

Argumentkwaliteit in experimenteel onderzoek

Een corpusonderzoek naar de verschillen tussen sterke en
zwakke argumentatie

Robert Weijers

Communicatie- en Informatiewetenschappen

S4224574

Prof. dr. H. Hoeken

Dr. J. Hornikx

A. Weerman, MSc

Inhoudsopgave

Abstract	Pagina 3
Theoretisch kader	Pagina 4
1. Het Elaboration Likelihood Model	Pagina 4
2. Argumentkwaliteit: Problemen met operationalisatie	Pagina 5
3. Argumentkwaliteit: De invloed van argumentstructuur	Pagina 6
4. Argumentkwaliteit: De invloed van argumentinhoud	Pagina 7
5. Hoofdvraag	Pagina 9
Methode	Pagina 10
1. Materiaal	Pagina 10
2. Procedure	Pagina 11
3. Statistische toetsing	Pagina 15
Resultaten	Pagina 16
Conclusie	Pagina 18
1. Het prototypische argumentatiekoppel	Pagina 18
2. Ongeschikte manipulaties	Pagina 19
Discussie	Pagina 22
Referentielijst	Pagina 24
Appendix A. Codeboek	Pagina 26
Appendix B. Frequentietabellen	Pagina 27
Appendix C. Pretest en argumentatiemaker	Pagina 29

Abstract

Veel onderzoek is gedaan op basis van het Elaboration Likelihood Model naar de verschillen in overtuigingskracht tussen sterke en zwakke argumentatie. Met behulp van gemanipuleerd materiaal zijn hier experimenten naar gedaan. Na verloop van tijd ontstond er echter kritiek op de manier waarop deze manipulaties plaatshadden; namelijk dat hier geen normatief proces voor was, en de vraag kwam naar voren wat nu een zwak argument zwak maakt, of een sterk argument sterk. In dit corpusonderzoek wordt gekeken naar gebruikte argumentatieparen (sterk en zwak) in eerder gebruikte onderzoeken om uit te zoeken op basis waarvan sterk van zwak verschilde.

Het prototypische argumentatiekoppel bleek samengesteld te zijn door de onderzoekers met behulp van aspectwaardering op Likertschalen van proefpersonen. Dit koppel bestaat uit drie sterke en drie zwakke argumenten, die zowel dezelfde als verschillende gevolgen van een product ondersteunden. De sterke argumentatie ondersteunt wenselijke gevolgen, en de zwakke argumentatie ondersteunt gevolgen die wel positief zijn, maar minder wenselijk dan die van de sterke argumentatie. Er wordt geen gebruik gemaakt van waarschijnlijkheidsmanipulatie.

Er bleek in bijna een derde van de materialen onder de noemer 'zwak argument' negatieve gevolgen te zijn ondersteund. Bij nader onderzoek van deze specifieke materialen bleek ook de sterke argumentatie niet alleen betere, maar ook méér argumenten te bevatten dan hun zwakke tegenhanger. Omdat deze manipulatie van 'sterk-zwak' ongeschikt is, moeten wellicht de resultaten van de desbetreffende onderzoeken opnieuw worden bekeken.

Naar aanleiding van dit resultaat zijn suggesties voor vervolgonderzoek te geven. Allereerst is dat het op argumentniveau en op grotere schaal uitvoeren van een soortgelijk corpusonderzoek. Daarnaast is het aan te raden de experimenten waarin tegenargumentatie voorkwam opnieuw uit te voeren, na het verwijderen van de ongeschikte manipulaties, en de resultaten te vergelijken met het oorspronkelijke experiment.

Theoretisch kader

Al sinds de oudheid is de mensheid gefascineerd door het overtuigen van zijn medemens met woorden alleen. Het is ook iets wonderlijks om iemands mening of zelfs gedrag radicaal te kunnen aanpassen, en dat alles zonder daadwerkelijk iets aan de wereld te veranderen. Het is dan ook niet verwonderlijk dat er grote interesse is voor dit onderwerp, ook vanuit de wetenschap. Modern persuasieonderzoek kijkt naar het overtuigingsproces en probeert uit te vinden wat overtuigend werkt en waarom.

1. Het Elaboration Likelihood Model

Binnen persuasieonderzoek kan je niet om het *Elaboration Likelihood Model* (ELM) van Petty en Cacioppo (1986) heen. Dit model gaat van het volgende uit: de ontvanger van een boodschap kan een boodschap niet op één manier, maar op twee manieren verwerken (het zogeheten *dual process* model). Deze twee manieren heten de centrale en de perifere route, waarbij respectievelijk veel en weinig moeite in het verwerkingsproces wordt gestopt.

Bij de centrale route worden argumenten afgewogen en op basis hiervan wordt een attitude gevormd. In de perifere route daarentegen wordt aan de hand van enkele vuistregels een attitude gevormd. Enkele voorbeelden hiervan zijn de aantrekkelijkheidsvuistregel ("de bron is aantrekkelijk, dus wat gezegd wordt zal wel kloppen") en de meer-argumentenvuistregel ("er staan veel argumenten, dus het zal wel kloppen") (Hoeken, Hornikx & Hustinx, 2012).

Er wordt alleen centraal verwerkt als de verwerker zowel geïnteresseerd als capabel is hiervoor. Als de verwerker niet geïnteresseerd, capabel of zelfs allebei niet is, dan wordt een oordeel geveld op basis van de heuristische cues (Hoeken, Hornikx & Hustinx, 2012).

Bij centrale verwerking wordt, zoals gezegd, kritisch gekeken naar de gebruikte argumentatie. De verwerker reageert op deze argumentatie door hier een waarde aan te geven, positief of negatief. Dit verschilt per verwerker. Zo kan een argument als "Het hotel is erg kindvriendelijk!" bij de ene de cognitieve respons "Mooi, dan is het geschikt voor ons hele gezin" teweeg brengen, maar juist bij een ander "Vervelend, dat zal wel een boel herrie en drukte betekenen". De waardering, het vertrouwen in de juistheid, en de hoeveelheid van de gegeven argumenten zijn van grote invloed op de cognitieve reacties en, als gevolg daarvan, op attitude (Petty & Cacioppo, 1986).

Centrale verwerkers hebben een sterkere voorkeur voor iets wat met sterke argumenten wordt aangeprezen dan voor iets wat met zwakke argumenten wordt aangeprezen.

Argumentkwaliteit is dus van significant belang bij attitudeverandering bij centrale verwerking (Petty & Cacioppo, 2005).

