

Hoe vergelijkbaar is de Net Promoter Score over landen heen?

Friso Janmaat en Hester van Herk

SAMENVATTING Het is gebruikelijk dat managers scores van klanttevredenheidsonderzoeken vergelijken tussen vestigingen om na te gaan welke vestiging de hoogste score heeft behaald. Doordat steeds meer organisaties de klant centraal stellen binnen hun organisatie, worden Key Performance Indicators (KPI's) zoals de Net Promoter Score (NPS), steeds belangrijker. De NPS wordt vaak opgenomen in prestatiebeoordelingen van vestigingen in binnen- en buitenland. De vraag is echter of scores die zijn vergaard onder klanten in verschillende landen wel met elkaar kunnen worden vergeleken. Dit onderzoek laat zien dat dit aan strenge restricties is gebonden.

RELEVANTIE VOOR DE PRAKTIJK Grote accountantsorganisaties zoals KPMG, PwC en BDO werken met de NPS, maar ook multinationals zoals Aegon, KPN, Philips en ING hebben de NPS in hun jaarverslagen opgenomen. In sommige gevallen zijn er ook prestatieafspraken verbonden aan deze beroemde KPI, waar zelfs de Autoriteit Financiële Markten zich onder voorwaarden positief over uitsprekt. Ons onderzoek toont aan dat het behalen van een hoge NPS gemakkelijker of moeilijker is afhankelijk van het land waar de NPS wordt gemeten.

1 Inleiding

Tegenwoordig maken veel marketeers gebruik van verschillende maatstaven om het effect van hun marketinginspanningen te meten. Scores op zulke maatstaven worden vaak verzameld onder klanten van de onderneming door middel van een enquête of survey. De Net Promoter Score (zie kader 1) is een veelgebruikte maatstaf (Hoekstra & Leeflang, 2010). Ondanks alle kritiek vanuit de wetenschap (o.a. Polloack & Alexandrov, 2013) op de NPS is deze maatstaf in de praktijk door een groot aantal multinationals opgenomen in het jaarverslag.

De NPS-theorie gaat uit van de gedachte dat klanten die positief scoren op de maatstaf in de toekomst klant zullen blijven. Daarnaast zullen nieuwe klanten worden aangetrokken door positieve mond-tot-mond reclame van bestaande klanten. Op basis hiervan wordt

geconcludeerd dat hoe hoger de Net Promoter Score, des te groter de kans dat de onderneming een positieve financiële groei in de toekomst zal doormaken (Reichheld, 2003). Onderzoek wijst echter uit dat de NPS niet beter of slechter is in het voorspellen van toekomstige financiële groei ten opzichte van andere metrics, zoals klanttevredenheid of loyaliteitintenties (Van Doorn, Leeflang & Tijs, 2012). Ondanks deze wetenschappelijke inzichten gaan organisaties door met het implementeren van de NPS. De kritiek op de NPS en het gebruik in internationale context roept de vraag op hoe bruikbaar de NPS is bij vergelijking tussen landen.

Internationaal opererende organisaties die voor de NPS-methode kiezen worden geconfronteerd met nieuwe uitdagingen. Zij willen graag de Net promoter Scores (NPSs) met elkaar vergelijken en hebben vragen als: Welke nationale vestiging scoort het beste? Wie heeft de hoogste en wie de laagste score? ING vermeldt bijvoorbeeld in haar jaarverslag (2012) dat zij in tien verschillende landen de (op één na) hoogste NPS heeft in haar branche. Wanneer de NPS is verzameld in verschillende landen kan dit grote invloed hebben op de validiteit van de vergelijking. Dit komt doordat mensen uit diverse landen verschillen in de manier waarop zij vragen gemeten met antwoordschalen beantwoorden (Van Herk, Poortinga & Verhallen, 2004). De aangekruiste antwoordcategorie is niet alleen bepaald door de inhoud van de vraag, maar ook door de manier waarop men gewend is vragen in te vullen.

Het volgende, aan het Forresterblog ontleende citaat illustreert de verwachte verschillen bij het beantwoorden van de NPS-vraag in internationale context: *“Onlangs gaf een Europese klant mij een NPS beoordeling. Bij de beantwoording van de vraag zei hij tegen mij: ‘Een tien is voor God, een negen krijgt alleen El Maestro, dus een acht is het hoogst haalbare voor jou’. Ik was in shock! De klant was helemaal tevreden, desondanks was ik volgens NPS begrippen gewoon niet goed genoeg.”*

Een ander voorbeeld wordt aangehaald op Customergauge.com. Hier wordt besproken of er zoiets bestaat als het *“Dutch Effect”*, wat inhoudt dat Nederlanders van nature niet graag een 10 geven. Het effect blijkt

niet alleen voor Nederland te gelden. Inwoners van de meeste Noordwest Europese landen lijken terughoudend in het geven van tienens (Van Herk, Poortinga & Verhallen, 2004).

Het belangrijkste doel van dit onderzoek is het beantwoorden van de vraag: “hoe kunnen NPSs worden vergeleken wanneer ze verzameld zijn in verschillende landen?” In het theoretisch raamwerk zal worden besproken dat respondenten (in meer of mindere mate) op een systematische manier reageren op uitspraken. Een dergelijke systematiek wordt aangeduid met de term antwoordstijl. Allereerst wordt uitgelegd wat een antwoordstijl precies is. Vervolgens zullen verschillende typen antwoordstijlen worden toegelicht. Daarna wordt er besproken welke individuele en landkenmerken antwoordstijlen beïnvloeden. In paragraaf 3 volgt de beschrijving van de door ons gebruikte data en de analyse, waarna de resultaten worden besproken. In paragraaf 5 besteden we aandacht aan de mogelijke effecten van antwoordstijlen op de NPS. We sluiten af met een conclusie waarin de belangrijkste bevindingen worden weergegeven.

2 Antwoordstijlen

2.1 Effecten van antwoordstijlen

Vergelijkbaarheid van gegevens is nodig wanneer er onderzoek wordt gedaan in verschillende landen of in subgroepen met een verschillende (culturele) achtergrond. De vergelijkbaarheid van gegevens kan worden omschreven als de systematische detectie, identificatie en interpretatie van overeenkomsten en verschillen tussen mensen of landen (Van Herk et al., 2004). In dit artikel definiëren wij vergelijkbaarheid van gegevens als de mate waarin een schaal of item dezelfde betekenis heeft en op dezelfde manier wordt geïnterpreteerd door alle respondenten in het onderzoek.

