

8. *Buttons of balken, klikken of slepen: wat werkt er nu het beste op mobiele telefoons, tablets of PCs?*¹

VERA TOEPOEL²

SAMENVATTING

Tegenwoordig weet je niet op welk apparaat respondenten je online vragenlijst invullen. PC's, tablets, en mobiele telefoons variëren sterk in schermgrootte. Een onderzoeker kan uit verschillende antwoordformats kiezen zoals buttons, slider balken, en Visual Analogue Scales (VAS). Verschillen in schermgrootte en operationalisatie van antwoordformats kunnen voor verschillen in antwoordgedrag zorgen. In een studie, uitgevoerd onder meer dan 5000 panelleden van GfK, ben ik nagegaan hoe panelleden reageren op verschillende antwoordformats. Ik heb daarbij het aantal schaalpunten gevarieerd, van 5, 7 tot 11. Het doel was na te gaan hoe antwoord format, aantal schaalpunten, en apparaten (mobiel, tablet, en PC) interacteren. Ik heb ook evaluatievragen toegevoegd, zodat ik niet alleen kan kijken naar datakwaliteit, maar ook naar het gebruiksgemak van de verschillende formats

Trefwoorden: mobiele surveys, design, antwoordformats, schaaltype, datakwaliteit

1. INLEIDING

Meer en meer mensen hebben een mobiele telefoon of tablet. Dit zorgt ervoor dat online vragenlijsten tegenwoordig ingevuld worden op veel verschillende soorten apparaten. Als onderzoeker weet je niet van te voren op welk apparaat respondenten je online vragenlijst in gaan vullen. PC's, tablets, en mobiele telefoons variëren sterk in schermgrootte en maken gebruik van een toetsenbord of touchscreen. De layout en zichtbaarheid van de vragen, alsmede de manier waarop antwoorden gegeven worden, varieert. Bovendien gaat elk platform anders om met antwoord formats. Dit werpt de vraag op of we antwoorden gegeven op bijvoorbeeld mobiele telefoons kunnen vergelijken met antwoorden gegeven op traditionele PC's. De literatuur laat immers zien dat zichtbaarheid en operationalisatie (bijvoorbeeld klikken of slepen) van antwoordformats zorgen voor verschillen in antwoordgedrag (zie Couper et al., 2004; Toepoel & Dillman, 2011).

Wanneer je een online survey programmeert, kun je uit vele antwoordformats kiezen. In online vragenlijsten wordt vaak gebruik gemaakt van radio buttons, knoppen waarin je kunt klikken om een antwoord te geven. Alternatieven zijn balken of lijnen waarop je een antwoord kunt geven. De medische sector maakt vaak

A.E. Bronner et al. (red.), Ontwikkelingen in het marktonderzoek: Jaarboek MarktOnderzoek-Associatie, dl. 41, 2016. Haarlem: Spaar^{en}Hout.

gebruik van zogenaamde Visual Analogue Scales (VAS) waarbij de respondent op een lijn kan klikken om een antwoord te geven. Marktonderzoek maakt ook vaak gebruik van Slider Balken, waarmee een zogenaamde hendel verslept kan worden naar het gewenste antwoord (Funke, 2015). Het is alom bekend dat de keuze voor een antwoordformat antwoordgedrag beïnvloed (zie Smith, 1995; Jenkens & Dillman, 1995; Couper, Tourangeau & Conrad, 2004; Toepoel & Dillman, 2011, en Toepoel, Das & Van Soest, 2009) en ook de zichtbaarheid van antwoorden heeft een effect op de resultaten (Couper et al., 2004). Vooral zichtbaarheid kan een probleem zijn wanneer mensen vragenlijsten invullen op hun mobiele telefoon, omdat mobiele telefoons kleine schermpjes hebben. Bovendien kan zichtbaarheid gerelateerd zijn aan de lengte van de antwoordschaal, in het bijzonder het aantal schaalpunten. Een 11-puntsschaal neemt in traditionele formats meer ruimte in dan een 5-puntsschaal. Vooral radio buttons zijn erg inefficiënt als het op ruimte aankomt. Bij invullen op een mobiele telefoon heeft de keuze voor een 5- of 11-puntsschaal met radio buttons een direct effect op zichtbaarheid en gebruiksvriendelijkheid van de antwoordmogelijkheden. Onderzoek moet nagaan hoe apparaten, aantal schaalpunten en antwoordformats van elkaar afhankelijk zijn, om na te gaan of we antwoorden gegeven op PC's, tablets en mobiele telefoons wel met elkaar kunnen vergelijken. In dit hoofdstuk bespreek ik een experiment dat ik in mei 2014 gedaan heb in het GfK Online Panel. Respondenten werden random toegewezen aan een conditie waarin hen werd verzocht de vragenlijst in te vullen op respectievelijk een traditionele PC, tablet of mobiele telefoon. Verder varieerde ik de antwoordformats (verschillende vormen van buttons en balken) en het aantal schaalpunten (5, 7, of 11). Ik onderzoek meetfouten en item nonrespon. Ik heb bovendien evaluatievragen toegevoegd zodat ik ook kan kijken naar het gebruiksgemak vanuit het oogpunt van de respondent.

2. ACHTERGROND

2.1. *Antwoordformats*

Bij het maken van een online vragenlijst moet je als onderzoeker/programmeur keuzes maken voor wat betreft het soort antwoord dat respondenten kunnen geven. De meeste online software heeft de mogelijkheid om gebruik te maken van radio buttons. Met de komst van de mobiele technologie, en het toenemende gebruik van mobiele telefoons en tablets, zijn veel marktonderzoekbureaus hun software aan gaan passen. Eén van deze aanpassingen is het gebruik van grote vakken (cellen) in plaats van de traditioneel kleine radio buttons. Hierdoor is het makkelijker om een antwoordoptie te kiezen op een klein scherm met touchscreen. Immers, de verhouding vinger-mobiele telefoonscherm is heel anders dan muis-PCscherm, en het aanklikken van een verkeerde optie is een veelgeplaagd probleem op apparaten met een klein scherm en touchscreen als methode voor navigatie. Ook het gebruik van balken -VAS of slider balken- wint aan populariteit omdat ze minder ruimte innemen dan traditionele radio buttons. Al met al zijn er veel formats waar je als onderzoeker/programmeur uit kunt kiezen, en de keuze voor een bepaald format wordt vaak gemaakt op basis van persoonlijke voorkeuren van de ontwerper. Hierbij wordt

vaak weinig rekening gehouden met eventuele meetfouten (zie de Leeuw, 2010). Dit terwijl de literatuur op het gebied van meetfouten en antwoordformats omvangrijk is. Zie bijvoorbeeld het werk van Smith (1995), Jenkens & Dillman (1995), Couper, Tourangeau & Conrad (2004), Toepoel & Dillman (2011), Toepoel, Das & Van Soest (2009) en Sikkel, Gras, Steenbergen en Oosterveld (2014). Voor een goed begrip ga ik even dieper in op de verschillende mogelijkheden.

Radio buttons (zie Appendix A, Figuur 1) zijn ronde cirkels waarin een respondent kan klikken om een antwoord te geven. Radio buttons maken gebruik van standaard HTML en werken in elke browser. Ze zijn een low-tech antwoordformat en elke respondent weet hoe ze te gebruiken. Het nadeel van radio buttons is dat ze veel ruimte nodig hebben. Traditionele radio buttons zijn namelijk niet schaalbaar.