2. Argumentkwaliteit: problemen met operationalisatie

Sterk en zwak verschillen dus van elkaar in argumentkwaliteit. De vraag rijst dan wat argumentkwaliteit precies is, en hoe het geoperationaliseerd kan worden. In de literatuur komen verschillende operationalisaties voor, gebaseerd op criteria uit de logica (bijvoorbeeld McGuire, 1960), of empirische criteria (zoals Petty & Cacioppo, 1986), maar niet één universeel gehanteerde operationalisatie.

O'Keefe en Jackson (1995) constateerden, door dit gebrek aan een grote analyse waar definitieve, normatieve criteria op gebaseerd kunnen worden, ook een wildgroei aan soorten pretesten die moesten meten of argumentkwaliteit succesvol gemanipuleerd waren - allemaal uitgaand van verschillende operationalisaties van argumentkwaliteit. O'Keefe en Jackson onderscheidden drie soorten pretesten, waar zij alle drie kritiek op leverden, namelijk *thoughtlisting*, kwaliteitsbeoordeling, en onsystematische variaties.

Thoughtlisting is het houden van een pretest door proefpersonen te vragen naar hun mening terwijl ze argumentatie lezen, en ze als sterker te tellen naarmate ze meer positieve respons opleveren. Deze manier meet echter, volgens O'Keefe en Jackson (1995), niet zozeer de argumentkwaliteit zelf, maar is een operationalisatie op basis van het effect ervan, wat een stap verder is.

Kwaliteitsbeoordeling houdt in dat participanten gevraagd worden uit een lijst argumenten de beste te kiezen. Nadeel hiervan is dat het ervan uitgaat dat perceptie van leken van argumentkwaliteit hetzelfde is als de argumentkwaliteit. Hoewel ook leken er toe in staat zijn op basis van intuïtie een onderscheid te maken tussen sterke en zwakke argumenten (van Eemeren et al., 2009) en zelfs onbewust normen hebben waar zij argumentaties aan toetsen (Schellens et al., 2014), is het een te grote stap om argumentkwaliteit te definiëren als wat leken als zodanig percipiëren.

Onsystematische variaties is de laatste categorie van O'Keefe en Jackson (1995). Hierbij brachten onderzoekers op basis van hun intuïtie of een zelfgekozen aspect (bijvoorbeeld "logische geldendheid") zelf manipulaties aan in het materiaal. Dit is problematisch omdat dit

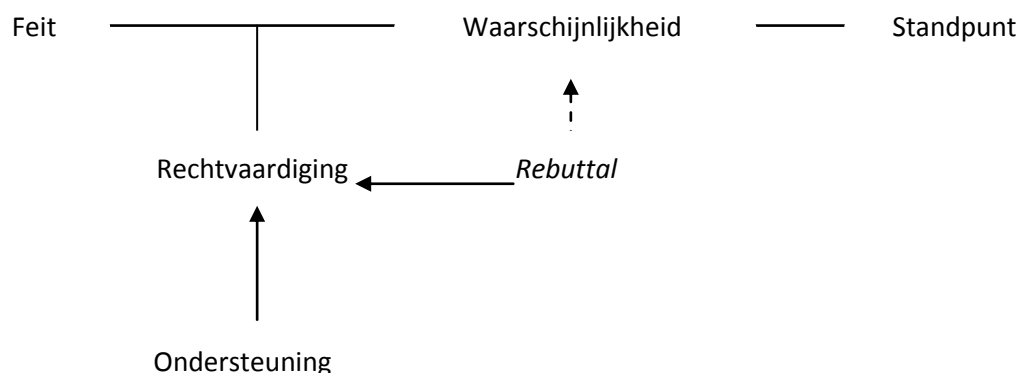
niet op basis van algemeen gehanteerde normatieve criteria gebeurd is, en omdat argumentkwaliteit niet uit één, maar uit een samenhang van meerdere aspecten bestaat.

Om argumentkwaliteit te kunnen operationaliseren moet gekeken worden naar het actieve ingrediënt tussen sterk en zwak (O'Keefe & Jackson, 1995); namelijk wat het verschil maakt tussen de twee. O'Keefe en Jackson merkten zelf al op dat dit niet één aspect zal zijn, maar meerdere aspecten. Zoals gezegd is er niet één normatieve definitie, maar er zijn wel factoren te identificeren, betreffende structuur en inhoud, die wellicht onderdeel van het 'actieve ingrediënt' zijn. In dit theoretisch kader wordt gekeken naar de invloed van zowel de structuur als de inhoud van het argument.

3. Argumentkwaliteit: de invloed van argumentstructuur

Het Toulmin-model wordt veelvuldig gebruikt binnen de argumentatieleer (Boller, Swasy & Munch, 1990). Het Toulmin-model bestaat uit zes onderdelen. Ten eerste het *standpunt*, dat verdedigd moet worden door argumentatie. Vervolgens een *feit*, dat ter ondersteuning moet dienen van het standpunt, en een *rechtvaardiging*, iets wat de link tussen het feit en het standpunt legt. Daarna kan een rechtvaardiging aan kracht winnen door een *ondersteuning*: een feit dat de rechtvaardiging ondersteunt. Er is in het model nog plaats voor de *rebuttal*, een feit wat de rechtvaardiging zou kunnen ondergraven, en, op basis van de ernst van de rebuttal, de *waarschijnlijkheid*, een aanduiding van hoe waarschijnlijk het standpunt is. (Toulmin, 1958; in Schellens & Verhoeven, 1994, p25-27)

Hieronder een visuele representatie van de onderlinge relaties in het model.



Figuur 1. Het Toulmin-model (Gebaseerd op Toulmin, 1958; in Schellens & Verhoeven, 1994, p.25-27)

Of het Toulminmodel compleet is, is van belang. Een argumentatieschema moet kloppend zijn, anders is het bijbehorende standpunt niet verdedigbaar (Van Eemeren, Snoeck & Henkemans, 2006, p. 192). Het blijkt ook uit een evaluatie van de studie van Areni en Lutz (1988) gedaan door Boller et al. (1990) dat het argumentatieschema van sterke argumenten compleet zijn, maar dat bij zwakke argumenten vaak de ondersteuning wordt weggelaten.

4. Argumentkwaliteit: de invloed van argumentinhoud

Schellens en Verhoeven (1994) stelden een model op met zes verschillende typen argumentatie. Dit theoretisch kader beperkt zich vanwege het gebruikte materiaal tot één type van Schellens en Verhoeven, de pragmatische argumentatie, ook wel argumentatie op basis van voor- en nadelen genoemd.