Wanneer respondenten vragen in enquêtes invullen kunnen hun antwoorden worden beschouwd als de werkelijke mening over het onderwerp. Echter, naast deze mening bevat het antwoord van de respondent op de antwoordschaal ook zogenaamde content-irrelevante factoren, de antwoordstijlen. De quote uit de inleiding illustreert dit, in Nederland geeft een respondent nooit een negen of een tien ongeacht het onderwerp van de vraag of stelling.

Samengevat, wanneer een respondent de neiging heeft om op een systematische manier positief (of negatief) op vragen/stellingen te antwoorden, op basis van iets anders dan dat wat de vragen zouden moeten meten, bestaat er een zeer grote kans dat dit een antwoordstijl is (Baumgartner & Steenkamp, 2001).

Zulke content-irrelevante factoren hebben uiteraard ook hun uitwerking op een NPS-meting.

Wanneer klanten uit een bepaalde subgroep bijvoor-

Kader 1 De Net Promoter Score, wat is dat eigenlijk?

“Hoe waarschijnlijk is het dat U bedrijf X zal aanbevelen bij een vriend of collega?”. Dit is de NPS- vraag die vaak wordt beantwoord op een elf-puntsschaal (van 0 t/m 10).

De NPS wordt als volgt berekend: alle klanten die participeren in het onderzoek worden onderverdeeld in drie categorieën: de Criticasters, de Passief Loyalen en de Promoters.

Promoters: Dit zijn klanten die de NPS-vraag met een 9 of een 10 hebben beantwoord. Deze klanten zijn lyrisch over het bedrijf. De verwachting is dat deze klanten in de toekomst klant zullen blijven. Daarnaast is het zeer waarschijnlijk dat zij de organisatie zullen aanbevelen in hun netwerk.

Passief Loyalen: Dit is een groep klanten die tevreden is over bedrijf X. Zij hebben de NPS-vraag met een 7 of een 8 beantwoord. De kans dat zij klant zullen blijven is groot, maar ze zullen geen aanbevelingen doen binnen hun netwerk.

Criticasters: Klanten in deze categorie zijn niet tevreden over bedrijf X. Zij hebben de vraag met een 6 of lager beantwoord. De kans bestaat dat zij in de toekomst klant zullen blijven, echter zodra er een alternatief op de markt komt is de kans groot dat ze zullen overstappen.

De berekening $\% \text{ Promoters} - \% \text{ Criticasters} = \text{NPS}$

De Net Promoter Score wordt berekend door het percentage Criticasters van het percentage Promoters af te trekken. (NB: Er zijn dus ook bedrijven met een negatieve NPS.)

Waarom?

Na het stellen van de NPS-vraag volgt vaak de eenvoudige vervolgvraag: “Waarom?”. Deze vraag is eigenlijk net zo belangrijk als de aanbevelingsvraag. Na beantwoording van deze vraag weet de organisatie immers waar haar sterke punten liggen, maar ook wat de punten van aandacht zijn per klantloyaliteitscategorie (Reichheld & Markey, 2011).

beeld snel geneigd zijn om de positieve antwoordcategorieën op de schaal te kiezen, ongeacht de inhoud van de vraag, zal het gemiddelde op de NPS automatisch hoger liggen. Daarnaast kan het zo zijn dat mensen de neiging hebben om juist de extreem positieve of negatieve antwoordcategorieën te kiezen, ongeacht de inhoud van de vraag. In een dergelijk geval zal de standaarddeviatie hoger zijn dan van een groep klanten die juist de neiging heeft om antwoordcategorieën rond het middelpunt te gebruiken. Een neiging om de meest extreme antwoorden te kiezen kan leiden tot een erg hoge of juist erg lage NPS in vergelijking met een land waarin men juist extreme antwoorden vermijdt.

2.2 Verschillende typen antwoordstijlen

In dit onderzoek zullen de twee meest voorkomende antwoordstijlen worden meegenomen, namelijk: Acquiescence Response Style (ARS) en Extreme Response Style (ERS) (Paulhus, 1991). In tabel 1 is een illustratie van ARS en ERS weergegeven. ARS kan worden omschreven als de neiging om het (helemaal) eens te zijn met een uitspraak ongeacht de inhoud van de vraag of

stelling. Bij ARS worden ook negatief geformuleerde items positief beantwoord (Baumgartner & Steenkamp, 2001). Een hoge score op ARS betekent dat de respondent een sterke neiging heeft tot het overmatig gebruik van de meest positieve antwoordcategorieën. Aangezien onderzoeken veelal gebruik maken van positief geformuleerde items is ARS een vaak voorkomende antwoordstijl (Van Rosmalen, Van Herk & Groenen, 2010). Respondenten die de neiging hebben om automatisch voor de positieve antwoordcategorieën te kiezen (ongeacht de inhoud) en dus ARS vertonen, beoordelen een item waarover zij negatief zijn bijvoorbeeld met een zeven in plaats van een vijf. Uiteraard heeft dit zijn uitwerking op de uiteindelijk berekende NPS.

Naast ARS is ERS een belangrijke antwoordstijl (Paulhus, 1991). Het is een vaak onderzochte antwoordstijl (Baumgartner & Steenkamp, 2001; De Jong, Steenkamp, Fox & Baumgartner, 2008). ERS kan worden omschreven als de neiging om de eindpunten (meest extreme punten) van een antwoordschaal overmatig te gebruiken. Een hoge score op ERS betekent dat de respondent een sterke neiging heeft tot het overmatig gebruik van de meest extreme antwoordcategorieën, bij de NPS de 0 en de 10. Als mensen ontevreden zijn zullen ze het bevroegde onderwerp eerder met een nul beoordelen dan met een vier.

Daarentegen wordt er een tien gegeven wanneer het individu tevreden is, in plaats van bijvoorbeeld een acht. Respondenten die de neiging hebben tot het gebruik van ERS, zullen daardoor sneller in de hoogste of laagste NPS-categorie vallen en minder snel in de middelste categorie.