Big buttons (zie Appendix A, Figuur 5) zijn een nieuwe vorm van de traditionele radio buttons. Het verschil zit in het klikbare oppervlak. Bij big buttons kan de respondent op een hele cel klikken (bijvoorbeeld op de tekst van de antwoordoptie) en hierdoor zijn ze makkelijker te gebruiken op apparaten die gebruik maken van touch screen en een klein scherm. Het is onbekend of het gebruik van big buttons resulteert in andere antwoorddistributies dan het gebruik van traditionele radio buttons. VAS (zie Appendix A, Figuur 3) bevatten een (meestal horizontale) lijn. Respondenten geven een antwoord door ergens op die lijn te klikken, zodat er een marker gezet wordt. VAS gaan efficiënt met ruimte om. De lengte van de lijn in pixels kan de range van de antwoordschaal zijn. Zo kan een VAS met 50 antwoordopties evenveel ruimte innemen als een 3-puntsschaal met radio buttons (Funke, 2015). VAS kunnen zowel geoperationaliseerd worden in een continue schaal (waarbij elke pixel een antwoordoptie is) als in een discrete n-punt schaal. Nadeel van VAS is dat ze meer technologie vereisen aan de kant van de respondent (JavaScript) en daardoor meer high-tech zijn dan de traditionele radiobuttons. Wanneer de respondent de vereiste software niet heeft geïnstalleerd zorgt dit voor problemen.

Slider balken (zie Appendix A, Figuur 2) zijn vergelijkbaar met VAS omdat ze ook uit een lijn bestaan. Het verschil met VAS is dat sliders een hendel hebben die verschoven moet worden om een antwoord te geven. Sliders werken met een zogenaamd drag-and-drop principe; respondenten moeten de hendel verslepen en loslaten op het gewenste antwoordpunt. De initiële positie van de hendel is een groot probleem van sliders. Funke, Reips en Thomas (2011) hebben aangetoond dat de initiële positie van de hendel tot meetfouten kan leiden. Sliders kunnen geoperationaliseerd worden met HTML5 of gebruikerssoftware zoals JavaScript. Slider balken en VAS worden vaak door elkaar gebruikt (zie Couper, Tourangeau, Conrad & Singer, 2006).

Sikkel, Gras, Steenbergen en Oosterveld (2014) experimenteerden met klikken en slepen en vonden dat de formats die gebruik maakten van slepen beter geëvalueerd werden door respondenten. Echter, de voordelen vielen weg in een her-test. De auteurs concludeerden dat een sleep-principe alleen positief werkt als het spaarzaam wordt gebruikt, anders gaat de nieuwigheid eraf. Funke (2015) laat zien dat sliders negatief presteren ten opzichte van radio buttons en VAS met klik. Hij beargumenteert dat een klein verschil in operationalisatie – klikken versus slepen – een groot

effect kan hebben op de data. Zijn onderzoek laat zien dat sliders een negatief effect hebben op de respons, de samenstelling van de steekproef, de frequentieverdelingen van variabelen, en de duur van de vragenlijst. In zijn experimenten laat hij ook zien dat VAS geen negatieve bijeffecten hebben. Zelfs niet op touch screen apparaten zoals mobiele telefoons. De equivalentie van radio buttons en VAS met klikken is een bevestiging van eerder onderzoek (Funke en Reips, 2012).

2.2. Schermgrootte

Wanneer respondenten op verschillende apparaten met verschillende schermgroottes vragenlijsten invullen, vraag je je als onderzoeker af of dit meetfouten met zich meebrengt. In die zin lijkt een web survey tegenwoordig meer weg te hebben van een mixed-mode survey waarbij verschillende manieren van afname (telefoon, face-to-face, papier, web) gebruikt worden. De keuze van een onderzoeker voor een specifieke afname hangt af van zaken zoals verwachte respons, kosten en snelheid. Op dezelfde manier maakt de respondent de keuze voor een bepaald apparaat. De keuze voor het gebruik van een mobiele telefoon of gewone PC hangt af van beschikbaarheid, gebruiksgemak, en locatie. Onderzoek naar mixed-mode surveys is de afgelopen jaren flink toegenomen (Couper, 2011; de Leeuw, 2010; Olivier, 2008). Studies naar mixed-device surveys, waarbij verwezen wordt naar online surveys die op verschillende apparaten ingevuld kunnen worden, zijn vrij recent. We kunnen echter dezelfde selectie en meetfouten verwachten als bij mixed-mode surveys. Mixed-mode onderzoek laat zien dat selectie en meetfouten het moeilijk maken om antwoorden in verschillende formats te vergelijken (Schouten et al., 2013).

Het gebruik van mobiele apparaten is de afgelopen paar jaar toegenomen. Poggio, Bosnjak, en Weyandt (2015) concluderen dat 4% van de panelleden in het GESIS Pilot Panel (Duitsland) surveys invult op hun mobiele telefoon in 2012. De Bruijne & Wijnant (2014b) noemen 16% in het CentERPanel (Nederland) in 2013. Toepoel, Lugtig en Amin (2015) rapporteren dat 10% van alle panelleden in het American Life Panel een vragenlijst via de mobiele telefoon heeft ingevuld gedurende 5 waves van het panel in 2014. Alle vragenlijsten in deze panels waren niet geoptimaliseerd voor mobiele telefoons. Ze maakten gebruik van designs die waren geoptimaliseerd voor traditionele PCs met grote schermen. Toepoel & Lugtig (2014) maakten gebruik van een zogenaamd responsive design – waar vragenlijsten automatisch werden geoptimaliseerd al naar gelang het type apparaat – in het MarketResponse panel. Na een bewuste opmerking dat de vragenlijst was geoptimaliseerd voor mobiele telefoons, werd de vragenlijst door 57% van de panelleden (met een mobiele telefoon) op een mobiele telefoon met internettoegang ingevuld.

We zijn aangekomen in een tijdperk wanneer het niet meer volstaat om vragenlijsten alleen gebruiksvriendelijk te maken voor grote schermen zoals van PCs. Met de groei van mobiel internet verwachten consumenten dat surveys aangepast worden op mobiele telefoons. Dit brengt nieuwe aandachtspunten met zich mee, omdat we weten dat de lay-out van een vragenlijst antwoorden kan vertekenen. Als de vragenlijst er op een PC niet hetzelfde uitziet als op een mobiele telefoon, kun je antwoorden dan nog wel met elkaar vergelijken? Tourangeau, Couper en Conrad (2013)

laten zien dat de positie van een item op een scherm antwoordgedrag beïnvloed. Echter, De Bruijne & Wijnant (2013a) vinden geen systematische verschillen in antwoordgedrag wanneer ze een PC en een mobiele web lay-out met elkaar vergelijken. Respondenten waren wel minder tevreden met de mobiele web lay-out.

2.3. Aantal schaalpunten

In het begin van dit hoofdstuk heb ik geconcludeerd dat de traditionele radio buttons niet schaalbaar zijn en niet effectief met ruimte omgaan. Veel onderzoekers werken met zogenaamde Likert-schalen. Het is dan gebruikelijk om tussen de vijf en elf schaalpunten te nemen (De Beuckelaer, Toonen & Davidov, 2013) Wanneer je in een traditioneel radio button format een keuze maakt voor een 5-- of 11-puntsschaal, heeft dit effect op de lengte die de antwoordschaal inneemt.