Veel onderzoekers rondom het thema argumentkwaliteit hebben gebruikt gemaakt van dit soort argumentatie (Petty & Cacioppo, 1986; Areni & Lutz, 1988; Van Enschoot, Hustinx en Hoeken, 2003). Ook komt deze argumentatiesoort veelvuldig voor in overtuigende brochures (Schellens & de Jong, 2000). In pragmatische argumentatie wordt een actie aanbevolen of afgeraden op basis van een positief of een negatief gevolg, dus deze aanpak ligt voor reclame of voorlichting voor de hand.

Omdat met argumentatie geprobeerd wordt attitudeverandering teweeg te brengen is het nuttig te kijken naar attitudemodellen. Een attitude bestaat volgens het model van Fishbein en Ajzen (1975), uit twee componenten; namelijk de *valence* en *strength* van het gegeven product of dienst. Deze definitie met deze componenten wordt ook in andere attitudemodellen gebruikt (Fishbein & Yzer, 2003; Witte, 1998). In deze modellen worden *valence* en *strength* respectievelijk wenselijkheid en waarschijnlijkheid genoemd. Om attitude te veranderen zal pragmatische argumentatie dan ook proberen de waargenomen waarschijnlijkheid of wenselijkheid van een voordeel of nadeel te veranderen.

Wenselijkheid

Wenselijkheid is te omschrijven als de mate waarin de verwerker een gevolg als positief ervaart: hoe positiever de respons, hoe hoger de wenselijkheid.

Niet elk gevolg levert een even positieve respons op. Er is een onderscheid te maken tussen primaire en secundaire overtuigingen. De primaire overtuigingen zijn een beperkte groep overtuigingen over de gevolgen van een product of dienst die daadwerkelijk invloed hebben

op attitudevorming (Hoeken, Hornikx en Hustinx, 2012). De secundaire overtuigingen zijn daarvoor minder relevant. Informatie over primaire overtuigingen wordt dan ook gemist als deze ontbreekt, maar informatie over secundaire overtuigingen niet (Jaccard en Wood, 1988; in Hoeken, Hornikx en Hustinx, 2012).

Het veranderen van een primaire overtuiging kan attitudeverandering tot gevolg hebben (Hoeken, Hornikx en Hustinx, 2012). Idealiter richt een argumentatie zich dus op het veranderen van primaire overtuigingen. Het richten van argumentatie op primaire argumenten zorgt namelijk voor een positievere respons dan wanneer er op secundaire argumenten wordt gericht (Hoeken, 1995).

Onderscheid tussen deze soorten overtuigingen over een product of dienst wordt vaak gedaan door een vooronderzoek (Hoeken, Hornikx en Hustinx, 2012).

Waarschijnlijkheid

Waarschijnlijkheid beslaat de inschatting van de verwerker van wat de kans is dat het genoemde gevolg ook daadwerkelijk zal optreden. Hoe groter die kans wordt ingeschat, hoe hoger de waarschijnlijkheid.

Ter ondersteuning van een waarschijnlijkheidsargument wordt vaak evidentie aangehaald, dat dient om het wenselijke gevolg waarschijnlijker te maken. Een onderzoek van Hoeken en Hustinx (2002) maakt een onderscheid tussen vier soorten evidentie: anekdotische evidentie (wanneer een specifiek voorbeeld wordt aangehaald), statistische evidentie (getalsmatige samenvatting van specifieke gevallen), causale evidentie (voorspelling op basis van oorzaak-gevolg), en autoriteitsevidentie (door het aanhalen van een deskundige op het gebied). Uit het onderzoek bleek dat anekdotische evidentie minder effectief was dan de andere drie. Hornikx (2005) bevestigde dit.

Waarschijnlijkheid is echter moeilijk te herkennen, zo bleek na evaluatie van het onderzoek Areni en Lutz (1988) door van Enschoot, Hustinx en Hoeken (2003). Leken hadden daar in eerste instantie geen waarschijnlijkheidsmanipulatie in herkend (Areni & Lutz, 1988), maar het materiaal bleek dat wel te bevatten (van Enschoot, Hustinx en Hoeken, 2003). Dit komt overeen met de bevindingen van Hustinx, Enschoot en Hoeken (2007; in Hoeken, Hornikx en Hustinx, 2012, p.122) dat leken gemakkelijker de wenselijkheid van een gevolg herkennen dan de waarschijnlijkheid ervan. Hustinx et al. (2006) bevestigden ook dat manipulatie van alleen waarschijnlijkheid inderdaad geen invloed had op de attitude.

Verhouding wenselijkheid/waarschijnlijkheid

Verskillende onderzoeken tonen aan dat een waarschijnlijkheidsargument vaker aanvullende evidentie meekrijgt dan een wenselijkheidsargument (Hornikx, Starren & Hoeken, 2003, Schellens & de Jong, 2004; in Hoeken et al., 2012, p. 123). Wenselijkheid is namelijk vaak op zich al duidelijk: waarom een betere gezondheid wenselijk is behoeft geen argumentatie, waarom meer vis eten die gezondheid waarschijnlijker maakt wel. Alles wijst er dus op dat wenselijkheid vaker herkend wordt en dus meer effect lijkt te hebben op argumentkwaliteit dan waarschijnlijkheid. Dat ze wel allebei invloed hebben, bleek uit een onderzoek van Hoeken, Timmers en Schellens (2012).

5. Hoofdvraag

Zoals hierboven vermeld zijn er veel onderzoeken geweest naar persuasie, en veel verschillende definities van argumentkwaliteit gesteld. De meeste experimenten die zijn gedaan naar persuasie en argumentkwaliteit, bevatten materialen die op veel verschillende manieren zijn samengesteld. Dit is geen ideale situatie en er is dan ook vraag naar normatieve criteria om het onderscheid tussen sterk en zwak objectiever te kunnen maken. Om dat onderscheid te kunnen maken moet onderzocht worden op welke kenmerken precies deze materialen verschillen.

De vraag is nu of er tussen deze sterke en zwakke materialen verschillen te vinden zijn. De hoofdvraag van dit onderzoek luidt dan ook:

Welke kwaliteitsverschillen tussen sterke en zwakke argumenten, zoals gebruikt in eerder empirisch onderzoek naar argumentkwaliteit, zijn er aan te wijzen?

Antwoord op deze vraag zou een concrete aanwijzing kunnen geven naar welke factoren een rol spelen in het verschil tussen een sterk en een zwak argument, wat op zijn beurt weer op talloze vakgebieden, van reclamemaker tot speechschrijver, nuttig is.