Tabel 1 Visualisatie van ARS en ERS. De donkergekleurde antwoordcategorieën worden overmatig gebruikt ongeacht de inhoud van de vraag.

ARS – Acquiescence: de neiging om de meest positieve antwoordcategorieën overmatig te gebruiken.									
Helemaal mee oneens									Helemaal mee eens
ERS – Extreme Response Style: de neiging om de meest extreme antwoordcategorieën overmatig te gebruiken.									
Helemaal mee oneens									Helemaal mee eens

De vraag die vervolgens wordt gesteld is welke mensen meer, dan wel minder, geneigd zijn antwoordstijlen te vertonen. Op individueel niveau hangt dit onder andere samen met demografische factoren. Er is bijvoorbeeld gevonden dat oude, laagopgeleide respondenten een veel sterkere neiging hebben om positieve antwoor-

den te geven, dan jonge, hoogopgeleide respondenten (Knauper, 1999). Dit maakt het voor een vestiging, gelokaliseerd in een yuppenbuurt van Amsterdam, lastiger om hoog op klant-KPI's te scoren in vergelijking met een vestiging in een klein dorp. Het gebruik van antwoordstijlen kan ook door factoren op landniveau worden beïnvloed, bijvoorbeeld door cultuur. Doordat antwoordstijlen invloed op de uitkomsten van onderzoek hebben is het noodzakelijk om hiervoor te corrigeren (zie bijvoorbeeld De Jong et al., 2010). De antwoordstijl ARS kan er bijvoorbeeld toe leiden dat de NPS veel positiever uitvalt, onafhankelijk van de daadwerkelijke mening van de klant.

2.3 Antecedenten van antwoordstijlen

Demografische factoren en de culturele en economische omgeving waarin mensen leven zijn van invloed op de wijze waarop respondenten vragen beantwoorden; deze factoren beïnvloeden de mate waarin mensen antwoordstijlen vertonen.

2.3.1 Demografische kenmerken

Verschillende onderzoeken naar antwoordstijlen hebben aangetoond dat ouderen eerder geneigd zijn tot ARS dan jongeren (Johnson, et al., 2005; Harzing, 2006) en dat opleiding negatief samenhangt met het gebruik van antwoordstijlen, waarbij hoger opgeleiden minder ARS en ERS vertonen (Marin et al., 1992; Greenleaf, 1992). Een reden hiervoor zou kunnen zijn dat hoger opgeleide respondenten geneigd zijn om stellingen en vragen intensiever te evalueren (Weijters, Geuens & Schillewaert, 2010). Het geslacht van de respondent is niet eenduidig gerelateerd aan ARS en ERS; sommige onderzoeken vonden meer ARS bij vrouwen en andere bij mannen. Voor de NPS zouden antwoordstijlen kunnen betekenen dat vestigingen van bijvoorbeeld zakelijke dienstverleners met relatief veel jonge en/of hoogopgeleide klanten hierdoor automatisch een kleinere kans op hoge NPSs hebben.

2.3.2 Landkenmerken

In het afgelopen decennium is meer onderzoek gedaan naar verschillen in antwoordstijlen tussen mensen uit verschillende landen. In deze onderzoeken is steeds een beperkt aantal landen met elkaar vergeleken, maar er zijn wel tendensen waar te nemen. Er is bijvoorbeeld gevonden dat Grieken veel meer geneigd zijn tot ARS dan Noord-Europeanen (Van Herk et al., 2004) en dat mensen in Azië minder ERS vertonen dan mensen in de USA (Chen, Lee & Stevenson, 1995; De Jong et al., 2010). Een belangrijke verklaring voor de gevonden verschillen in antwoordstijlen is cultuur en dan specifiek de Hofstede-dimensies individualisme versus collectivisme (IND) en onzekerheidsvermijding (Hofstede, 2001). In collectivistische culturen wordt het bewaren van de harmonie tussen mensen belangrijk

gevonden evenals het conformeren aan de groep. Dit leidt tot meer positieve antwoorden, omdat mensen harmonie prefereren (Harzing, 2006). In culturen die hoog scoren op onzekerheidsvermijding hebben mensen een sterkere behoefte aan structuur en regels en houden mensen niet van ambiguïteit en dubbelzinnigheid. Griekenland scoort hoog op onzekerheidsvermijding en het Verenigd Koninkrijk laag, wat de hogere score op ERS en ARS verklaart in Griekenland (Van Herk et al., 2004; Baumgartner & Steenkamp, 2001). Een andere belangrijke verklaring voor verschillen in ERS en ARS lijkt het Bruto Nationaal Product per hoofd van de bevolking. Hoe lager het BNP per hoofd van de bevolking is des te meer ARS en ERS er voorkomt (Johnson et al., 2005). Uit het bovenstaande zijn geen wereldwijd geldende regels af te leiden, wel is duidelijk dat culturele en economische factoren van invloed zijn op antwoordstijlen.

2.4 Antwoordstijlen en de NPS

Mensen verschillen in de mate waarin zij een specifieke antwoordstijl gebruiken. Dit komt doordat zowel demografische kenmerken van het individu als de culturele en economische kenmerken van het land invloed hebben op de beantwoording van enquêtevragen. Het is aannemelijk dat antwoordstijlen invloed hebben op de wijze waarop de NPS-vraag wordt beantwoord. Mensen die een sterke neiging hebben tot ARS kunnen hierdoor sneller in een hoge NPS-categorie vallen. Mensen die een sterke neiging tot ERS hebben zullen sneller in de Promoter- of Criticaster-groep vallen.

In ons onderzoek is de NPS zelf niet gemeten. Door in een groot aantal landen vragen te bestuderen die gemeten zijn op 10-punsschalen wordt inzicht verkregen in de antwoordstijlen van de mensen in die landen. De verschillen in de antwoordstijlen geven inzicht in hoe de NPS in die landen systematisch hoger of lager zou kunnen uitvallen in vergelijking met bijvoorbeeld Nederland.