Elf-puntsschalen laten een hoger precisieniveau toe, maar ze vragen ook een behoorlijk inspanningsvermogen van de respondent. Dit kan leiden tot meetfouten. De Beuckelaer et al. (2013) laten zien dat een 11-puntsschaal voordelen heeft qua analyses en resulteert in een hogere betrouwbaarheid. Zij laten echter ook zien dat een 7-puntsschaal een redelijk alternatief is. Een verdere terugschaling naar een 5-puntsschaal werd problematisch en resulteerde in een hogere mate van inconsistentie in antwoordgedrag. Choudhury en Bhattacharjee (2014) merken ook een toename in betrouwbaarheid met een toename in het aantal schaalpunten, hoewel de verschillen klein waren. Preston en Coleman (2000) laten eveneens zien dat een schaal met minimaal zeven schaalpunten beter presteerde op indices voor betrouwbaarheid en validiteit. De voorkeur van respondenten lag bij de 10-puntsschaal, gevolgd door een 7- en 9-puntsschaal. Maitland (2009) komt in een literatuurstudie tot de conclusie dat de voorkeur uitgaat naar een 7-puntsschaal. Dit is in lijn met de theoretische analyse over het informatieverwerkingsvermogen van mensen in het algemeen (Miller, 1956). Literatuur laat zien dat antwoordopties die initieel zichtbaar zijn, vaker worden gekozen dan opties die niet meteen zichtbaar zijn (waarbij respondenten eerst een actie moeten ondernemen – zoals scrollen of klikken- om de overige antwoordopties te zien. Couper et al., 2004). De Bruijne & Wijnant (2014) concluderen dat de zichtbaarheid van antwoordopties aanzienlijk verschilt op mobiele telefoons: een 5-puntsschaal was geheel zichtbaar op de mobiele telefoon, terwijl een 11-puntsschaal alleen geheel zichtbaar was voor 59% van de respondenten. De gemiddelde duur van de vragenlijst was langer voor de 11-puntsschaal maar dat had geen effect op het cijfer dat de respondent aan de vragenlijst toekende. De Bruijne and Wijnant vonden geen verschil tussen een horizontaal en verticaal uitgelijnde schaal. Dit wordt bevestigd door Funke, Reips & Thomas (2011) die ook geen relatie vonden tussen oriëntatie en data kwaliteit.

Literatuur suggereert dat de effecten in schaalengte voornamelijk veroorzaakt worden door respondentkenmerken. Weathers, Sharma en Niedrich (2005) beargumenteren dat mensen met een groot informatieverwerkingsvermogen beter in staat zijn om veel schaalpunten te hanteren.

3. METHODE

In deze studie wil ik nagaan of de verschillende antwoordformats equivalente meet-instrumenten zijn. Ik heb daarbij geprobeerd om de functionaliteit van de slider balk te verbeteren, door de hendel van de slider buiten het antwoordveld te plaatsen (zie Appendix A, Figuur 2). Ik voeg daarbij nog een antwoord format toe waarbij de slider werkt met zowel een klik als een sleep-principe (zie Appendix A, Figuur 4, ik noem dit de “combinatie slider/VAS”. Ik vergelijk antwoorden gegeven op verschillende apparaten (PC, tablet, mobiele telefoon) en op verschillende schaallengtes (5-, 7-, en 11 puntsschaal).

Data werden verzameld in het GfK Online Panel. De vragenlijst bestond uit drie blokken met vragen: drie vragen naar attitude ten opzichte van vragenlijsten (niet gerapporteerd), 16 items over vakantie-ervaringen die dienst deden als onze test items, en zes evaluatievragen. De 16 test items varieerden in antwoordformat en schaallengte. De vragenlijst duurde ongeveer vijf minuten.

We gebruikten een 3*5*3 design waarin panelleden random werden toegewezen aan de volgende condities:

- Desktop, tablet of mobiele telefoon
- Radiobuttons, big buttons, slider (slepen), VAS (klikken), combinatie slider/VAS (zowel klikken als slepen)³
- 5, 7-, 11 puntsschaal

Elk format was identiek qua lengte maar varieerde in functionaliteit (klikken of slepen) en antwoordtype (buttons of balken), zie Appendix A. Zodoende varieerden de schalen alleen in het aantal schaalpunten en niet in lengte van de antwoordmogelijkheden.

Het veldwerk werd uitgevoerd van 8-16 april 2014. Aangezien de respons op de mobiele apparaten achterbleef bij de PC variant, werd een extra uitnodiging gestuurd op 11 april voor tablets en mobiele telefoons. Het aantal uitgenodigden en respondenten staan in Tabel 1. We hebben een quota gebruikt voor het aantal respondenten op elk apparaat; 5.339 panelleden hebben de vragenlijst niet ingevuld vanwege deze quota. Het aantal mensen dat uitviel tijdens de vragenlijst was 74. Vanwege dit geringe aantal ga ik niet in op uitval tijdens de vragenlijst.

Tabel 1. Uitnodiging en respons per apparaat.

	Desktop	Tablet	Mobiele telefoon	Totaal
Uitnodiging 8 april	2.694	2.695	2.696	8.085
Uitnodiging 11 april		2.618	6.228	8.846
Totaal				16.931
respons	1709	1702	1666	5.007

4. RESULTATEN

Een aanzienlijk aantal respondenten negeerde de instructie om een bepaald apparaat te gebruiken. Zie Tabel 2. Van de panelleden die aan een desktop conditie waren toegewezen, vulde 76% de vragenlijst op een desktop PC in. De overige 24% negeerde de instructie en vulde de vragenlijst op een tablet (15%) of mobiele telefoon (9%) in. Ongeveer de helft van de panelleden die toegewezen waren aan een tablet of mobiele telefoon vulde daadwerkelijk de vragenlijst op het gewenste apparaat in. In onderstaande analyses heb ik panelleden ingediend op basis van het apparaat dat ze daadwerkelijk gebruikt hebben.

Tabel 2. *tussen toegewezen conditie en daadwerkelijk apparaat gebruik.*

Gebruikt	Toegewezen			
	Desktop	tablet	mobiele telefoon totaal	
desktop	76%	28%	19%	1709
tablet	15%	54%	28%	1702
mobiele telefoon	9%	18%	53%	1666

4.1. Data kwaliteit

4.1.1. Gemiddelde scores

Tabel 3 geeft de gemiddelde antwoordscores in de verschillende formats over de verschillende apparaten en schaalpunten. Wanneer we naar de totalen kijken over de verschillende apparaten, dan zien we dat er voor zowel de 5, 7, als 11-puntsschaal significante verschillen zijn tussen de traditionele radio buttons en de slider balken. Dit betekent dat dit niet equivalente meetinstrumenten zijn. In alle gevallen is het gemiddelde van de sliders significant lager dan de radiobuttons. Dit wordt waarschijnlijk veroorzaakt door de positie van de hendel, die links buiten de antwoordlijn was geplaatst (zie Appendix A, Figuur 2. Het verslepen van de hendel van links naar rechts vereist meer oog-hand coördinatie, waardoor de hendel niet ver genoeg naar rechts wordt gesleept (in vergelijking met de gemiddelde scores op de radiobutton schaal). Blijkbaar heeft de operationalisatie van de sliders – met de plaatsing van de hendel links buiten het antwoordveld - er niet voor gezorgd dat de slider balken beter presteren dan de literatuur ons uitwees.

