



Universiteit
Leiden
The Netherlands

Computerized adaptive testing in Dutch mental health care: a new tool to assess depression and anxiety

Slok-Flens, G.

Citation

Slok-Flens, G. (2022, October 5). *Computerized adaptive testing in Dutch mental health care: a new tool to assess depression and anxiety*. Retrieved from <https://hdl.handle.net/1887/3466118>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3466118>

Note: To cite this publication please use the final published version (if applicable).

Dutch Summary

Computergestuurd Adaptief Testen in de Nederlandse Geestelijke Gezondheidszorg

Een nieuw hulpmiddel voor het meten van Depressie en Angst

Hoofdstuk 1: Introductie

In Nederland hebben meer dan 4 op de 10 personen in hun leven last gehad van een of meerdere psychische aandoeningen. De prevalentie van de meest voorkomende psychische aandoeningen is bovendien niet veel veranderd in de afgelopen decennia. Met andere woorden: psychische aandoeningen komen frequent voor in Nederland zonder duidelijke afname over tijd (Bijl, van Zessen, Ravelli, de Rijk, & Langendoen, 1998; Bijl, Ravelli, & van Zessen 1998; de Graaf, ten Have, & van Dorsselaer; 2010, 2012). Het is daarom relevant dat we blijven zoeken naar nieuwe oplossingen om dit maatschappelijke probleem te verkleinen.

Elk jaar wordt ongeveer 6% van de Nederlandse bevolking behandeld voor psychische problemen (de Beurs, Barendregt, & Warmerdam, 2017). Om patiënten effectief en efficiënt te behandelen kunnen klinici gebruik maken van verschillende hulpmiddelen. Een van die hulpmiddelen is het gebruik van zelfrapportage-meetinstrumenten. Een zelfrapportage-meetinstrument bestaat uit een reeks vragen (items) over een of meerdere constructen die wordt beantwoord door de patiënt, bijvoorbeeld over depressie of angst. Door relevante constructen periodiek te meten tijdens de behandeling worden klinici ondersteund bij de klinische besluitvorming over de aard en de ernst van de klachten van de patiënt, de invulling van de behandeling en de beëindiging daarvan, de behandeling van non-responders, en bij terugvalpreventie (de Beurs et al., 2018; Greenhalgh et al., 2018; Lambert, 2010; Lewis et al., 2015; Martin-Cook et al., 2021). Hierdoor kan de motivatie van patiënten toenemen om de behandeling voort te zetten en kunnen behandelresultaten verbeterd worden (de Jong et al., 2021; Fortney et al., 2017; Guo et al., 2015; Rush & Thase, 2018; Scott & Lewis, 2015). Wordt het gebruik van meetinstrumenten daarnaast gecombineerd met samen beslissen (d.w.z., het ondersteunen van patiënten om te participeren in de besluitvorming over de behandeling), dan kunnen behandelresultaten nog verder worden verbeterd (Metz et al., 2019; van der Feltz-Cornelis et al., 2014). Ten slotte maakt de aggregatie van meetinstrumentcores het mogelijk om de algehele kwaliteit en de waarde van zorg voor patiënten te verbeteren door groepen met elkaar te vergelijken en hiervan te leren (de Beurs et al., 2018; Porter, 2009).

Meetinstrumenten zijn het meest geschikt als hulpmiddel voor het evalueren van de behandeling als ze betrouwbaar zijn (d.w.z., de resultaten zijn consistent over replicaties; Crocker & Algina, 1986), valide (d.w.z., het te meten construct wordt adequaat gemeten; Cook & Campbell, 1979), responsief (d.w.z., verandering in de tijd wordt voldoende gedetecteerd in het gemeten construct; Mokkink et al., 2010), en efficiënt (d.w.z., het aantal items is zo klein mogelijk om de patiënt zo min mogelijk te belasten). Voor veel Nederlandse meetinstrumenten is het echter onduidelijk of ze voldoen aan deze criteria, omdat relevante psychometrische