Methode

1. Materiaal

Voor het beantwoorden van de onderzoeksvraag is gekozen voor een corpusonderzoek van eerdere wetenschappelijke onderzoeken naar argumentkwaliteit in het kader van reclameonderzoek. Dit materiaal is op een systematische manier bemachtigd: er is gezocht in de literatuur naar de trefwoorden "argument quality", "argument strength", "weak arguments" en "strong arguments", in combinatie met "advertisements" en enkele synoniemen en variaties daarvan. De artikelen die bij deze zoektocht naar boven kwamen zijn doorzocht op manipulatie van argumentkwaliteit, en indien dit gebeurd was, werd het gebruikte materiaal opgenomen in het corpus. Indien het materiaal niet beschikbaar was in het artikel, werd het opgevraagd bij de auteurs.

In totaal zijn de materialen van 44 experimenten, afkomstig uit 40 artikelen, opgenomen in het corpus. Het materiaal bestond uit een verzameling sterke en een verzameling zwakke argumenten. In tabel 1 is een voorbeeld te vinden van gebruikt materiaal.

Tabel 1. Voorbeeld van gevonden materiaal (Lowrey, 1998)

<i>Sterke argumentatie</i>	<i>Zwakke argumentatie</i>
BRAN-NEW cereal is a healthy choice for breakfast because it's high in fiber. BRAN-NEW is also preservative-free. More people prefer this all-natural cereal because it has crispy flakes and tastes great. It's also naturally fat-free and cholesterol-free. Try BRAN-NEW-the new, healthy cereal.	BRAN-NEW cereal is a convenient choice for breakfast because it's available at many stores. BRAN-NEW also has lightweight packaging. More people prefer this convenient cereal because it's easy to store in small kitchens. It comes in only one small size. Try BRAN-NEW- the new, convenient cereal.

2. Procedure

In het codeboek (zie Appendix A) zijn verschillende aspecten opgenomen. Alle categorieën zijn geoperationaliseerd in het codeboek. Alle gebruikte voorbeelden komen uit de materialen.

Allereerst werd van al het gebruikte materiaal beschrijvende criteria als auteur, jaartal en het aantal woorden genoteerd.

De categorie "Argumentatiemaker" beschrijft wie de argumentatie gemaakt heeft, namelijk de onderzoekers zelf (1), participanten tijdens een pretest (2) of een ander (3). Indien door een ander gemaakt, werd de auteur zo mogelijk erbij genoteerd is.

De categorie "Pretest" is gebaseerd op O'Keefe en Jackson (1995). Er waren drie opties: de argumenten zijn gekozen door middel van thoughtlisting van proefpersonen (1), in een pretest werd geoordeeld over de argumentkwaliteit, veelal op Likert-schalen (2) of de onderzoekers brachten zelf variaties aan in de argumentatie (3).

Tabel 2. Overzicht van artikelclassificatie.

Naam categorie	Opties
Auteur	<i>Naam auteur(s)</i>
Jaartal	<i>Jaar</i>
Tijdschrift	<i>Naam tijdschrift</i>
Experimentnummer	<i>Getal</i>
Argumentatiemaker	<i>Onderzoekers, Participanten of Anders</i>
Pretest	<i>Thoughtlisting, Likertschalen of Zelfgemaakt</i>
Aantal woorden sterke argumentatie	<i>Getal</i>
Aantal woorden zwakke argumentatie	<i>Getal</i>

Hierna werden categorieën opgesteld om aspecten van de argumentatie te bepalen; alle materialen werden namelijk op studieniveau bekeken, niet op argumentniveau. Het materiaal werd verdeeld onder drie codeurs, die elk een derde codeerden, en een vierde codeur codeerde alles. Op basis hiervan werd de interbeoordelaarsbetrouwbaarheid bepaald. In tabel 3 is een overzicht te vinden van de gebruikte categorieën. Daaronder worden de categorieën toegelicht.

Tabel 3. Overzicht van argumentatieclassificatie.

Naam categorie	Opties
Geadverteerd product	<i>Product</i>
Aantal sterke argumenten	<i>Getal</i>
Aantal zwakke argumenten	<i>Getal</i>
Domein hetzelfde	<i>ja/nee (1/0)</i>
Domein verschillend	<i>ja/nee (1/0)</i>
Waarschijnlijkheid gemanipuleerd	<i>ja/nee (1/0)</i>
Waarschijnlijkheid gemanipuleerd: hoe?	<i>Tekst</i>
Wenselijk - minder wenselijk	<i>ja/nee (1/0)</i>
Wenselijk - neutraal	<i>ja/nee (1/0)</i>
Wenselijk - onwenselijk	<i>ja/nee (1/0)</i>

Na de eerste codeerronde bleek de interbeoordelaarsbetrouwbaarheid dusdanig laag dat een tweede codeerronde nodig was. Na deze tweede ronde werden de overgebleven verschillen besproken en opgelost.

Product

In deze categorie werd het aangeprezen product of de aangeprezen dienst vermeld.

Aantal argumenten

In deze categorie keken de codeurs naar hoeveel producteigenschappen (domeinen) werden beargumenteerd. Losse argumenten die op hetzelfde productdomein sloegen, werden als één argument geteld. Zo werden "affordable" en "a good choice for the price" geschaard onder hetzelfde argument, omdat zij beiden refereerden naar het productdomein 'goedkoop'.

De correlatie van het aantal sterke argumenten tussen de codeurs was na de eerste codeerronde hoog ($r(41) = .88, p < .001$) en na de tweede codeerronde hoog ($r = .98, p < .001$).

De correlatie van het aantal zwakke argumenten tussen de codeurs was na de eerste codeerronde hoog ($r(41) = .91, p < .001$) en na de tweede codeerronde hoog ($r = .98, p < .001$).

Domein

Bij deze categorie werd gekeken welke productdomeinen werden genoemd bij sterk en zwak. Deze categorisering is gebaseerd op het verschil tussen primaire en secundaire overtuigingen, om te kijken of hier ook tussen wordt gevarieerd. Als één of meerdere argumenten in de set van de sterke argumentatie en van de zwakke argumentatie over hetzelfde productdomein gaat valt dit in de categorie 'hetzelfde domein'. Een voorbeeld van 'hetzelfde domein' is over een alarmsysteem "dual sensors" (sterk) versus "single sensor" (zwak).

Als één of meerdere argumenten van de sterke argumentatie of de zwakke argumentatie uniek is in zijn domein (en geen tegenhanger heeft in de zwakke dan wel sterke argumentatie), valt de argumentatie binnen "domein verschillend". Een voorbeeld van "domein verschillend" is, over een bril, "high technology materials" (sterk) en "free case" (zwak).

Omdat de argumentaties op studieniveau, en niet op argumentniveau, bekeken worden, sluiten deze twee categorieën elkaar niet uit.