Wanneer we inzoomen op de conditie met mobiele telefoons, dan zien we bij de 5- en 7-puntsschaal geen significante verschillen tussen de antwoordformats. Alle formats presteren dus identiek. Bij de 11-puntsschaal zien we weer dat de sliderbalken geen equivalent meetinstrument zijn met de radiobuttons. Ook de VAS laat een significant lagere score zien ten opzichte van de radiobuttons. Dit suggereert dat een 11-puntsschaal met balken (slider en VAS) niet vergelijkbaar is met een 11-puntsschaal met knoppen op de mobiele telefoon.

Voor de vragenlijsten ingevuld op tablets is het verhaal net iets anders. Hier presteren alle antwoordformats identiek bij de 7- en 11-puntsschalen, maar niet bij de 5-puntsschaal. Zowel bij de tablets als bij de desktop PC's – apparaten met relatief grotere schermen- genereren de sliderbalken significant andere antwoorden dan de

Tabel 3. Gemiddelde scores en ANOVA testen voor het aantal schaalpunten, antwoordformats en apparaten. Standaard deviatie tussen parentheses.

	Mobiele telefoon	Tablet	Desktop	Totaal
5 puntsschaal	F(489,4)=,37	F(500,4)=2,32*	F(502,4)=2,47**	F(1493,4)=3,67***
Radio buttons	3,00 (.68) N=98	3,14 (.66) N=100	3,17 (.71) ^a N=100	3,10 (.69) ^a N=298
Big buttons	3,00 (.64) N=98	3,19 (.60) ^a N=100	3,15 (.19) N=101	3,11 (.73) ^a N=299
Slider	2,99 (.64) N=98	2,91 (.69) ^b N=100	2,87 (.82) ^b N=102	2,92 (.72) ^b N=300
VAS	3,06 (.68) N=98	3,08 (.75) N=101	3,10 (.70) N=100	3,08 (.71) N=299
Slider/VAS	2,96 (.54) N=98	3,03 (.71) N=100	3,07 (.71) N=100	3,02 (.66) N=298
7 puntsschaal	F(489,4)=,64	F(500,4)=1,49	F(500,4)=3,16**	F(1491,4)=4,23***
Radio buttons	4,25 (.81) N=98	4,33 (.85) N=100	4,38 (1,06) N=100	4,32 (.91) ^a N=298
Big buttons	4,11 (.88) N=98	4,30 (1,02) N=100	4,31 (1,08) N=100	4,24 (1,00) N=298
Slider	4,11 (.94) N=98	4,14 (1,11) N=100	4,00 (1,28) N=100	4,08 (1,12) ^b N=298
VAS	4,08 (.94) N=98	4,19 (1,08) N=100	4,35 (.93) N=100	4,21 (.99) N=298
Slider/VAS	4,06 (1,07) N=98	4,01 (1,21) N=101	3,98 (1,19) N=101	4,02 (1,16) ^b N=300
11 puntsschaal	F(489,4)=3,46***	F(499,4)=1,15	F(503,4)=1,59	F(1493,4)=3,96***
Radio buttons	6,48 (1,40) ^a N=98	6,30 (1,35) N=100	6,58 (1,35) N=100	6,45 (1,37) ^a N=298
Big buttons	6,13 (1,34) N=98	6,16 (1,56) N=100	6,33 (1,82) N=102	6,21 (1,58) N=300
Slider	5,78 (1,49) ^b N=98	5,95 (1,49) N=100	6,19 (1,62) N=100	5,97 (1,54) ^b N=298
VAS	5,88 (1,72) ^b N=98	6,30 (1,29) N=100	6,54 (1,33) N=101	6,24 (1,48) N=299
Slider/VAS	6,25 (1,52) N=98	6,02 (1,67) N=100	6,18 (1,35) N=101	6,15 (1,51) N=299

*p<.1, **p<.05, ***p<.01

Noot: abcd=homogeneous subsets, subset a verschilt significant van subset b. radio buttons.

De resultaten suggereren dat sliders andere antwoorden genereren dan de andere formats. Dit geldt vooral voor 11-puntsschalen op mobiele telefoons en 5-puntsschalen op apparaten met grotere schermen (tablet en desktop PC).

4.2. Nonresponse

Zoals eerder opgemerkt waren er 74 breakoffs. Door dit lage aantal kan ik geen analyses over de verschillende condities maken. Ik heb wel gekeken naar de ontbrekende waarden (missings) per item. Tabel 4 geeft de resultaten van een regressie-analyse met als afhankelijke variabele het gemiddelde aantal missende antwoorden over de

zestien items heen. Opvallend is dat de verschillende antwoordformats op zich in vergelijking met traditionele radio buttons (referentie categorie) minder item non-respons met zich mee brengen, maar dat bepaalde interacties tussen de verschillende antwoordformats en apparaten/schaalpunten *meer* item nonresponse laten zien. Zo zorgt een slider an sich voor minder nonrespons, maar zien we voor een slider op 5 en 7 puntsschalen en op mobiele apparaten (tablet/mobiel) meer item nonresponse. Ook een vijfpuntsschaal op de mobiele telefoon genereert gemiddeld gezien meer item nonresponse. Opvallend genoeg zijn er geen significante verbanden tussen item nonrespons en type apparaat en ervaring op de verschillende apparaten.

4.3. Evaluatie van respondenten

In totaal waren er 6 evaluatievragen aan de vragenlijst toegevoegd (duidelijkheid van de vragen, hoe leuk de vragenlijst was, de vormgeving, gebruiksvriendelijkheid, motivatie van respondenten en hoe serieus de vragenlijst is ingevuld). Tabel 4 geeft de totale somscore over de zes items heen, alsmede de individuele resultaten voor de items “design” en “gebruiksvriendelijkheid”. Alle evaluatievragen werden gemeten met een rapportcijfer van 0 tot en met 10.

In een lineaire regressie met de totale somscore als onafhankelijke variable, en de antwoordformats, aantal schaalpunten, apparaten, zelfselectie naar een bepaald apparaat, ervaring met surveys op de verschillende apparaten, en interactietermen, blijkt dat respondenten een voorkeur hebben voor een 11-puntsschaal. Vragenlijsten die zijn ingevuld op een tablet of mobiele telefoon worden significant lager beoordeeld dan vragenlijsten ingevuld op een desktop PC. Opvallend is dat mensen die zich niet aan het toegewezen apparaat hielden, negatiever zijn over de vragenlijst. Hoe meer ervaring panelleden hebben met het invullen van een vragenlijst op PC of mobiele telefoon, hoe positiever ze zijn over de vragenlijst. Dit geldt niet voor tablets. Wanneer we kijken naar de interacties, dan zien we dat panelleden die een vragenlijst kregen met balken (VAS of Slider) op hun mobiele telefoon, positiever zijn over de vragenlijst. Dit suggereert dat balken prettiger door panelleden worden ervaren op mobiele telefoons. Ook was er een significante interactie tussen een 5-puntsschaal en de slider/VAS combinatie. Verder waren er geen significante interacties tussen antwoordformats en apparaten met het aantal schaalpunten.