3 Empirisch onderzoek

3.1 Databeschrijving

Wij gebruiken in dit onderzoek vragen die gemeten zijn op een met de NPS vergelijkbare meetschaal. Een geschikt en in het publieke domein beschikbaar bestand dat wij gebruiken in dit onderzoek is van de World Values Survey (WVS). Wij gebruiken de data van Wave 5, die tussen 2005 en 2007 is verzameld. Het WVS-bestand bevat representatieve steekproeven uit 57 landen wereldwijd. Van de respondenten zijn achtergrondkenmerken zoals geslacht, leeftijd en opleiding (aantal jaren dat men op school is geweest) bekend. Aan de landen in de WVS hebben wij gegevens over cultuur (Hofstede, 2001) en BNP per hoofd van de bevolking in 2007 (Worldbank, 2014) toegevoegd.

Een basisvoorwaarde om antwoordstijlen te bepalen is dat een dataset veel heterogene items (vragen die laag met elkaar zijn gecorreleerd) bevat (Baumgartner & Steenkamp, 2001). De WVS voldoet aan deze voorwaarde, zo is er een vraag over inkomensgelijkheid meegenomen, maar ook de vraag of criminelen wel hard genoeg worden gestraft. Deze heterogene items maken de dataset zeer geschikt voor het bepalen van het voorkomen van antwoordstijlen binnen landen en de effecten ervan op variabelen die gemeten zijn op een vergelijkbare manier als de NPS. Om tot de uiteindelijke set items voor onze analyse te komen hebben wij een aantal criteria opgesteld waaraan de items moeten voldoen. Allereerst moet een item (om bruikbaar te zijn) in Nederland zijn beantwoord. Dit maakt het mogelijk om Nederland met de rest van de wereld te vergelijken. Daarnaast moeten de items zijn beantwoord op een 10-puntsschaal, om de meest gebruikte NPS-schaal zo dicht mogelijk te benaderen. Dit resulteerde in totaal 36 geschikte items (32 items om de antwoordstijlen te bepalen en 4 om te valideren). De gemiddelde inter-item correlatie in deze set van 36 items is 0,067, hetgeen duidt op een heterogene set items die bruikbaar is voor het bepalen van antwoordstijlen (Van Herk et al., 2004). Het aantal landen waarin al deze items zijn meegenomen is 47 (zie grafiek 1 voor een overzicht).

In totaal zijn er na het verwijderen van respondenten met missing data op deze 36 items 30.662 respondenten over. Deze respondenten worden meegenomen voor de volgende analyses. De kleinste steekproef is van Bulgarije ($N = 227$) en de grootste van Zuid-Afrika ($N = 1420$), er zijn 657 Nederlanders in de dataset.

3.2 Data-analyse

De eerste stap bij de analyses is het berekenen van de antwoordstijlen ARS en ERS voor elk individu. Voor het berekenen van de antwoordstijlen hebben we de methode gehanteerd van Baumgartner en Steenkamp (2001). Bij deze methode worden voor ARS de twee meest positieve antwoordcategorieën (10 en 9) gebruikt, waarbij de hoogste categorie het gewicht 2 en de op een na hoogste het gewicht 1 krijgt. Voor ERS worden de twee extreme antwoordcategorieën (1 en 10) gebruikt, beide categorieën ontvangen het gewicht 1. Door deze methode liggen de uitkomsten voor iedere individuele respondent voor ARS tussen 0 en 2 en voor ERS tussen 0 en 1. Hoe hoger de score, des te sterker de neiging van de respondent om in de desbetreffende antwoordstijl te vervallen.

Vervolgens analyseren we de effecten van de demografische factoren en culturele dimensies op de ARS en ERS. Bij deze stap maken wij gebruik van multi-niveau-analyse (Hox & De Leeuw, 1997). Wij gebruiken deze techniek, omdat er in de WVS sprake is van een hiërarchische

structuur. Hiërarchische data kunnen niet met standaardanalysemethoden worden geanalyseerd omdat de steekproef niet enkelvoudig aslect is. Er is sprake van een getrapte clustersteekproef. Mensen in een land delen kenmerken en zijn niet onafhankelijk van elkaar, bijvoorbeeld door gedeelde normen en waarden. Multi-niveau-analyse houdt rekening met deze afhankelijkheid.

Tot slot berekenen we op een viertal items de NPS (voor berekening zie kader 1); ook hier kijken we naar verschillen tussen landen. Deze vier items zijn gekozen omdat er voor elk ervan een extern criterium bestaat: we weten gebruikmakend van andere gegevens hoe landen ten opzichte van elkaar scoren op deze items. Het gaat specifiek om: de tevredenheid van het leven, corruptie binnen landen, gelijkheid tussen mannen en vrouwen en het belang van religie in het dagelijks leven. In paragraaf 4.4 (Correctie Gewenst) zullen we het gebruik van deze items nader toelichten.

4 Resultaten

4.1 ARS: de neiging om positief te antwoorden

De gemiddelde score op ARS is ,711 met een standaarddeviatie van ,363; deze standaarddeviatie duidt op verschillen tussen individuen. Uit de multi-niveau-analyse (PROC MIXED in SPSS) blijkt dat leeftijd een positief effect heeft op het vertonen van ARS ($\gamma = ,003$, $p < .001$), en dat opleiding een negatief effect laat zien ($\gamma = -,002$, $p < .001$). Dit betekent dat een persoon die 60 jaar is een ,09 hogere ARS heeft dan iemand van 30 jaar; als deze persoon ook tien jaar korter op school heeft gezeten gaat er nog ,02 op. Er is geen significant verschil tussen vrouwen en mannen. In totaal wordt er door deze socio-demografische kenmerken 2% van de variantie in ARS verklaard. Dit is in overeenstemming met eerder in de literatuur gevonden effecten.

Op landniveau blijkt dat vooral BNP per hoofd van de bevolking (in duizendtallen) een negatief effect heeft op ARS ($\gamma = -,005$, $p < .001$), hetgeen vergelijkbaar is met eerder in de literatuur gevonden effecten (Johnson et al., 2005). Een vergelijkbaar negatief effect is gevonden voor individualisme, IDV ($\gamma = -,004$, $p < ,01$), dat sterk samenhangt met BNP per hoofd van de bevolking. De cultuurdimensie onzekerheidsvermijding liet geen significante effecten zien. Op landniveau wordt er 27% van de variantie in ARS verklaard. Hieruit kan worden geconcludeerd dat mensen in meer welvarende (veelal individualistische) landen zoals Nederland, Verenigd Koninkrijk en de Verenigde Staten relatief lager zullen scoren op ARS dan mensen in minder welvarende (doorgaans collectivistische) landen zoals landen in Azië of Zuid-Europa. Een stijging in BNP per hoofd van de bevolking van 12.000 tot 22.000 dollar leidt tot een daling van ARS met ,05.