eigenschappen onvoldoende zijn onderzocht. Daarnaast is het een uitdaging om meetinstrumenten te ontwikkelen die zowel zeer betrouwbaar zijn als zeer efficiënt, omdat nagenoeg alle instrumenten uit een vaste set items bestaan. Hierdoor gaat een hogere betrouwbaarheid vaak ten koste van de efficiëntie en een hogere efficiëntie vaak ten koste van de betrouwbaarheid. Ten slotte maakt het grote aantal beschikbare meetinstrumenten het lastig voor veel zorgaanbieders om van elkaar te leren omdat meetinstrumenten altijd wel wat verschillen in content, antwoordcategorieën en psychometrische eigenschappen.

Het doel van dit proefschrift is het neerzetten van de basis voor een nieuw type meetinstrumentarium dat de genoemde beperkingen oplost door moderne meettechnieken toe te passen. Om dit doel te bereiken wordt een uitgebreide psychometrische evaluatie uitgevoerd op twee nieuwe meetinstrumenten die zijn ontwikkeld voor het meten van generieke depressie en angst. Depressie en angst zijn gekozen omdat het wereldwijde problemen zijn (Baxter, Scott, Vos, & Whiteford, 2013; Marcus, Yasamy, van Ommeren, Chisholm, & Saxena, 2012) en de meest voorkomende geestelijke gezondheidsproblemen in Nederland (de Graaf, ten Have, van Gool, & van Dorsselaer, 2012). Bovendien zijn depressie en angst vaak aanwezig bij andere psychische aandoeningen en zijn ze meestal de belangrijkste reden om geestelijke gezondheidszorg te zoeken (Clarke & Kissane, 2002; Frank, 1974).

De geëvalueerde meetinstrumenten zijn ontwikkeld door het Patient-Reported Outcomes Measurement Information System (PROMIS®) initiatief. PROMIS is aan het begin van deze eeuw opgericht in de Verenigde Staten (VS) om een nieuw, uniform zelfrapportage-meetinstrumentarium te ontwikkelen voor het meten van een groot deel van het gezondheidsspectrum (Cella et al., 2007, 2010). Ze hebben hiervoor tientallen item banken ontwikkeld die onder andere afgenomen kunnen worden als computergestuurde adaptieve test (CAT). CAT-instrumenten selecteren nieuwe items uit een set items (de item bank), meestal op zo een manier dat het volgende item het meest informatief is om het construct met een zo hoog mogelijke betrouwbaarheid te meten. Het selecteren van nieuwe items stopt zodra de meting een vooraf ingestelde betrouwbaarheid heeft bereikt. Met deze methodiek zou de clinicus niet alleen de *gewenste* informatie voor de behandeling kunnen verzamelen omdat de meetinstrumenten efficiënt zijn, maar in combinatie met de PROMIS item banken ook *kwalitatieve* informatie omdat de meetinstrumenten betrouwbaar, valide en responsief zijn.

In 2009 is de Nederlands-Vlaamse (NV) PROMIS groep opgericht om te onderzoeken of PROMIS ook geïmplementeerd kan worden in Nederland en België. Als startpunt hiervoor vertaalden ze 17 van de PROMIS item banken voor volwassenen (Terwee et al., 2014) en 9 van de PROMIS item banken voor kinderen (Haverman et al., 2016). Voor het meten van generieke depressie en angst bij volwassenen vertaalden ze de PROMIS v1.0 item banken voor Depressie en Angst. Verwacht wordt op basis van Amerikaans onderzoek dat CAT-afnames van deze item banken depressie en angst efficiënt, betrouwbaar, valide en responsief zullen meten in de Nederlandse en Vlaamse bevolking (Kroenke, Baye, & Lourens, 2019; Pilkonis et al., 2011, 2014; Schalet et al., 2016). Daarnaast wordt ook verwacht dat de PROMIS CATs efficiënter en betrouwbaarder zullen zijn dan andere meetinstrumenten (Pilkonis et al., 2014). In dit proefschrift worden deze verwachtingen onderzocht voor de algemene Nederlandse bevolking

en voor Nederlandse patiënten die worden behandeld voor veelvoorkomende, psychische stoornissen.