Wanneer we specifiek inzoomen op het design van de vragenlijst (“Hoe vond u de vormgeving van de vragen?”), dan zien we dat alle antwoordformats slechter worden geëvalueerd dan de traditionele radiobuttons (referentie). Vooral de VAS wordt veel slechter geëvalueerd. Panelleden die de vragenlijst via tablet of mobiel hebben ingevuld, zijn ook veel negatiever over het design van de vragenlijst; de eerste groep geeft gemiddeld een tiende (0,1) rapportcijfer lager aan het design in vergelijking met de panelleden die de vragenlijst op een desktop PC hebben ingevuld. De panelleden die de vragenlijst op hun mobiele telefoon hebben ingevuld zelfs twee tiende lager (0,2). Desondanks worden balken op een mobiel apparaat – zowel tablet als mobiele telefoon- significant beter geëvalueerd. Balken lijken dus goed op mobiele apparaten te werken. Panelleden zijn minder positief over het gebruik van een 5-puntsschaal op de mobiele telefoon. Panelleden die zich niet hebben gehouden aan de instructie en

Tabel 4. Beta coëfficiënten van een lineaire regressie op item nonrespons en evaluatie van de vragenlijst.

	Item nonresponse	Totale evaluatie	Design	Gebruiks-vriendelijkheid
Radiobuttons	Ref.			
Big buttons	-,089***	-,008	-,07**	-,018
Slider	-,061*	-,018	-,075**	-,052
Vas	-,122***	-,043	-,104***	-,065*
Slidervas	-,070**	-,051	-,080**	-,051
Fivepoint	-,190***	,008	-,027	,044
Sevenpoint	Ref.	Ref.	Ref.	Ref.
elevenpoint	-,198***	,090**	,046	,090**
Pc	Ref.	Ref.	Ref.	Ref.
Tablet	,049	-,070*	-,099***	-,070*
Mobile	-,021	-,243***	-,233***	-,286***
Selfselect	-,011	-,048**	-,038	-,044*
Bigbuttons*fivepoint	,087***	-,013	,021	-,022
Bigbuttons*elevenpoint	,093***	-,037	-,022	-,030
Bigbuttons*tablet	-,032	,011	,027	,022
Bigbuttons*mobile	-,041*	0,28	,037	-,038
Slider*fivepoint	,125***	,014	,052**	,009
Slider*elevenpoint	,049**	-,018	,009	-,018
Slider*tablet	,045*	,022	,048*	,049**
Slider*mobile	,049**	,045*	,076***	,044*
VAS*fivepoint	,083***	-,006	,024	-,008
VAS*elevenpoint	,094***	-,008	,017	-,006
VAS*tablet	,024	,050**	,067***	,036
VAS*mobile	,024	,045*	,070***	,080***
s/v*fivepoint	,113***	,049*	,068***	,036
s/v*elevenpoint	,053**	,002	,009	,014
s/v*tablet	,062**	,037	,059**	,042*
s/v*mobile	,027	,058**	,063**	,068***
Fivepoint*tablet	,015	-,011	-,016	-,014
Fivepoint*mobile	,042*	-,033	-,050**	-,027
Elevenpoint*tablet	,016	-,035	-,008	-,029
Elevenpoint*mobile	,038	-,028	-,031	-,023
Selfselect*tablet	,012	,030	,027	,031
Selfselect*mobile	-,004	,061***	,040**	,099***
Survey experience desktop	-,001	,062***	,041***	,031**
Survey experience tablet	-,021	,013	-,023	,000
Survey experience mobile	,018	,040**	,037**	,066***
ANOVA	F(34,5076)= 5,95***	F(34, 4688)= 5,09***	F(34, 4980)= 4,82***	F(34, 4971)= 6,20***

*p<.1, **p<.05, ***p<.01

de vragenlijst op een mobiele telefoon hebben ingevuld (en niet op de gevraagde PC of tablet) zijn positiever over het design van de vragenlijst. Verder zijn ook panelleden met veel ervaring met het invullen van vragenlijsten op desktop computers en mobiele telefoons positiever over het design van de vragenlijst. We zien eenzelfde patroon voor de vraag over de gebruiksvriendelijkheid van de vragen.

5. DISCUSSIE EN CONCLUSIE

In dit hoofdstuk heb ik verschillende antwoordformats en apparaten met elkaar vergeleken om na te gaan of het equivalente meetinstrumenten zijn. We zeven panelleden van het GfK Online Panel random toe aan:

- a) Desktop PC, tablet of mobiele telefoon
- b) Radiobuttons, big buttons, slider balken, VAS, en een combinatie van slider/VAS, oftewel knoppen versus balken.
- c) 5-, 7-, of 11 puntsschaal.

De resultaten laten zien dat de sliderbalken geen equivalent meetinstrument is in vergelijking met de andere antwoordformats. Dit is in lijn met de literatuur, hoewel we geprobeerd hadden de operationalisatie van de slider te verbeteren. We plaatsten de hendel aan de linkerkant *buiten* de antwoordbalk. Antwoorden op de sliderbalken waren significant lager (meer naar links) in vergelijking met de antwoordformats met knoppen. Dit was vooral duidelijk op de lange schaal (11-puntsschaal) op mobiele telefoons en de korte antwoordschaal (5-puntsschaal) op apparaten met een groter scherm (desktop PC en mobiel).

Onze vragenlijst was slechts vijf minuten en de breakoff was dan ook beperkt (totaal 74 cases). Item nonresponse hing vooral samen met de interactie tussen antwoordformats (buttons of balken) en het aantal schaalpunten en apparaten. Het soort apparaat en de ervaring van panelleden met de verschillende apparaten had geen relatie met item nonrespons.

Ook werd in dit onderzoek duidelijk dat panelleden instructies om een bepaald apparaat te gebruiken vaak niet opvolgen. Ongeveer 1 op de 4 panelleden vulden de vragenlijst niet op een desktop PC in terwijl dit werd gevraagd, in vergelijking met de helft van de panelleden toegewezen aan een tablet of mobiele telefoon conditie.

Panelleden evalueerden de vragenlijsten significant beter wanneer ze een 11-puntsschaal voorgelegd hadden gekregen. Panelleden die de vragenlijst op hun tablet of mobiele telefoon hadden ingevuld waren negatiever over de vragenlijst, hoewel dit gecompenseerd lijkt te worden met ervaring. Panelleden met veel ervaring met het invullen van vragenlijsten op mobiele telefoons waren namelijk positiever dan mensen met minder ervaring. Alle formats werden slechter geëvalueerd dan de traditionele radio buttons. Echter, het gebruik van balken op mobiele apparaten wordt wel goed geëvalueerd. In een toekomst met veel mobiele apparaten lijken balken dus weldegelijk een goed alternatief.

Slider balken moeten vermeden worden in vragenlijstonderzoek. VAS, aan de andere kant, die werken met een klikprincipe op een balk, presteren over het algemeen in lijn met radio buttons. Aangezien respondenten de balken significant beter evalueerden dan de radio buttons, en VAS efficiënter met de ruimte op een scherm omgaan, zouden VAS een aantrekkelijk alternatief zijn voor de traditionele radio buttons. De traditionele niet-schaalbare radiobuttons genereerden dezelfde antwoorden als de buttons waarbij op het hele veld geklikt kon worden (wat makkelijker is op kleine schermen zoals de mobiele telefoon).

Respondenten evalueerden de 11-puntsschaal beter dan de 5- en 7-puntsschaal. Aangezien de literatuur suggereert dat een 11-puntsschaal beter presteert qua betrouwbaarheid en variantie, suggereren deze resultaten een voorkeur voor een langere antwoordschaal.