De interbeoordelaarsbetrouwbaarheid van "Domein hetzelfde" was na de eerste codeerronde passabel ($\kappa = .69$, $p < .001$) en na de tweede ronde goed ($\kappa = .95$, $p < .001$).

De interbeoordelaarsbetrouwbaarheid van "Domein verschillend" was na de eerste codeerronde onvoldoende ($\kappa = .56$, $p < .001$) en na de tweede ronde goed ($\kappa = .95$, $p < .001$).

Ten behoeve van de resultaten is na de codering een categorie gemaakt genaamd "Domein", waarin de twee domeincategorieën zijn samengevoegd. Deze categorie heeft vier mogelijkheden: alleen domein hetzelfde (1), alleen domein verschillend (2), zowel domein hetzelfde als domein verschillend (3) en geen van beide (4).

Waarschijnlijkheid gemanipuleerd

Indien de waarschijnlijkheid van (ten minste) één argument was gemanipuleerd, viel dat als een 'ja' in deze categorie. Een voorbeeld hiervan is "A U.S. Health Survey reports..." vergeleken met "Home magazine reports...".

De interbeoordelaarsbetrouwbaarheid van "Waarschijnlijkheid gemanipuleerd" was na de eerste codeerronde onvoldoende ($\kappa = .49$, $p = .002$) en na de tweede ronde goed ($\kappa = .91$, $p < .001$).

Soort waarschijnlijkheidsmanipulatie

Als waarschijnlijkheidsmanipulatie had plaatsgevonden, werd genoteerd op welke manier dit gebeurd was in de categorie "Soort waarschijnlijkheidsmanipulatie". Dit werd woordelijk genoteerd en later ingedeeld in één van de volgende categorieën, gebaseerd op Hoeken en Hustinx (2002) en Boller et al. (1990), namelijk door manipulatie van anekdotische evidentie (1), statistische evidentie (2), causale evidentie (3) of autoriteitsevidentie (4), of door het ontbreken van evidentie (5).

Wenselijkheid

In de categorieën over wenselijkheid werd de sterke en zwakke argumentatie tegenover elkaar gezet. Het uitgangspunt was dat alles wat in de sterke argumentatie stond wenselijk was, en de vraag was hoe de zwakke argumentatie was in relatie tot het sterke. Indien er in de zwakke argumentatie argumenten voorkwamen die minder wenselijk waren dan het sterke, maar nog steeds op zichzelf positief waren, kwamen ze in de categorie "Wenselijk - minder wenselijk". Een voorbeeld over een tekstverwerker is "Over 100 typefaces included" (sterk) en "Includes standard fonts" (zwak).

Was een zwak argument niet negatief, maar ook niet positief, dan viel het in de categorie "wenselijk - neutraal". Een voorbeeld hiervan is "Requires 20M of hard disc space and 8MB of RAM", over eenzelfde tekstverwerker.

Tenslotte kwam het voor dat in de zwakke argumentatie tegenargumenten genoemd werden. Dat viel de argumentatie in de categorie "wenselijk - onwenselijk". Het voorbeeld met de tekstverwerker hierbij is "Installation from 30 floppy disks". Hoewel dit niet tevoren verwacht werd, maar toch bleek voor te komen, is deze categorie opgenomen.

De interbeoordelaarsbetrouwbaarheid van de categorie "wenselijk - minder wenselijk" was na de eerste codeerronde onvoldoende ($\kappa = .48$, $p = .001$) en na de tweede codeerronde onvoldoende ($\kappa = .63$, $p < .001$). Hier hoort echter de kanttekening bij dat 40 van de 43 antwoorden van de codeurs overeen kwamen.

De interbeoordelaarsbetrouwbaarheid van de categorie "wenselijk - neutraal" was na de eerste codeerronde onvoldoende ($\kappa = .02$, $p = .911$) en na de tweede codeerronde passabel ($\kappa = .68$, $p < .001$).

De interbeoordelaarsbetrouwbaarheid van de categorie "wenselijk - onwenselijk" was na de eerste codeerronde onvoldoende ($\kappa = .35$, $p = .015$) en na de tweede codeerronde goed ($\kappa = .78$, $p < .001$).

Een overzicht van de interbeoordelaarsbetrouwbaarheid van de categorieën is te vinden in tabel 4.

Tabel 4. *Interbeoordelaarsbetrouwbaarheid (kans tussen haakjes)*

Naam categorie	Na eerste codeerronde	Na tweede codeerronde
Aantal sterke argumenten	$r = .88$ ($p < .001$)	$r = .98$ ($p < .001$)
Aantal zwakke argumenten	$r = .91$ ($p < .001$)	$r = .98$ ($p < .001$)
Domein hetzelfde	$\kappa = .69$ ($p < .001$)	$\kappa = .95$ ($p < .001$)
Domein verschillend	$\kappa = .56$ ($p < .001$)	$\kappa = .95$ ($p < .001$)
Waarschijnlijkheid gemanipuleerd	$\kappa = .49$ ($p = .002$)	$\kappa = .91$ ($p < .001$)
Wenselijk - minder wenselijk	$\kappa = .48$ ($p = .001$)	$\kappa = .63$ ($p < .001$)
Wenselijk - neutraal	$\kappa = .02$ ($p = .911$)	$\kappa = .68$ ($p < .001$)
Wenselijk - onwenselijk	$\kappa = .35$ ($p = .015$)	$\kappa = .78$ ($p < .001$)

3. *Statistische toetsing*

Voor het verwerken van het corpus is gebruik gemaakt van een *paired-sample* T-toets, eenweg variantie-toetsen, bivariate correlaties en de hiervoor al gebruikte toets voor interbeoordelaarsbetrouwbaarheid om Cohen's kappa te berekenen.

Resultaten

Allereerst is gekeken naar het aantal woorden en argumenten, zoals te zien in tabel 5.

Tabel 5. Aantal woorden en argumenten per argumentatiesoort (SD tussen haakjes) (N = 44)

	Sterke argumentatie	Zwakke argumentatie
Aantal argumenten	5,00 (2,40)	4,82 (2,50)
<i>Range</i>	1 - 15	1 - 15
Aantal woorden	85,52 (78,48)	83,48 (82,83)
<i>Range</i>	8 - 418	8 - 421

Uit een T-toets bleek een significant verschil in het aantal argumenten tussen sterke en zwakke argumentatie ($t(43) = 2.07, p = .044$). Sterke argumentatie ($M = 5,00, SD = 2,40$) bleek meer argumenten te bevatten dan zwakke argumentatie ($M = 4,82, SD = 2,50$).