Hoofdstuk 2: Het CAT potentieel evalueren met de MASQ

In Hoofdstuk 2 is eerst het *potentieel* van CAT onderzocht met simulatieonderzoek. Hiervoor is gebruikt gemaakt van de drie subschalen van de Mood and Anxiety Symptom Questionnaire (MASQ; Watson & Clark, 1991) en een bestaande dataset van patiënten die zijn behandeld door de zorgaanbieder Rivierduinen ($N = 3.597$). Het CAT-potentieel is vervolgens onderzocht met behulp van een psychometrische evaluatie en meerdere post hoc CAT-simulaties. Een post hoc CAT-simulatie is geen echte CAT-afname, maar gebruikt de antwoorden op de items uit de volledige item bank en evalueert deze alsof ze adaptief zijn verzameld. Eerder onderzoek heeft laten zien dat de uitkomsten van post hoc CAT-simulaties veel lijken op die van echte CAT-afnames (Kocalevent et al., 2009).

De resultaten lieten zien dat de drie subschalen van de MASQ gunstige psychometrische eigenschappen hebben voor een CAT-afname. Alleen voor de somatische angst schaal werd een item verwijderd vanwege de grote gelijkensis met een ander item. De CAT-simulaties lieten vervolgens voor alle MASQ schalen zien dat het aantal items aanzienlijk verminderd kan worden met een minimaal verlies aan betrouwbaarheid. Bij een acceptabele betrouwbaarheidsstandaard voor individuele metingen (standaard error [SE] $\leq 0,30$) varieerde de percentuele item afname van 74% tot 56%. Bovendien bleven de scores en de voorspellende diagnostische waarde van de gesimuleerde CATs vergelijkbaar met die van de volledige MASQ schalen. Deze resultaten suggereren dat een CAT-versie van de MASQ kan leiden tot substantieel efficiëntere metingen met behoud van gunstige psychometrische eigenschappen.

Hoofdstuk 3: Psychometrische eigenschappen van de Nederlands-Vlaamse

PROMIS Depressie item bank

In Hoofdstuk 3 is de PROMIS Depressie item bank onderzocht om een valide set items vast te stellen voor CAT-afnames (Reeve et al., 2007). Als startpunt hiervoor zijn de 28 items uit de NV PROMIS v1.0 volwassene item bank voor Depressie aangevuld met 28 vertaalde items die de VS versie net niet hebben gehaald door ongunstige psychometrische eigenschappen. Deze item set is vervolgens psychometrisch geëvalueerd met behulp van een gecombineerde algemene populatiesteekproef ($n = 1.002$) en een klinische steekproef van patiënten die zijn behandeld door de zorgaanbieder Parnassia Groep ($n = 1.008$). Op basis van de resultaten zijn acht van de aanvullende items verwijderd. De overgebleven item set is ten slotte gebruikt als input voor meerdere post hoc CAT-simulaties om de uitgebreide item bank (48 items) te vergelijken met de originele item bank (28 items). De resultaten suggereerden dat beide item banken geschikt zijn om depressie efficiënt, betrouwbaar en valide te meten met CAT. Voor de rest van het proefschrift is daarom gebruikt gemaakt van de originele PROMIS item bank omdat het gebruik hiervan internationale vergelijkingen mogelijk maakt. De CAT-simulaties voor deze item bank selecteerden voor de klinische steekproef gemiddeld 8,40 items bij een hoge

betrouwbaarheidsstandaard voor individuele metingen ($SE \leq 0,20$) en gemiddeld 3,40 items bij een acceptabele betrouwbaarheidsstandaard voor individuele metingen ($SE \leq 0,30$).