In ons experiment gebruikten we exact dezelfde lengte voor de 5-, 7-, en 11-puntsschaal (zie de Appendix). Vaak zijn antwoordformats anders geprogrammeerd en zorgt een langere antwoord schaal voor een langer antwoordveld. Een sterker effect kan dan ook verwacht worden in situaties waar een langere antwoordschaal ook een langer antwoordveld genereert.

Dit onderzoek laat zien dat respondenten het invullen van vragenlijsten op mobiele telefoons lager waarden dan het invullen van vragenlijsten op andere apparaten. Echter, mensen met meer ervaring met het invullen van vragenlijsten op mobiele telefoons zijn positiever, dus de gebruiksvriendelijkheid van mobiele telefoons voor het invullen van vragenlijsten is voornamelijk een kwestie van gewoon doen.

6. IMPLICATIES VOOR MARKETING EN MARKTONDERZOEK

Met een groeiende diversiteit in apparaten waarmee respondenten vragenlijsten invullen, is het voor marktonderzoekers belangrijk om goed na te denken over antwoordformats. Dit onderzoek laat zien dat sliderbalken, die werken met een sleep-principe- geen equivalente meetinstrumenten zijn met radio buttons of Visual Analogue Scales (VAS) met klikprincipe. Aangezien traditionele radio buttons niet schaalbaar zijn, is het van belang in het tijdperk van de mobiele telefoon te werken met schaalbare buttons (“big buttons”) of met VAS, die efficiënt met ruimte om kunnen gaan. Dit kan wel gevolgen hebben voor nonrespons en de evaluatie van de vragenlijst vanuit het oogpunt van de respondent. Het toevoegen van evaluatievragen is nuttig om na te gaan hoe respondenten vragenlijsten ervaren. Zodoende kan de gebruiksvriendelijkheid geoptimaliseerd worden en de nonrespons beperkt.

NOTEN

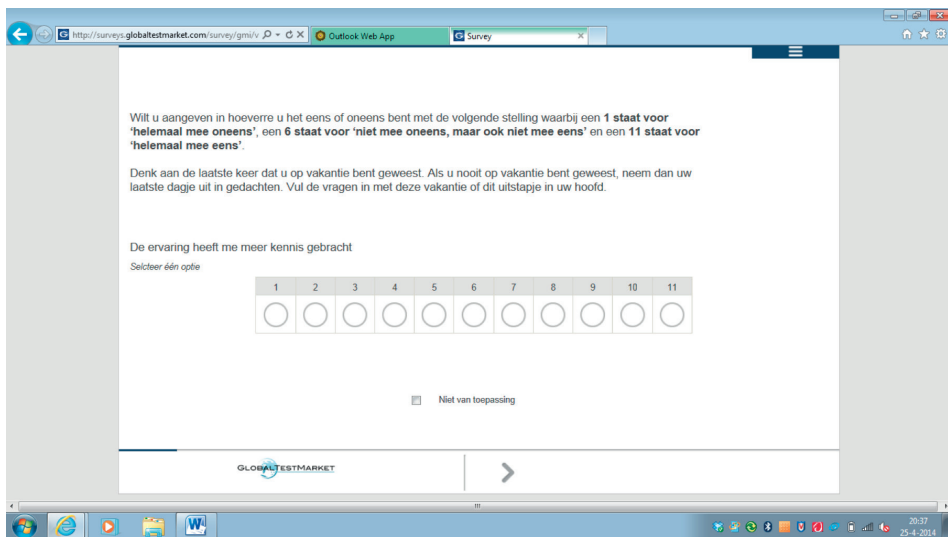
- ¹ Ik wil graag GfK en in het bijzonder Randy Streng en Frans Louwen bedanken voor hun hulp bij de dataverzameling. Ook wil ik graag Frederik Funke bedanken voor zijn hulp bij het ontwerpen van het experiment en in het bijzonder de antwoord formats.
- ² V.toepoel@uu.nl. Universiteit Utrecht. Padualaan 14. Postbus 80140 3508TC Utrecht. Tel: 030-2539075
- ³ We gebruikten ook een format met een continue schaal, maar deze resultaten worden niet gerapporteerd in dit paper.

LITERATUUR

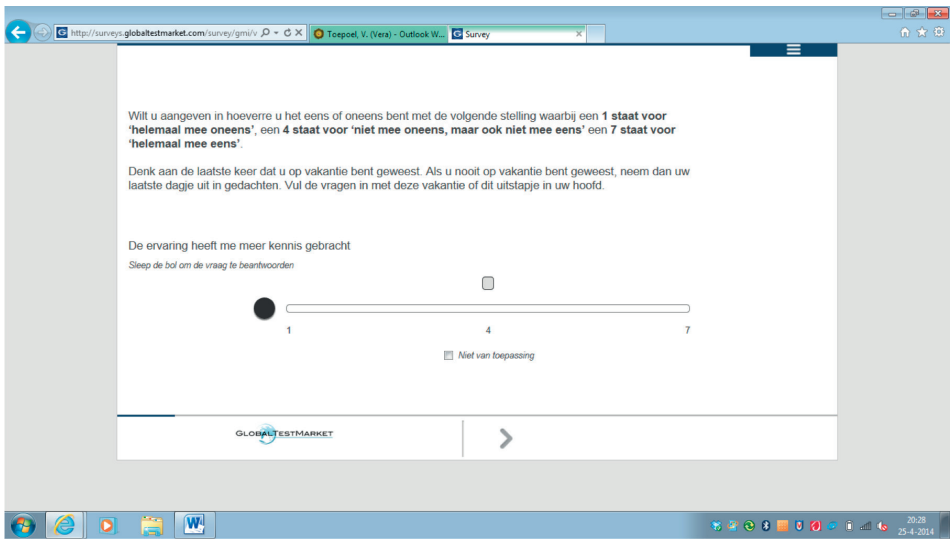
- Choudhury, S. & D. Bhattacharjee (2014) Optimal Number of Scale Points in Likert Type Scales for Quantifying Compulsive Buying Behaviour. *Asian Journal of Management Research*, 4, 432-440.
- Couper, M. P. (2011) The Future of Modes of Data Collection. *Public Opinion Quarterly*, 75, 889-908.
- Couper, M.P., F. Tourangeau, F. Conrad & E. Singer (2006). Evaluating the effectiveness of visual analog scales: A web experiment. *Social Science Computer Review*, 24, 227-245.
- Couper, M.P., F. Tourangeau, F. Conrad & S.D. Crawford (2004) What They See is What We Get. Response Options for Web Surveys. *Social Science Computer Review*, 22, 111-127.
- De Beuckelaer, Al, S. Toonen & E. Davidov (2013) On the Optimal Number of Scale Points in Graded Pair Comparisons. *Qual Quant*, 47, 2869-2882.
- De Bruijne, M. & A. Wijnant (2013a) Can Mobile Web Surveys Be Taken on Computers? A Discussion on a Multi-Device Survey Design. *Survey Practice*, 6, 1-8.
- De Bruijne, M. & A. Wijnant (2013b) Comparing Survey Results Obtained via Mobile Devices and Computers: An Experiment With a Mobile Web Survey on a Heterogeneous Group of Mobile Devices Versus a Computer-Assisted Web Survey. *Social Science Computer Review*, 1-23.
- De Bruijne, M and A. Wijnant (2014) Improving Response Rates and Questionnaire Design for Mobile Web Surveys. *Public Opinion Quarterly*, 78, 4, 951-962.
- De Bruijne, M. & A. Wijnant (2014b) Mobile Response in Web Panels. *Social Science Computer Review*, 1-15.
- Funke, F. (2015) A Web Experiment Showing Negative Effects of Slider Scales Compared to Visual Analogue Scales and Radio Buttons. *Social Science Computer Review*, 1-11.
- Funke, F & R.-D. Reips. (2012) Why Semantic Differentials in Web-Based Research Should Be Made from Visual Analogue Scales and Not from 5-Point Scales. *Fieldmethods*, 24, 310-327.
- Funke, F., U.-D. Reips & R.K. Thomas. (2011) Sliders for the Smart: Type of Rating Scale on the Web Interacts With Education Level. *Social Science Computer Review*, 29, 221-231.
- Jenkins, C.R. & D.A. Dillman. (1995) Towards a Theory of Self-Administered Questionnaire Design." In Llyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz & Trewin (Eds.) *Survey Measurement and Process Quality* (pp. 165-196). New York: Wiley.
- De Leeuw, E.D. (2010) Passen en meten online: de kwaliteit van Internet-enquetes. In A.E. Bronner, P. Dekker, E. de Leeuw, L.J. Paas, K. de Ruyter, A. Smidts, J.E. Wieringa. *Ontwikkelingen in het Marktonderzoek 2010*. Haarlem: SpaarenHout, pp 9-23.
- Maitland, A. (2009) How Many Scale Points Should I Include for Attitudinal Questions? *Survey Practice*, 2, 1-2.
- Miller, G.A. (1956) The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63, 81-97.
- Poggio, T. M. Bosnak & K. Weyandt (2013) Online Survey Participation via Mobile Devices: Implications for Nonresponse. Paper Presented at the ESRA Conference, Ljubljana, July 15-19, 2013.
- Olivier, A.J. (2008) Mixed-mode. De aanstaande revolutie binnen het marktonderzoek. *CLOU*, nr. 39, p48.
- Preston, C.C. & A.M. Colman (2000) Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, and Respondent Preferences. *Acta Psychologica*, 104, 1-15.