Na uitsluiting van de materialen waar het aspect "Wenselijk - onwenselijk" naar voren kwam, was er geen significant verschil tussen het aantal argumenten in de sterke en zwakke argumentatie te ontdekken ($t(29) = 1.000, p = .326$).

Uit een T-toets bleek geen significant verschil in het aantal woorden tussen sterke en zwakke argumentatie ($t(43) = .768, p = .446$).

In tabel 6 is de frequentie en het percentage te vinden van het voorkomen van een categorie.

Tabel 6. Frequenties aanwezigheid van een categorie (N = 44)

Categorie	Aantal voorkomend	Percentage voorkomend
Domein hetzelfde	30	68,2
Domein verschillend	27	61,4
Waarschijnlijkheid gemanipuleerd	21	47,7
Wenselijk - minder wenselijk	39	88,6
Wenselijk - neutraal	13	29,5
Wenselijk - onwenselijk	14	31,8

Frequenties zijn berekend van de resultaten per categorie. Een compleet overzicht is te vinden in Appendix B.

In de samengestelde categorie Domein waren de uitslagen vrij gelijk verdeeld. "Alleen hetzelfde" (36,4%) kwam het meest voor, daarna "Alleen verschillend" (31,8%). 29,5% van de materialen viel in de categorie "Beide". Tenslotte behoorde 2,3% in "Geen van beide".

In de categorie Pretest viel 61,4% van het materiaal in "Likert-schalen". Daarna volgde "Zelfgemaakte variaties" (27,3%) en Thoughtlisting (4,5%). Tenslotte viel 6,8% in de categorie "Anders".

De frequenties binnen de categorie "Argumentatiemaker" zijn berekend. Het materiaal is in 72,7% van de gevallen door onderzoekers samengesteld, in 18,2% van de gevallen door participanten in een pretest, en 9,1% viel in de categorie "Anders".

De frequenties binnen "Soort waarschijnlijkheidsmanipulatie" zijn berekend. Van het materiaal bevatte 52,3% geen waarschijnlijkheidsmanipulatie. "Autoriteitsevidentie" kwam voor in 18,2% van de gevallen voor, en "Ontbreken van evidentie" in 11,4% van de gevallen. Daarna volgde "Statistische evidentie" (15,9%), en "Anekdotische evidentie" (2,3%) en tenslotte kwam "Causale evidentie" niet voor.

Daarnaast is berekend of een verschillende pretest leidde tot een verschillende manipulatie. Dit bleek niet het geval te zijn ($p > .05$). De argumentatiemaker speelde ook geen rol bij de gevonden verschillen in de manipulatie ($p > .05$). De complete uitslagen van deze statistische analyse zijn te vinden in Appendix C.

Conclusie

Met behulp van de resultaten kan de hoofdvraag beantwoord worden: Welke kwaliteitsverschillen tussen sterke en zwakke argumenten, zoals gebruikt in eerder empirisch onderzoek naar argumentkwaliteit, zijn er aan te wijzen?

Met deze resultaten is het mogelijk het prototypische argumentatiekoppel samen te stellen, namelijk wat er over het algemeen gebruikt wordt in empirisch onderzoek naar argumentkwaliteit, maar valt er ook aan te wijzen welke ongeschikte manipulaties voorkomen bij het samenstellen van materiaal voor deze onderzoeken.

1. Het prototypische argumentatiekoppel

De prototypische argumentkwaliteitsmanipulatie is samengesteld door de onderzoekers, en de pretest is uitgevoerd door een groep participanten die op Likert-schalen de gegeven argumenten beoordeelden, waaruit de drie beste en slechtste scorende werden geselecteerd.

De sterke argumentatieset bevat drie argumenten die alle drie wenselijk zijn. De tegenhanger, de zwakke argumentatie, bevat argumenten die wel wenselijk zijn, maar vergeleken met de sterke set argumenten minder wenselijk. De productdomeinen die door de zwakke argumenten ondersteund worden, komen deels overeen met die van de sterke argumenten, maar er is ook ruimte voor een zwak argument dat een productdomein verdedigt wat niet in de sterke argumentatie verdedigd wordt. De sterke argumentatie bestaat uit ongeveer 86 woorden, en de zwakke argumentatie uit om en nabij 83 woorden.

In de prototypische argumentatiemanipulatie wordt in principe waarschijnlijkheid niet gemanipuleerd. Vindt deze waarschijnlijkheidsmanipulatie toch plaats, zoals in 47% van de gevallen, wordt er veelal gekozen voor manipulatie van autoriteit van de bron.

Een voorbeeld van deze prototypische argumentatie is te vinden in het materiaal, namelijk het tweede experiment van Chang (2012), tabel 7. Deze argumentatie prijst een mobiele telefoon aan. Dit voorbeeld past nagenoeg op de gegeven schets van 'het prototypische argumentatiekoppel', alleen zijn de onderzoekers niet de argumentatiemakers, maar hebben participanten de argumentatie gemaakt, en bevat dit voorbeeld vele malen minder woorden dan de prototypische argumentatie. Dat er geen enkele gebruikte argumentatie in het corpus precies overlapt met de mal van 'prototypische argumentatie', geeft goed aan hoeveel verscheidenheid er te vinden is in het onderzochte materiaal.

Tabel 7. Materiaal gebruikt door Chang (2012)

<i>Sterke argumentatie</i>	<i>Zwakke argumentatie</i>
Good design, built-in high-pixel camera, quick repair service	Two display windows, exchangeable front and back covers, clamshell design

2. Ongeschikte manipulaties

Allereerst worden iets meer argumenten gebruikt in sterke argumentatie dan bij hun zwakke tegenhanger binnen hetzelfde experiment. Dit is ongewenst, omdat het enige beoogde verschil tussen de sterke en zwakke argumentatie de kwaliteit ervan zou moeten zijn. Als de sterke argumentatie niet alleen kwalitatief hoger is, maar ook meer argumenten bevat, is het dus niet uit te sluiten dat de oorzaak van de gevonden resultaten niet het gevolg zijn van kwaliteitsmanipulatie, maar van onbedoelde manipulatie van het aantal argumenten.

Vervolgens bleek al tijdens het opstellen van het codeboek bleek dat sommige onderzoeken op een ongeschikte manier manipulaties gedaan hadden van hun materiaal. Sommige onderzoeken hebben namelijk niet zozeer sterke met zwakke argumenten vergeleken, maar ook sterke argumenten met tegenargumenten. Dit kwam terug in de categorie "Wenselijk - onwenselijk", wat in bijna een derde (31,8%) van de onderzochte materialen naar voren kwam. Dit is natuurlijk ongewenst, omdat een tegenargument natuurlijk heel snel een negatievere attitude veroorzaakt dan een zwak, maar nog steeds positief argument. Een dergelijke zwakke argumentatie is geen teken van een manipulatie tussen sterk en zwak, maar van voor en tegen.