4.2 ERS: de neiging om in extremen te antwoorden

De gemiddelde score op ERS is ,379 met een standaarddeviatie van ,237. ERS is sterk gecorreleerd met ARS ($r = ,91$), hetgeen te verwachten is aangezien alle items in de WVS positief geformuleerd zijn, deze correlatie is in lijn met eerder onderzoek (bijv. Van Herk et al., 2004). Uit de multi-niveau-analyse (PROC MIXED in SPSS) blijkt dat het effect van de demografische factoren op ERS in lijn is met de uitkomsten van ARS. Ook hier zien we dat oudere respondenten eerder geneigd zijn om voor extreme antwoordcategorieën te kiezen dan jongere respondenten ($\gamma = ,001$, $p < .001$). Zowel opleiding als geslacht hebben geen significante invloed op ERS. Op landniveau bleken zowel individualisme als onzekerheidsvermijding geen significant effect te hebben op ERS. BNP per hoofd van de bevolking heeft wel een zwak negatief effect op ERS ($\gamma = -,002$, $p < ,01$), vergelijkbaar met het effect op ARS. In totaal wordt er op individueel niveau 1% variantie verklaard en op landniveau 14%.

De gevonden verschillen in ARS en ERS tussen landen hebben duidelijke implicaties voor de vergelijkbaarheid van NPSs. Hier komen wij op terug in paragraaf 5.

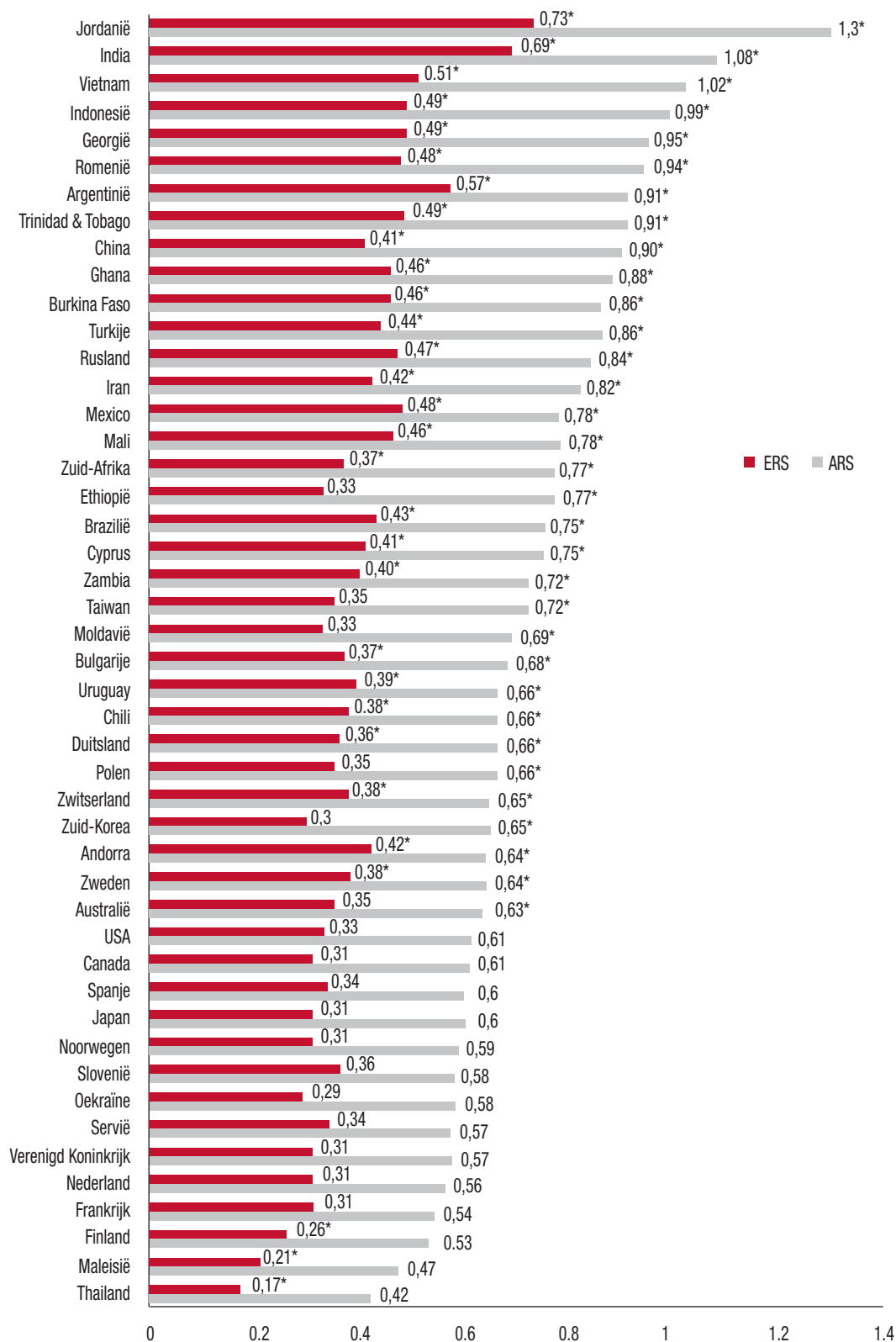
4.3 Verschillen in antwoordstijlen tussen landen nader bekeken

Er is vervolgens een variantieanalyse uitgevoerd om na te gaan hoe de scores op ARS en ERS verschillen over de landen heen. De resultaten laten significante verschillen zien (ARS $F_{(46,30661)} = 171,25$, $p < .001$ en ERS $F_{(46,30661)} = 144,17$, $p < .001$), die duiden op verschillen tussen de landen. De gemiddelde ERS en ARS scores per land zijn in figuur 1 weergegeven.