- Sikkel, D., R. Steenbergen & S. Gras (2014) Clicking vs. Dragging. Different Uses of the Mouse and Their Implications for Online Surveys. *Public Opinion Quarterly*, 78, 177-190.
- Sikkel, D., S. Gras, R. Steenbergen & P. Oosterveld (2014) Klikken of slepen: een vergelijking tussen vraagformats, gebaseerd op een verschillend gebruik van de muis. In A.E. Bronner, P. Dekker, E. de Leeuw, L.J. Paas, K. de Ruyter, A. Smidts, J.E. Wieringa. *Ontwikkelingen in het Marktonderzoek 2014*. Haarlem: SpaarenHout, pp 85-102.
- Smith, T. (1995) Little Things Matter: A Sampler of How Differences in questionnaire Format Can Affect Survey Responses.” In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 187-2010. New York/Berlin: Springer-Verlag.
- Toepoel, V., & Dillman, D.A. (2011). Words, numbers, and visual heuristics in web surveys. *Social Science Computer Review*, 29(2), 193-207.
- Toepoel, V., Das, J.W.M., & Soest, A.H.O. van (2009). Design of web questionnaires: The effect of layout in rating scales. *Journal of Official Statistics*, 25(4), 509-528.
- Tourangeau, R., M. Couper, and F.G. Conrad (2013) “Up Means Good” The Effect of Screen Position on Evaluative Ratings in Web Surveys. *Public Opinion Quarterly*, 77, 69-88.
- Weathers, D., S Sharma & R.W. Niedrich (2005) The Impact of the Number of Scale Points, Dispositional Factors, and the Status Qua Decision Heuristic on Scale Reliability and Response Accuracy. *Journal of Business Research*, 58, 1516-1524.

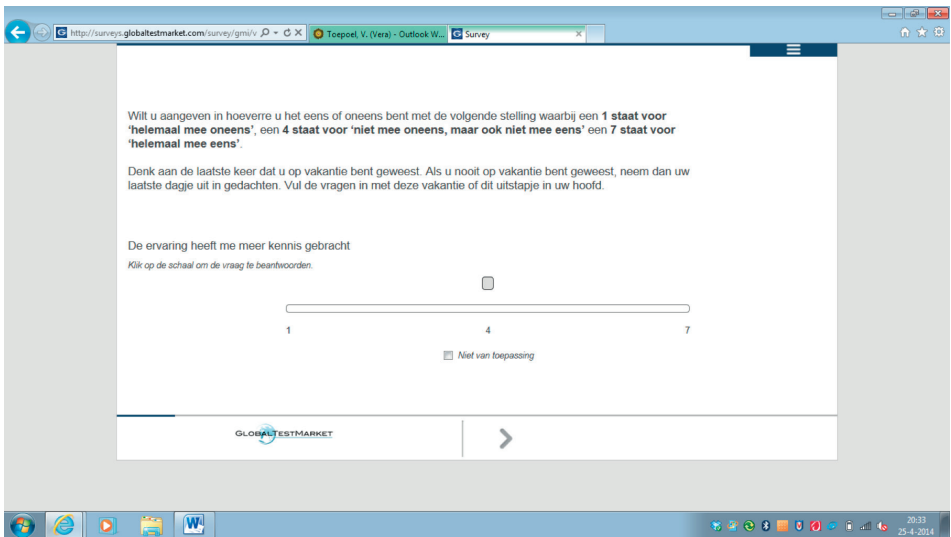
APPENDIX A: SCREENSHOTS VAN DE VERSCHILLENDE EXPERIMENTELE CONDITIES



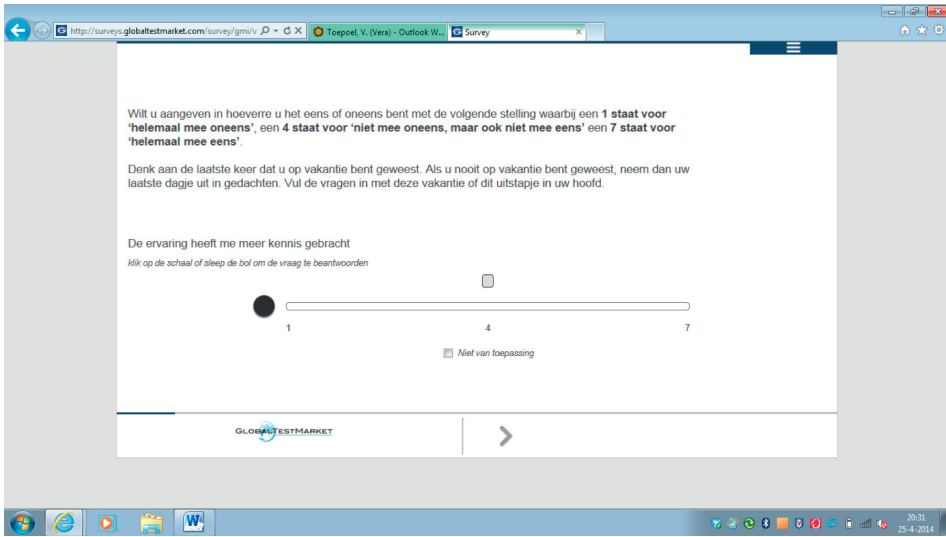
Figuur 1. Desktop, Radiobuttons, 11-puntsschaal.



