. Informatie over eisen die de redactie stelt aan boekbesprekingen vindt u in de auteursrichtlijnen ([www.tijdschriftgedragstherapie.nl/auteurs/auteursrichtlijnen](http://www.tijdschriftgedragstherapie.nl/auteurs/auteursrichtlijnen)).

den Daas, N. (2024). *Zo ongelukkig mogelijk in 8 stappen: Leefregels van een eigenzinnige psycholoog*. Boom.

Derkman, M., Roos, S., & van Tetering, E. (2024). *Mijn ID van ADHD*. Bohn Stafleu van Loghum.

Derkman, M., Roos, S., & van Tetering, E. (2024). *Psycho-educatie over ADHD aan kinderen, jongeren en hun omgeving: Handleiding bij Mijn ID van ADHD*. Bohn Stafleu van Loghum.

Draaisma, N., Kramer, A., & Pameijer, N. (2024). *Handelingsgerichte diagnostiek in de Jeugdzorg en de JGGZ: Samen van vraag naar plan*. Acco.

Soderlind Warren, C. (2024). *Afkicken van je ex: Zelfhulp bij (toxisch) liefdesverdriet*. Boom.

Spuij, M. (2024). *Rouw bij kinderen en jongeren: Over het begeleiden van verliesverwerking* (Tweede, herziene editie). Uitgeverij Nieuwezijds.

van der Wijngaart, R., & van Genderen, H. (2024). *Schematherapie voor cluster C-persoonlijkheidsstoornissen: Behandeling van afhankelijke, vermijdende en dwangmatige cliënten*. Bohn Stafleu van Loghum.

Verbeek, I. (2024). *Rustig slapen: Je leven lang een goede nachtrust*. Boom.

# Erratum

---

In het artikel van van Bentum en collega's (2019) in *Gedragstherapie* willen de auteurs de volgende wijziging aanbrengen.

Op pagina 337 staat de zin: 'EMDT is gebaseerd op de bekende EMDR-therapie (Shapiro, 2001).'

Dit moet zijn: 'EMDT volgde grotendeels de procedure van EMDR-therapie (Shapiro, 2001), zoals beschreven in het *Handboek EMDR* (de Jongh & ten Broeke, 2012).'

## Referenties

de Jongh, A., & ten Broeke, E. (2012).  
*Handboek EMDR: Een geprotocolleerde  
 behandelmethode voor de gevolgen van  
 psychotrauma*. Pearson.

van Bentum, J., Sijbrandij, M., Kerkhof,  
 A., van Schaik, A., & Huibers, M.  
 (2019). Eye Movement Dual Task  
 (EMDT) als behandeling van suïci-

dale intrusies bij depressieve patiën-  
 ten: Pilotstudie naar de veiligheid,  
 verdraagbaarheid en uitvoerbaarheid  
 van de interventie. *Gedragstherapie*,  
 52, 333-352. [www.tijdschriftgedrags-therapie.nl/inhoud/tijdschrift\\_](http://www.tijdschriftgedrags-therapie.nl/inhoud/tijdschrift_artikel/TG-2019-4-5/Eye-Movement-Dual-Task-als-behandeling-van-suicidale-intrusies-bij-depressieve-patienten)  
 artikel/TG-2019-4-5/Eye-Movement-  
 Dual-Task-als-behandeling-van-  
 suicidale-intrusies-bij-depressieve-  
 patienten

### **Abonnementenadministratie**

Uitgeverij Boom  
Postbus 15970, 1001 NL Amsterdam  
(020) 520 0120  
abonnementen@boom.nl

### **Abonnementen**

Leden en deelnemers van de VGCT en van de VVGT ontvangen *Gedragstherapie*, inclusief toegang tot het digitale archief, als onderdeel van hun lid- of deelnemerschap.

Voor de actuele prijzen en mogelijke abonnementsvormen, zie:  
[www.tijdschriftgedragstherapie.nl](http://www.tijdschriftgedragstherapie.nl).

### **Nieuwe abonnementen**

Abonnementen kunnen op elk gewenst tijdstip ingaan. Nieuwe abonnementen kunnen alleen via [www.tijdschriftgedragstherapie.nl/abonnementen/abonneren](http://www.tijdschriftgedragstherapie.nl/abonnementen/abonneren) worden opgegeven of per e-mail aan [abonnementen@boom.nl](mailto:abonnementen@boom.nl). Vanwege de aard en inhoud van de uitgave wordt u geacht het abonnement in het kader van uw beroep of bedrijf te ontvangen en niet als consument op te treden in de zin van de artikelen 236 en 237 boek 6 BW. Mocht dit anders zijn, dan bent u gehouden dit binnen één maand na ingang van het abonnement per e-mail ([abonnementen@boom.nl](mailto:abonnementen@boom.nl)) of schriftelijk bij de abonnementenadministratie van de uitgever aan te geven.

### **Wijziging of beëindiging abonnement**

Bij wijziging van naam en/of adres verzoeken wij u een e-mail te zenden aan de abonnementenadministratie.

Abonnementen worden aangegaan voor de duur van één jaar. Het abonnement wordt na verloop van één jaar telkens stilzwijgend verlengd. Na het eerste jaar geldt voor particulieren een opzegtermijn van 1 maand. Zakelijke abonnementen kunnen opgezegd worden aan het eind van een jaar, met inachtneming van 1 maand voor het einde van de termijn. Opzegging geschiedt schriftelijk of per e-mail. De opzegging kan worden gericht aan de abonnementenadministratie op [abonnementen@boom.nl](mailto:abonnementen@boom.nl).

### **Betaling**

U ontvangt een factuur voor de betaling.

## GEDRAGSTHERAPIE

Zevenvijftigste jaargang, nummer 2, juni 2024

Pagina 117-224

117 *Redactioneel*

---

118 **Yannick Vander Zwalmen, David Demeester, Kristof Hoorelbeke & Ernst Koster**

Effectiviteit van een adaptieve cognitieve training bij depressieve klachten:  
Een systematische review en meta-analyse

147 **Edwin de Beurs, Jan Boehnke & Eiko Fried**

Universele schalen voor testuitslagen: Een pleidooi voor T-scores en  
percentiel rangorde scores

179 **Anne Schimmel & Julie Krans**

Virtual reality voorafgaand aan exposure bij patiënten met een paniek-  
stoornis en/of sociale-angststoornis: Een pilotstudie

---

## FORUM

206 **Erik ten Broeke & Kees Korrelboom**

Voor welke problemen bieden regelgeleid gedrag en de functionele analyse  
2.0 eigenlijk oplossingen? Een reactie op De Houwer & Boddez (2023)

214 **Jan De Houwer & Yannick Boddez**

Verduidelijkingen en bijkomende overwegingen bij het belang van  
functionele theorieën: Een antwoord op de reactie van ten Broeke &  
Korrelboom (2024)

---

## VOOR U GELEZEN

220 Cuijpers, P., Miguel, C., Harrer, M., Ciharova, M., & Karyotaki, E. (2023).  
Does the use of pharmacotherapy interact with the effects of psycho-  
therapy? A meta-analytic review. *European Psychiatry*, 66, e63, 1-9.

223 *Ontvangen*

224 *Erratum*