Thailand (ARS = ,420) en Maleisië (ARS = ,472) hebben de laagste scores op ARS, Nederland scoort ook relatief laag met ARS = ,558. In Europa hebben Nederland, Frankrijk, Finland, Verenigd Koninkrijk, Spanje en Noorwegen de laagste ARS-scores. Deze scores zijn onderling niet significant verschillend van elkaar (significante verschillen zijn aangeduid met een * in de figuur 1). Wanneer er naar regio's wordt gekeken, blijken vooral Noord-Afrikaanse, Zuid- en Centraal-Europese landen bovengemiddeld te scoren op ARS. Landen die het hoogst scoren op ARS zijn Jordanië (1,30) en Indonesië (1,08). Kenmerkend voor de landen met een hoge ARS-score is het relatief lage BNP per hoofd van de bevolking.

De laagste score voor ERS komt ook voor in Thailand (,17) en Maleisië (,21), op de voet gevolgd door Finland (0,26). Op ERS scoort Nederland 0,31, waarmee het een land is dat in de wereldwijde context laag scoort. Op ERS had Jordanië ook de hoogste score (0,73), gevolgd door India (0,69). De hoogste scores op ERS komen voor in Azië en Zuid-Amerika waar deze scores boven ,50 liggen. Dit betekent dat in deze landen door

Figuur 1 Visualisatie van de antwoordstijlen ARS en ERS per land. Conform de definitie ligt de ARS score tussen 0 en 2 en de ERS tussen 0 en 1. Landen met een * verschillen significant met Nederland.



de respondenten bij meer dan de helft van de items een extreme antwoordcategorie (1 en 10) is aangekruist!

Eerder onderzoek waarin alleen westerse landen waren opgenomen (o.a. Harzing 2006; Van Herk et al., 2004), heeft de relatief lage scores op ARS en ERS van Noordwest-Europese landen al laten zien. Echter, in vergelijking met deze landen blijken niet alleen Zuid-Europese, maar ook Afrikaanse en Aziatische landen meer ERS en ARS te vertonen. Dit heeft direct invloed op de NPS: hoe meer ARS, des te groter de kans dat een NPS-responder een Promoter is. Hoe meer ERS, des te vaker een respondent in de Promoter- of Criticaster-categorie valt, hetgeen van invloed is op de berekende score van de NPS (zie kader 1).

4.4 Correctie gewenst

Om na te gaan in hoeverre voor antwoordstijlen gecorrigeerde gegevens overeenkomen met objectieve gegevens, vergelijken wij de gevonden scores op een aantal inhoudelijk geselecteerde items met gegevens uit externe bronnen over hetzelfde onderwerp. Het eerste geselecteerde item, gaat over de tevredenheid over het leven, dit item wordt vergeleken met de externe index World map of Happiness (White, 2006). De tweede testvariabele gaat over corruptie en wordt vergeleken met de internationale Corruption Perception Index (Lambsdorff, 2006). Bij het derde geselecteerde item staat de gelijkheid tussen mannen en vrouwen centraal; dit item wordt vergeleken met de externe index gender equality index: Global Gender Gap Report (Hausmann, Tyson & Zahidi, 2006). Het vierde geselecteerde item heeft betrekking op het belang van religie in het dagelijkse leven. Voor de vergelijkingsprocedure is de Religion Importance map (Gallup, 2009) gebruikt. Alle indices beslaan de periode 2005 en 2007, hetgeen in lijn is met de jaren waarin de data van de WVS zijn verzameld.

De vier items zijn gecorrigeerd voor antwoordstijlen (conform Baumgartner & Steenkamp, 2001) en zijn vervolgens geaggregeerd naar landniveau. De landscores op de oorspronkelijke items en met de voor antwoordstijl gecorrigeerde items zijn vergeleken met de indices. Na correctie voor ARS en ERS veranderde de correlatie tussen de externe indices en de items in positieve zin. De grootste verandering was voor het item over corruptie en de Corruption Perception Index: van 0,30 naar 0,70. Dit duidt erop dat de antwoordstijlen systematische variantie betreffen. De oorspronkelijk hoge score op de vier items is minder sterk gerelateerd aan de werkelijke verschillen tussen de landen. De verschillen in schaalgebruik wijzen op antwoordstijlen en niet op werkelijke verschillen. Correctie van scores is derhalve gewenst.

5 De verwachte NPS

Het uitgangspunt van dit artikel is het kijken naar de

vergelijkbaarheid van Net Promoter Scores over landen heen. De NPS wordt doorgaans gemeten op een 11-puntsschaal, een schaal die wij in dit onderzoek zoveel mogelijk hebben benaderd door uitsluitend items gemeten op tienpuntsschalen te gebruiken. De wijze van berekening van de NPS is vergelijkbaar met de berekening van antwoordstijlen zoals ARS en ERS. Het lijkt dus aannemelijk dat ook de NPS wordt verstoord door antwoordstijlen.

De score op de NPS wordt in dit onderzoek bepaald door het percentage antwoorden in de antwoordcategorieën 1 tot en met 6 af te trekken van het percentage antwoorden in de antwoordcategorieën 9 en 10. Het blijkt dat de NPS correleert met de scores op de antwoordstijlen ARS en ERS ($R^2 = ,88$ voor ARS en $R^2 = ,68$ voor ERS). Veel van de verschillen in scores op een NPS zouden derhalve kunnen worden verklaard door antwoordstijlen.

Wat betekent dit voor de NPS? De door een land gescoorde NPS wordt berekend door het percentage 0 tot en met 6 (= Criticasters), af te trekken van het percentage mensen dat een 9 of 10 gegeven heeft (Promoters), degenen die een 7 of 8 geven worden als Passief Loyaal gezien. Wij voeren deze berekening uit op de items uit onze dataset en kunnen zo de omvang van de drie NPS-groepen berekenen, bijvoorbeeld voor Nederland. Nederland heeft een laag percentage Promoters (34%) en ook een lage score op ARS, terwijl landen met een hoog percentage Promoters (54%) zoals India een hoge ARS-score hebben. Als wordt afgegaan op de scores op de antwoordschaal behoort de meerderheid van de Nederlanders tot de Criticasters (41%), terwijl de meerderheid van de Indiërs (54%) en vooral de Jordaniërs (68%) tot de Promoters behoort.