Een extreem voorbeeld van het voorkomen van ongeschikte manipulatie, is die zoals samengesteld in het experiment gedaan door Christensen, Ascione en Bagozzi (1997), te vinden in tabel 8. Bij dit experiment bestonden zowel de sterke als zwakke argumentatie uit tegenargumenten; het betrof een bijsluiter van medicijnen.

Tabel 8. Materiaal gebruikt door Christensen, Ascione en Bagozzi (1997)

<i>Sterke argumentatie</i>	<i>Zwakke argumentatie</i>
<p>Side effects should be considered. Only your doctor can tell if Dirovin is right for you. Be sure to tell you doctor if you are allergic to any medications. Along with its intended effects Dirovin may cause unwanted effects. Many are minor, such as heartburn, indigestion, constipation or diarrhea. However, if you notice any unusual reactions or side effects, check with your doctor or pharmacist</p>	<p>Side effects should be considered. Dirovin may cause a serious allergic reaction called anaphylaxis. In some cases this has been fatal. Be sure to tell your doctor if you are allergic to any medications. Other serious side effects such as stomach or intestinal bleeding, ulceration and peroration may occur with or without warning during long-term use of Dirovin. Confusion, forgetfulness, mental depression, and other mental and mood changes have also been reported with Dirovin use.</p>

Wat door deze onderzoekers als zwak is geclassificeerd, levert een sterkere respons op dan het als sterk geclassificeerde. Alleen omdat het beide tegenargumenten zijn, is de eigenlijk zwakkere, maar als sterk geclassificeerde, argumentatie wenselijker is dan de andere. Dit past zelfs niet in de categorie "Wenselijk - onwenselijk" omdat het als sterk geclassificeerde argument ook onwenselijk is, alleen in mindere mate. Dit onderzoek zegt ook resultaten te geven over sterke en zwakke argumentatie, maar doet dit met radicaal anders materiaal dan het hiervoor opgestelde 'prototype'. Dit radicaal anders opgestelde materiaal kan dus ook zorgen voor radicaal andere resultaten.

De aanwezigheid van materiaal met 'Wenselijk - onwenselijk' erin voorkomend is dus op zichzelf al ongewenst, en de implicaties hiervan zijn dat bijna een derde (31.8%) van de experimenten ongeschikte manipulaties bevat. Dit materiaal moet dus opnieuw worden bekeken. Hiervoor zijn de criteria waar O'Keefe en Jackson (1995) om vragen mogelijk een nuttig hulpmiddel. Door een vaste set regels, een procedure of normatieve criteria bij het opstellen van argumentatie voor een experiment is wellicht een wijdverspreide fout als deze voortaan te voorkomen.

Na weglating van het ongeschikte materiaal, bevatte de sterke argumentatie niet meer of minder argumenten dan haar zwakke tegenhanger. Het lijkt er dus op dat 68.2% van het materiaal geschikte manipulaties bevat.

Er was met behulp van dit onderzoek geen oorzaak aan te wijzen van deze wijdverspreide ongeschikte manipulaties. Zowel de argumentatiemaker als de soort pretest speelden geen rol bij de aanwezigheid van de gecodeerde aspecten.

Discussie

Bij het uitvoeren van dit onderzoek zijn kanttekeningen te plaatsen.

Er dient een behoorlijke kanttekening geplaatst te worden bij de aspecten die te maken hadden met wenselijkheid. Deze aspecten werden alleen binair gecodeerd (wel of niet aanwezig), in plaats van hun aantallen. Dit kan dus betekenen dat een zwakke argumentatie van tien argumenten het stempel "aanwezig" meekreeg, maar slechts één argument bevatte dat kon worden bestempeld als "Wenselijk - onwenselijk": er werd geen onderscheid gemaakt in hoeveel argumenten in de argumentatie onder dit aspect vielen. Een zwakke argumentset waarvan alle argumenten naar neutrale gevolgen verwezen, viel dus in dezelfde categorie als een zwakke argumentatieset waarvan maar één argument naar een neutraal gevolg verwees.

De argumentatie werd dus per studie geclassificeerd, en niet per argument; dit is gedaan omdat niet bij elke argumentatie koppels te maken waren van sterke en zwakke argumenten en dus zou dat problemen opleveren bij de codering. Dit maakt de resultaten wel wat grover. Dat betekent ook dat de conclusie die net getrokken is, namelijk dat bijna een derde van de experimenten gebruik maakte van ongeschikte manipulaties, wellicht niet zo erg is als het klinkt; er bestaat de mogelijkheid dat het in deze studies slechts om één onwenselijk argument ging wat opging in vergezellende minder wenselijke of neutrale argumenten.

Vervolgens was er onenigheid over de codering, zoals al te zien aan de resultaten van de eerste codeerronde. Dit had te maken met de vele factoren die bij een argument betrokken zijn; met name context speelde hier een grote rol in. Het bleek in het bijzonder moeilijk het onderdeel "Wenselijk - neutraal" te classificeren. Een voorbeeld hiervan is te vinden in de argumentatie van Helweg-Larsen & Howell (2000) over condooms: "Prevent pregnancy and Sexually Transmitted Diseases". Dit is aan de ene kant erg wenselijk, maar aan de andere kant zou dit ook beschouwd kunnen worden als neutraal; dit is namelijk de *raison d'être* van condooms, en zonder deze eigenschap zouden ze niet verkocht worden.

Deze problemen zijn besproken en er werd naar consensus gezocht om deze categorie toch succesvol in te kunnen vullen. Toch duidt de moeite van het bereiken van een hoge interbeoordelaarsbetrouwbaarheid wellicht op een codeboek dat te subjectief of niet voldoende duidelijk is.

Er bleek tijdens het opstellen van het codeboek ook een wildgroei aan aanpakken voor het pretesten te zijn gehanteerd in de onderzochte materialen. Dit bemoeilijkte de codering. Er was echter geen manier om al deze verschillende manieren los te coderen tot een overzichtelijk, betekenisvol geheel, dus is gekozen voor de simplistische aanpak zoals die uitgevoerd is. Problematisch was ook dat in sommige onderzoeken de rapportage van de totstandkoming van de manipulatie niet of niet voldoende gerapporteerd was, wat het coderen van de pretest en argumentmaker verder bemoeilijkte.