Hoofdstuk 4: Psychometrische eigenschappen van de Nederlands-Vlaamse

PROMIS Angst item bank

In Hoofdstuk 4 is de PROMIS Angst item bank onderzocht om een valide set items vast te stellen voor CAT-afnames. In deze studie is direct gekozen voor het gebruik van de items uit de originele NV PROMIS v1.0 volwassene item bank voor Angst omdat een pre-analyse van de data vergelijkbare resultaten liet zien als in Hoofdstuk 3: de efficiëntie, betrouwbaarheid en validiteit van een gesimuleerde CAT verschilt nauwelijks tussen de originele item bank en een uitgebreidere variant. Opnieuw werd de gecombineerde algemene en klinische populatiesteekproef gebruikt om aan te tonen dat ook de PROMIS Angst item bank gunstige psychometrische eigenschappen heeft voor een CAT-afname ($N = 2.010$). Daarnaast lieten verschillende post hoc CAT-simulaties zien dat angst efficiënt, betrouwbaar en valide gemeten kan worden met een CAT-afname van de item bank. Bij een hoge betrouwbaarheidsstandaard voor individuele metingen ($SE \leq 0,22$) werden gemiddeld 8,64 items geselecteerd voor de klinische steekproef en 9,48 items voor de Nederlandse bevolkingssteekproef. Werd deze betrouwbaarheidsstandaard verlaagd naar acceptabel voor individuele metingen ($SE \leq 0,32$), dan daalde het aantal geselecteerde items zelfs naar gemiddeld 4,25 voor de klinische steekproef en gemiddeld 6,06 items voor de Nederlandse bevolkingssteekproef.

Hoofdstuk 5: Longitudinale meetinvariantie in de PROMIS item banken

In hoofdstuk 5 zijn de NV PROMIS v1.0 volwassene item banken voor Depressie en Angst onderzocht op het validiteitsaspect longitudinale meetinvariantie (LMI; Fokkema et al., 2013; Fried et al., 2016). Hiervoor is gebruikt gemaakt van begin- en tussenmetingen, afgenomen bij patiënten die zijn behandeld door de zorgaanbieder Parnassia Groep voor stemmingsstoornissen ($N = 640$) en angststoornissen ($N = 528$). Met factoranalyse is vervolgens onderzocht of de item banken voldoende unidimensioneel zijn op beide meetmomenten en de gemeten constructen niet substantieel veranderen over tijd (Liu et al., 2017; Liu & West, 2018). De resultaten suggereerden dat de item banken voldoende unidimensioneel zijn op beide meetmomenten, maar de depressie en angst constructen veranderden wel over tijd. Op basis van voorlopige criteria lieten vervolganalyses echter zien dat de invariantieschendingen geen substantieel effect hebben op de antwoorden van patiënten en hun uiteindelijke scores. Hieruit kan geconcludeerd worden dat de invariantieschendingen waarschijnlijk van weinig betekenis zijn voor de klinische praktijk. Met andere woorden: de NV PROMIS v1.0 volwassene item banken voor Depressie en Angst kunnen beschouwd worden als voldoende invariant over tijd.

Hoofdstuk 6: Psychometrische eigenschappen van de PROMIS CATs voor Depressie en Angst