Om te corrigeren voor de antwoordstijlen zijn eerst NPSs berekend op de vier geselecteerde items (zie paragraaf 4.4). De scores zijn berekend conform de NPS (% hoge scores - % lage scores) en vervolgens ontdaan van de antwoordstijlvariantie door ARS en ERS (op basis van de onafhankelijke set met 32 items). De NPS na correctie voor antwoordstijlen (regressiebenadering, zie Baumgartner & Steenkamp, 2001) is weergegeven in tabel 2.

Dit biedt inzicht in de te verwachten verschillen in NPSs bij gebruik van de maatstaf over landen heen. Nederland is hier gekozen als de norm (NPS-percentages = 0%). Landen die significant van Nederland afwijken (Tukey-B, $p < .001$) zijn gemarkeerd met een *. Wanneer de scores in tabel 2 significant afwijken van Nederland, is directe vergelijking met de scores uit Nederland niet direct te maken. Het verdient dan aanbeveling om te kijken naar de jaarlijkse stijging van de

Tabel 2 Het te corrigeren verschil in NPS bij vergelijking over landen heen. Binnen de regio's zijn de landen geordend van laag naar hoog. Nederland is hier gekozen als referentiepunt (0%).

	Verwachte % NPS verandering		Verwachte % NPS verandering
West Europa		Zuid-Amerika	
Frankrijk	-1%	Uruguay*	5%
Finland	0%	Chili*	5%
Nederland	0%	Brazilië*	10%
Ver. Koninkrijk	1%	Mexico*	10%
Andorra	2%	Argentinië*	16%
Spanje	3%	Trinidad & Tobago*	30%
Noorwegen	3%		
Zweden	4%		
Zwitserland*	5%		
Duitsland*	6%		
Oost-Azië		Zuidoost Europa/Midden Oosten	
Japan	4%	Cyprus*	11%
Zuid Korea*	8%	Turkije*	19%
Taiwan*	11%	Jordanië*	40%
China*	22%		
Zuid-Azië		Midden-Azië	
Thailand*	-9%	Iran*	16%
Maleisië	-3%	India*	25%
Vietnam*	39%		
Indonesië*	40%		
Afrika		Centraal-Europa	
Zambia*	9%	Servië	0%
Mali*	14%	Slovenië	0%
Burkina Faso*	21%	Oekraïne*	3%
Zuid-Afrika*	22%	Polen*	6%
Ethiopia*	29%	Bulgarije*	7%
Ghana*	31%	Moldavië*	9%
		Rusland*	16%
		Roemenië*	22%
		Georgië*	23%
Oceanië		Noord-Amerika	
Australië*	5%	USA	4%
		Canada	4%

NPSs in plaats van naar de absolute scores, deze relatieve stijgingen kunnen wel direct tussen landen worden vergeleken.

De berekende gecorrigeerde scores geven aan dat een NPS in Nederland naar verwachting laag zal zijn in vergelijking met andere landen in de wereld. Verschillen met de meeste Westerse landen zijn minder groot, maar

de verschillen met opkomende markten zoals China, India en Turkije zijn substantieel. Bijvoorbeeld, bij een vergelijking van de NPS bij vergelijkbare bedrijven in Nederland en in China zal de NPS bij de Chinese bedrijven automatisch hoger zijn bij een in werkelijkheid gelijkwaardige kwaliteit van beide bedrijven. Het is te verwachten dat het verschil in antwoordstijl het verschil in NPSs in sterke mate bepaalt.

6 Samenvatting en conclusie

Veel bedrijven werken met maatstaven en klant-KPI's zoals de Net Promoter Score (NPS). Bedrijven vergelijken graag hun verkregen scores met andere bedrijven, om zo de uitkomsten in perspectief te kunnen zetten. Het vergelijken van Net Promoter Scores, die zijn verzameld in verschillende landen of onder verschillende demografische groepen is echter niet zonder risico. Antwoorden van respondenten uit verschillende groepen worden namelijk mede bepaald door antwoordstijlen. Hierdoor kunnen metingen in verschillende landen of demografische groepen systematisch een te hoge of te lage score opleveren.

In dit onderzoek hebben we de effecten van de twee meest prominente antwoordstijlen onderzocht, namelijk: ARS en ERS. Op landniveau blijkt dat een lager BNP per hoofd van de bevolking leidt tot meer ARS en ERS en dus tot een hoger te verwachten score op de NPS bij gelijke werkelijke kwaliteit.

Uit ons onderzoek blijkt dat Nederlanders minder geneigd zijn positieve antwoorden te geven op antwoordschalen en dus op de NPS. In tabel 2 wordt de manager een handvat geboden om verschillen in NPSs beter te kunnen vergelijken tussen vestigingen in verschillende landen. Bijvoorbeeld, ten opzichte van Nederland is in Duitsland een positieve afwijking van 6% te verwachten bij het beantwoorden van de NPS. Dus wanneer een vestiging van hetzelfde bedrijf in Nederland en in de Duitsland naar de NPS vraagt bij haar klanten, is de kans relatief groter dat de Duitse vestiging beter scoort, ongeacht de daadwerkelijke kwaliteit van beide vestigingen. Toch is het verschil tussen Nederland en Duitsland klein in vergelijking met het verschil tussen Nederland en de opkomende markten waar de NPSs meer dan 30% hoger kunnen liggen.

Naast factoren op landniveau hebben individuele factoren ook invloed op het gebruik van antwoordstijlen. Demografische gegevens zoals geslacht, leeftijd en opleidingsniveau hebben invloed op het gebruik van antwoordstijlen. Deze antwoordstijlen bepalen binnen elk land systematische verschillen tussen NPSs. ARS komt meer voor bij oudere klanten en dat betekent dat, bij een gelijke dienstverlening, een vestiging met oudere klanten positiever zal worden beoordeeld dan een vestiging met jonge klanten. Eenzelfde effect is er voor opleiding: een lagere opleiding leidt tot meer ARS. Wan-

neer er tussen vestigingen binnen landen verschillende klantengroepen worden vergeleken zijn er, bij gelijke dienstverlening, aanzienlijke verschillen te verwachten. Een vestiging met een lager opgeleide klantengroep zal positiever worden beoordeeld dan een vestiging met een hoger opgeleide klantengroep.