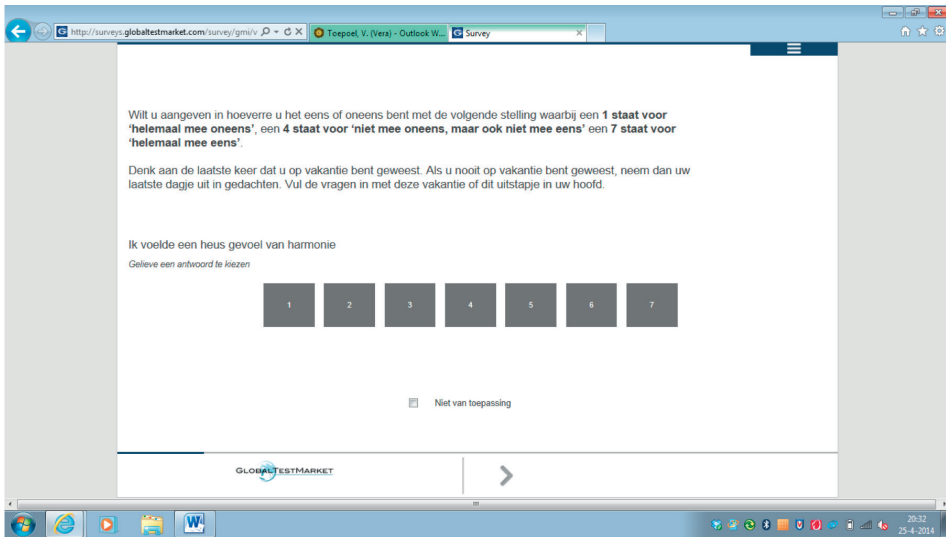
Figuur 2. Desktop, Slider bar, 7-puntsschaal. Sleep-principe. Hendel links buiten de antwoord-schaal. Zelfde lengte als de 11 puntsschaal met radio buttons in Figuur 1.



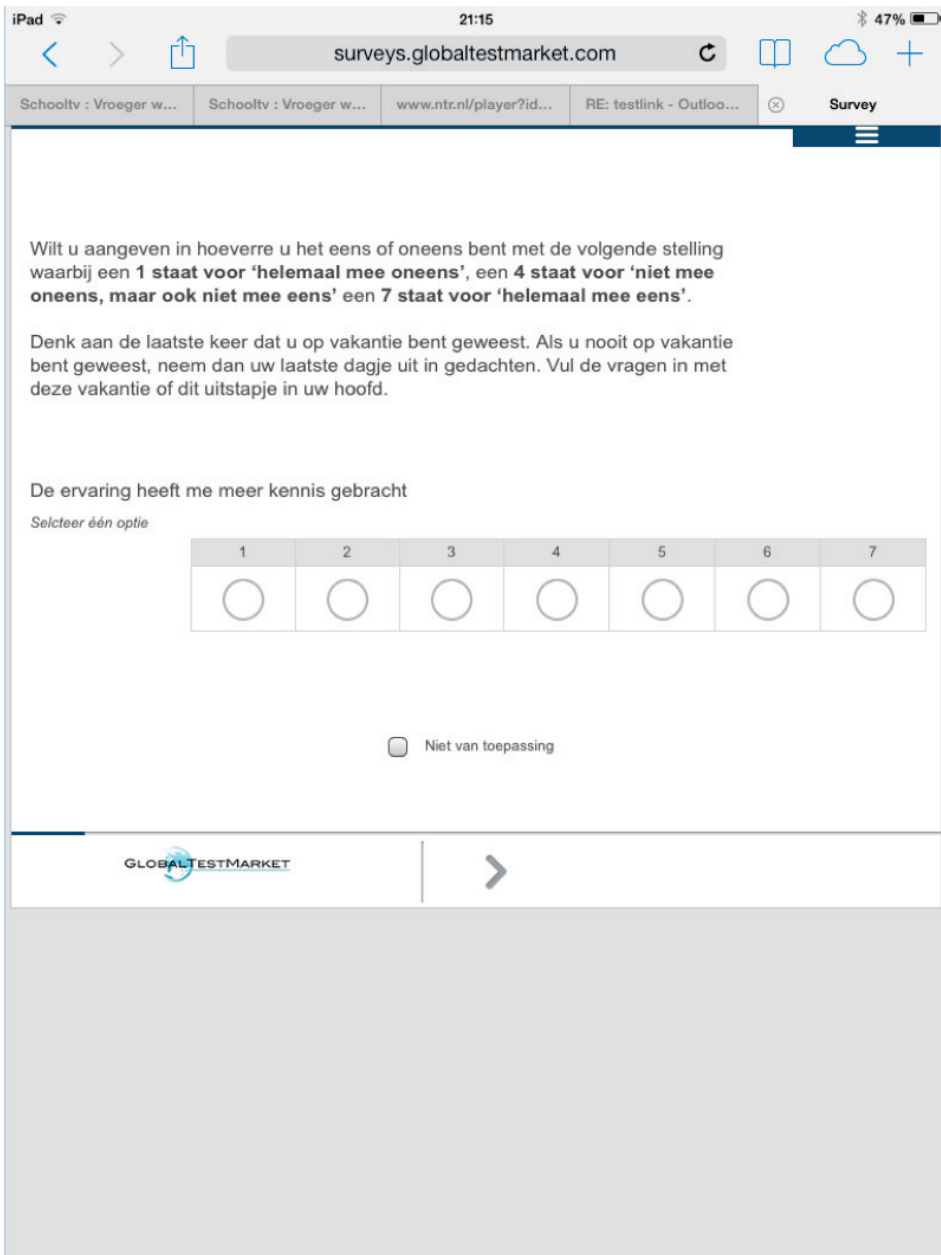
Figuur 3. Desktop, VAS, 7-puntsschaal. Klik-principe, geen hendel.



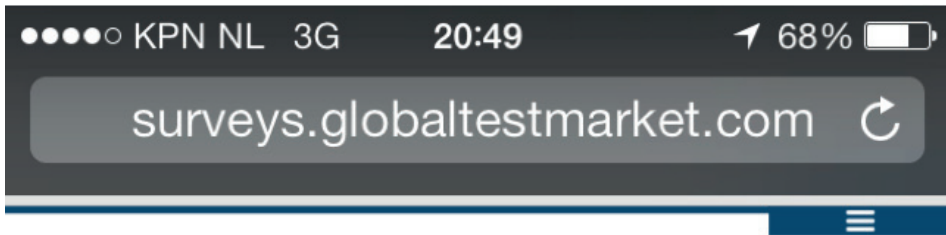
Figuur 4. Desktop, Combinatie VAS/Slider, 7-puntsschaal. Werkt met zowel klik als sleep principe.



Figuur 5. Desktop, Big buttons, 7-puntsschaal.



Figuur 6. Tablet, radio buttons, 7-puntsschaal.



Wilt u aangeven in hoeverre u het eens of oneens bent met de volgende stelling waarbij een **1 staat voor 'helemaal mee oneens'**, een **4 staat voor 'niet mee oneens, maar ook niet mee eens'** een **7 staat voor 'helemaal mee eens'**.

Denk aan de laatste keer dat u op vakantie bent geweest. Als u nooit op vakantie bent geweest, neem dan uw laatste dagje uit in gedachten. Vul de vragen in met deze vakantie of dit uitstapje in uw hoofd.

De ervaring heeft me meer kennis gebracht

Sleep de bol om de vraag te beantwoorden



Niet van toepassing



Figuur 7. Mobiele telefoon , Slider, 7-puntsschaal.