In Hoofdstuk 6 zijn echte CAT-afnames van de NV PROMIS volwassene v1.0 item banken voor Depressie en Angst vergeleken met de subschalen van de Brief Symptom Inventory (BSI; de Beurs & Zitman, 2005). Hiervoor is gebruikt gemaakt van begin- en tussenmetingen, afgenomen bij patiënten die zijn behandeld door de zorgaanbieder Dimence Groep ($N = 400$). Met verschillende analyses is vervolgens de validiteit en de responsiviteit van de PROMIS CATs onderzocht en zijn twee indicatoren voor verandering geëvalueerd. De resultaten suggereerden dat de PROMIS CATs vergelijkbare constructen meten als de BSI schalen voor Depressie en Angst. Onder deze aanname werd ook aangetoond dat de CAT Depressie volgens verwachting ongeveer even responsief is als de BSI Depressie schaal. De CAT Angst daarentegen was tegen de verwachting in substantieel responsiever dan de BSI Angst schaal. Ten slotte categoriseerden beide PROMIS CATs meer patiënten als daadwerkelijk veranderd dan de overeenkomende BSI schalen. Op basis hiervan kan geconcludeerd worden dat de PROMIS CATs voor Depressie en Angst een bescheiden verbetering zijn ten opzichte van de BSI schalen voor Depressie en Angst. De CAT-instrumenten zijn mogelijk beter in staat om daadwerkelijke verandering te detecteren, waarschijnlijk door een hogere betrouwbaarheid.

Hoofdstuk 7: Discussie

In dit proefschrift zijn de NV PROMIS volwassene v1.0 item banken voor Depressie en Angst psychometrisch geëvalueerd voor CAT-afnames in de algemene Nederlandse bevolking en de Nederlandse klinische populatie (d.w.z., patiënten die worden behandeld voor veelvoorkomende, psychische stoornissen). De resultaten suggereerden dat beide item banken zeer informatief zijn voor het meten van depressie en angst bij een groot deel van de Nederlandse (klinische) populatie. Dit werd bevestigd door zowel post hoc CAT-simulaties als echte CAT-afnames, die beide lieten zien dat depressie en angst betrouwbaar en efficiënt gemeten kunnen worden. Daarnaast werd aangetoond dat de CAT-instrumenten valide zijn voor het meten van depressie en angst in de Nederlandse (klinische) populatie, en responsief voor het meten van veranderingen in depressie en angst in de klinische populatie. Op basis van deze resultaten kan worden geconcludeerd dat het hoofddoel van dit proefschrift voldoende is bereikt: CAT-versies van de DF PROMIS volwassene v1.0 item banken voor Depressie en Angst zijn efficiënt, betrouwbaar, valide en responsief voor het meten van depressie en angst in de Nederlandse (klinische) populatie.

In de PROMIS CAT-studies vielen voornamelijk twee bevindingen op. Ten eerste lieten de PROMIS CATs slechts bescheiden verbeteringen zien ten opzichte van de BSI. Het lijkt daarom misschien niet erg aantrekkelijk voor klinici om de overstap te maken naar PROMIS CATs, zeker gezien het feit dat gebruikers moeten wennen aan nieuwe meetinstrumenten, wat mogelijk als last ervaren kan worden. Hierbij moet echter opgemerkt worden dat de PROMIS CATs (mogelijk) extra voordelen hebben vergeleken met andere meetinstrumenten. Ten eerste zijn PROMIS instrumenten ontwikkeld voor een universele toepassing (Belekkas et al., 2018; Lizzio et al., 2019; Papuga et al., 2018; Scholle et al., 2018; Wagner et al., 2015), terwijl veel

andere instrumenten vaak primair zijn ontwikkeld voor het meten van klinische populaties. Ten tweede kunnen PROMIS scores wellicht vergeleken worden tussen verschillende landen om van elkaar te leren (Elsman, Flens, de Beurs, Roorda, Terwee, 2022; Terwee et al., 2021; van Bebber et al., 2018), terwijl dit voor andere instrumenten onduidelijker is. Ten derde hebben PROMIS gebruikers toegang tot tal van andere PROMIS instrumenten voor het meten van een groot deel van het gezondheidsspectrum (bijv. Crins et al., 2015, 2016, 2017; Terwee et al., 2019), waardoor een set meetinstrumenten voor behandel-evaluatie flexibel samen te stellen is. Ten vierde is in dit proefschrift aangetoond dat de NV PROMIS volwassene v1.0 item banken voor Depressie en Angst voldoende invariant zijn over tijd, terwijl dit nog onduidelijk is voor veel andere Nederlandse instrumenten die mentale gezondheidsconstructen meten. Ten slotte zullen de voordelen van de PROMIS CATs waarschijnlijk groter zijn vergeleken met andere meetinstrumenten. In hoofdstuk 2 is bijvoorbeeld aangetoond dat CAT-versies van de drie MASQ schalen een item afname laten zien van 74% tot 56% met een minimaal verlies aan betrouwbaarheid.