Er zijn momenteel veel bedrijven die gebruik maken van de NPS. Ondanks alle kritiek op deze maatstaf is de NPS nog steeds een zeer populair dashboardtool voor vele organisaties. De NPS vergelijken kan, maar dan wel tussen cultureel en economisch vergelijkbare landen. Wanneer vergelijkbaarheid ontbreekt, kan ervoor worden gekozen om te kijken naar de procentuele groei van de NPS tussen één of meerdere tijdsintervallen. Op die manier is de vergelijking gerechtvaardigd, omdat de door antwoordstijlen ontstane systematische afwijking in beide tijdsintervallen gelijk is.

Tot slot, enerzijds is de verwachting dat de toenemende mondialisering verschillen tussen landen en tussen mensen zal verminderen (Ritzer, 2008). Dit geldt zeker voor managers van grote bedrijven die wereldwijd opereren, deze kosmopolieten lijken steeds meer op elkaar. Voor deze mensen zullen de verschillen op individueel niveau verminderen en naar verwachting ook de verschillen in antwoordstijlen. Anderzijds, bij het bepalen van de NPS, gebruikmakend van representatieve steekproeven van klanten, blijven verschillen tussen landen en tussen mensen bestaan. In veel landen neemt het verschil tussen arm en rijk en tussen hoog- en laagopgeleiden zelfs toe (The Economist, 2013). Voor de NPS en andere op antwoordschalen gebaseerde maatstaven zal correctie voor antwoordstijlen bij internationale vergelijkingen een aandachtspunt blijven. ■

F. Janmaat MSc. is onderzoeker en docent aan de Vrije Universiteit van Amsterdam. Hij studeerde afgelopen jaar af in Marketing met een specialisatie in cross-cultureel onderzoek. Daarnaast is hij eigenaar van onderzoeksbureau Insightful.

Prof. dr. H. van Herk is hoogleraar Cross-Cultural Marketing Research aan de Vrije Universiteit van Amsterdam. Daarnaast is zij voorzitter van de sectie Onderwijs van de MOA, Center for Information Based Decision Making & Marketing Research.

Noten

■ AFM regelgeving, inhoud & werking. (n.d.). Geraadpleegd op <http://www.afm.nl/nl/professionals/regelgeving/thema/beloningsbeleid/wanneer-beheerst/inhoud-en-werking.aspx>.

■ Evensen, R. (2011, April 18th). Stop Using NPS (Net Promoter Score) But Please Save The Question! Geraadpleegd op 4 April 2014, van [blogs.forrester.com: http://blogs.forrester.com/](http://blogs.forrester.com/richard_evensen/11-04-18-stop_using_nps_net_promoter_score_but_please_save_the_question)

[richard_evensen/11-04-18-stop_using_nps_net_promoter_score_but_please_save_the_question](http://blogs.forrester.com/richard_evensen/11-04-18-stop_using_nps_net_promoter_score_but_please_save_the_question).

- Baumgartner, H., & Steenkamp, J.B.E.M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2), 143-156.
- Chen, C., Lee, S.Y., & Stevenson, H.W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6(3), 170-175.
- Doorn, J. van, Leeflang, P.S.H., & Tijs, M. (2013). Satisfaction as a predictor of future performance: A replication. *International Journal of Research in Marketing*, 30(3), 314-318.
- Gallup (2009). Religion importance. Geraadpleegd op www.gallup.com.
- Greenleaf, E.A. (1992). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, 29(2), 176-188.
- Harzing, A.W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, 6(2), 243-266.
- Hausmann, R., Tyson, L.D., & Zahidi, S. (2006). *The global gender gap report 2006*. Cologne/Geneva: World Economic Forum.
- Herk, H. van, Poortinga, Y.H., & Verhallen, T.M.M. (2004). Response styles in rating scales: evidence of method bias in data from six EU countries. *Journal of Cross Cultural Psychology*, 35(3), 346-360.
- Hoekstra, J.C., & Leeflang, P.S.H. (2010). Waardering van en voor marketinginspanningen. *Maandblad voor Accountancy en Bedrijfsconomie*, 84(11), 588-595.
- Hofstede, H. (2001). *Culture's consequences: comparing values, behaviors, institutions and organizations across nations* (tweede editie). Thousand Oaks, CA: Sage Publications.
- Hox, J.J., & Leeuw, E.D. de (1997). Meta-analyse via multiniveau-modellen. *Tijdschrift voor Onderwijsresearch*, 22, 16-28.
- Johnson, T., Kulesa, P., Cho, Y.I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, 36(2), 264-277.
- Jong, M.G. de, Steenkamp, J.B.E.M., Fox, J.P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. (1), 104-115.
- Knauper, B. (1999). The impact of age and education on response order effects in attitude measurement. *Public Opinion Quarterly*, 63, 347 - 370.
- Lambsdorff, J.G. (2006). Corruption Perception Index 2006. Passau, Lower Bavaria, Germany.
- Marin, G., Gamba, R.J., & Marin, B.V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross Cultural Psychology*, 23(4), 498-509.
- Paulhus, D.L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver en L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diego, CA: Academic Press.
- Polloack, B.L., & Alexandrov, A. (2013). Nomological validity of the Net Promoter Index question. *Journal of Service Marketing*, 27(2), 118-129.
- Reichheld, F. (2003) The one number you need to grow. *Harvard Business Review*, 81(12), 46-55.
- Reichheld, F., & Markey, R. (2011). *The ultimate question 2.0: How companies thrive in a customerdriven world*. Boston, MA: Harvard Business School Publishing.
- Ritzer, G. (2008). *The McDonaldization of society* (vijfde editie). Los Angeles, CA: Pine Forge Press.
- Rosmalen, J. van, Herk, H. van, & Groenen, P.J.F. (2010). Identifying response styles: A latent-class bilinear multinomial logit model. *Journal of Marketing Research*, 47(1), 157-172.
- The Economist (2013). Inequality. Growing apart. 21 September 2013.
- The World Bank (2014). *GDP per capita*. Geraadpleegd op: <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>.
- Weijters, B., Geuens, M., & Schillewaert, N. (2010). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, 34(2), 105-121.
- White, A. (2006). *World map of happiness*. University of Leicester: EurekaAlert.