De tweede opvallende bevinding is het verschil tussen de PROMIS CATs met betrekking tot de onderzochte psychometrische eigenschappen. Voor beide meetinstrumenten zijn adequate psychometrische eigenschappen aangetoond, maar de CAT Angst presteerde net wat minder goed dan verwacht. Dit bleek voor zowel de efficiëntie, de betrouwbaarheid, de validiteit als de responsiviteit van het instrument. Op basis van deze resultaten kan het zinnig zijn om na te gaan of de psychometrische eigenschappen van de PROMIS CAT Angst nog verder verbeterd kunnen worden. Het kan bijvoorbeeld zijn dat het gemeten Angst construct beter onderverdeeld kan worden in meerdere subdomeinen, waardoor het meten van angst wellicht geschikter is voor een multidimensionale CAT (MCAT; Smits, Paap, & Böhnke, 2018). Misschien is het echter ook wel afdoende om een betere balans te realiseren in het aantal items waarmee elk subdomein wordt gemeten. Nader onderzoek moet hier uitsluitsel over geven.

De CAT-versies van de NV PROMIS volwassene v1.0 item banken voor Depressie en Angst kunnen vanaf nu geïmplementeerd worden in de klinische praktijk. Voor een succesvolle implementatie zijn verschillende aandachtspunten van belang. Voorbeelden hiervan zijn het vergroten van de beschikbaarheid en toegankelijkheid van CAT-technologie voor zorgaanbieders, het vergroten van kennis over CAT-instrumenten bij klinici, en het vergroten van de beschikbaarheid van gebruiksvriendelijke interfaces en feedbacktools. Misschien wel het meest dringende aandachtspunt is dat zorgaanbieders überhaupt gebruik gaan maken van meetinstrumenten als hulpmiddel voor het evalueren van de behandeling. In de Nederlandse geestelijke gezondheidszorg staat dit gebruik nog steeds in de kinderschoenen, ondanks dat tal van studies gunstige effecten hebben aangetoond op de algehele kwaliteit van zorg (de Jong et al., 2021; Fortney et al. 2017; Guo et al., 2015; Rush & Thase, 2018; Scott & Lewis, 2015). De belangrijkste reden hiervoor is dat een succesvolle implementatie van meetinstrument-gerichte zorg verschillende uitdagingen met zich meebrengt (Martin-Cook et al., 2021). Slechts één van deze uitdagingen is de beschikbaarheid van korte en betrouwbare zelfrapportage-meetinstrumenten zoals de PROMIS CATs voor Depressie en Angst. Andere voorbeelden zijn het tonen van actief leiderschap bij de implementatie, een actieve betrokkenheid van de clinicus

en patiënt, en de afstemming van het meetinstrumentarium op de aandoeningen en behandeldoelen van patiënten.

Terwijl de implementatie van CAT-instrumenten in de klinische praktijk opgestart wordt, kan de ontwikkeling van een volgende generatie CAT-instrumenten in gang worden gezet. Enkele voorbeelden van toekomstige onderzoeksrichtingen zijn het verbreden van het CAT-aanbod, het ontwikkelen van MCATs en het verbeteren van de stopregel van het CAT-algoritme. Door de CAT-instrumenten verder te verbeteren, wordt de kans vergroot dat dit moderne meetinstrumentarium uiteindelijk de nieuwe standaard wordt voor het evalueren van behandelingen in Nederland. Dit kan het gebruik van meetinstrumenten in de behandelkamer verder stimuleren, wat weer kan gaan leiden tot effectievere en efficiëntere behandelingen in het algemeen.