Suggesties voor vervolgonderzoek

Gegeven de resultaten van dit onderzoek is het mogelijk suggesties te maken voor toekomstig onderzoek.

Het uitvoeren van eenzelfde onderzoek op grotere schaal is de moeite waard. De opzet van dit onderzoek was in eerste instantie groter, maar moest door problemen met beschikbaarheid van codeurs worden ingekrompen. In het verlengde hiervan ligt een meer specifieke codering waarin wellicht per argument gecodeerd wordt in plaats van per argumentatie, of tenminste een codering waarin het aantal argumenten dat voldeed aan criteria van een aspect ook naar voren kwam, in plaats van het binaire ja/nee stelsel dat nu gebruikt is.

Daarnaast is het de moeite waard nogmaals te kijken naar de onderzoeken waarin "Wenselijk - onwenselijk" voorkwam, deze ongewenste manipulatie eruit te halen, de onderzoeken te herhalen en de resultaten van de onderzoeken te vergelijken. Dit zou nuttig zijn om de vraag te beantwoorden in hoeverre er nog sprake is van een effect van het verschil in argumentkwaliteit, als in de manipulatie dit ongewenste aspect niet voorkomt.

Dit onderzoek kan enerzijds gebruikt worden als een aanwijzing naar hoe de gemiddelde argumentatiemanipulatie in onderzoek eruit ziet, en dat kan dus als richtlijn worden gehanteerd bij het opstellen van nieuw materiaal (hoewel het te onvolledig is om het te bestempelen als de richtlijnen waar O'Keefe en Jackson (1995) om vroegen).

Anderzijds is dit onderzoek een waarschuwing voor de huidige gang van zaken, namelijk dat wijdverspreid manipulaties voorkomen die het liefst vermeden dienen te worden, omdat deze ongewenste manipulaties niet leiden tot een verhouding 'sterk-zwak' maar tot 'voor-tegen'.

Referentielijst

Areni, C. S., & Lutz, R. J. (1988). The role of argument quality in the elaboration likelihood model. *Advances in consumer research*, 15(1), pp. 197-203.

Boller, G. W., Swasy, J. L., & Munch, J. M. (1990). Conceptualizing argument quality via argument structure. *Advances in consumer research*, 17, pp. 321-328.

Eemeren, F. H. van, Garssen, B., & Meuffels, B. (2009). *Fallacies and judgments of reasonableness: Empirical research concerning the pragma-dialectical discussion rules*. Dordrecht: Springer.

Eemeren, F.H. van, & Snoeck Henkemans, A.F. (2006). *Argumentatie: Inleiding in het analyseren, beoordelen en houden van betogen*. Groningen: Wolters Noordhoff.

Enschot-Van Dijk, R. van, Hustinx, L., & Hoeken, H. (2003). The concept of argument quality in the elaboration likelihood model. In *Anyone Who Has a View* (pp. 319-335). Springer Netherlands.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, Massachusetts.

Fishbein, M. & Yzer, M. C. (2003). Using theory to design effective health behaviour interventions. *Communication theory*, 14 (2), pp. 164-183.

Hoeken, H. (1995) *The design of persuasive texts: Effects of content, structure and style*. Tilburg: Proefschrift K.U. Brabant.

Hoeken, H., & Hustinx, L. (2002). De relatieve overtuigingskracht van anekdotische, statistische, causale en autoriteitsevidentie. *Tijdschrift voor Taalbeheersing*, 24(3), pp. 226-236.

Hoeken, H., Hornikx, J., & Hustinx, L. (2012). *Overtuigende teksten. Onderzoek en ontwerp*. Bussum: Coutinho.

Hoeken, H., Timmers, R. & Schellens, P. J. (2012) Arguing about desirable consequences: What constitutes a convincing argument? *Thinking and reasoning*, 18 (3), 394-416.

Hornikx, J. (2005). A review of experimental research on the relative persuasiveness of anecdotal, statistical, causal, and expert evidence. *Studies in Communication Sciences* (5) pp. 205-216.

Hustinx, L. G. M. M., Enschoot, R. V. & Hoeken, H. (2006). Argumentkwaliteit en overtuigingskracht in het Elaboration Likelihood Model: Welke dimensies spelen een rol. *Tijdschrift voor taalbeheersing*, 28(1), 39-53.

Kahneman, D. (2011). *Thinking, fast and slow*. Penguin books: London.

McGuire, W. J. (1960). Direct and indirect persuasive effects of dissonance-producing messages. *The Journal of Abnormal and Social Psychology*, 60(3), p. 354.

O'Keefe, D. J., & Jackson, S. (1995). Argument quality and persuasive effects: A review of current approaches. In S. Jackson (Ed.), *Argumentation and values: Proceedings of the ninth Alta conference on argumentation* (pp. 88-92). Annandale, VA: Speech Communication Association.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in experimental social psychology*, 19, pp. 123-205.

Petty, R. E., Cacioppo, J. T., Strathman, A. J., & Priester, J. R. (2005). To think or not to think. *Persuasion: Psychological insights and perspectives*, pp. 81-116.

Schellens, P. J. & Verhoeven, G. (1994) *Argument en tegenargument. Een inleiding in de analyse en beoordeling van betogende teksten*. Groningen: Martinus Nijhoff.

Schellens, P. J., & Jong, M. D. (2000). Soorten argumenten in de voorlichting. *Tijdschrift voor Taalbeheersing*, 22(4), pp. 288-308.

Schellens, P. J., Šorm, E., Timmers, R., & Hoeken, H. (2014). Lekencriteria voor de evaluatie van vijf soorten argumentatie. *Tijdschrift voor Taalbeheersing*, 36(1), pp. 51-86.

Toulmin, S. E. (1958), *The Uses of Argument*. London: Cambridge University Press

Witte, K. (1998). Fear as motivator, fear as inhibitor: Using EPPM to explain fear appeal successes and failures. In P. A. Anderson, & L.K. Guerrero (eds.), *The handbook of communication and emotion* (pp.423-450). New York: Academic Press.

Appendix A. Codeboek

Auteurs	Auteurs
Jaartal	Jaar van publicatie
Experiment nr	Als er meerdere experimenten zijn, welk experiment of welke experimenten deze argumenten gebruiken.
Tijdschrift	Tijdschrift waarin is gepubliceerd
Pretest/variaties in argumentkwaliteit	hoe wordt onderscheid gemaakt in argumentkwaliteit, obv de indeling van O'Keefe & Jackson (p. 88-91)? 1) pretest voor overtuigingskracht dmv thoughtlisting, 2) pretest oordelen over argumentkwaliteit (meestal gemeten op schalen) 3) zelfgemaakte variaties in de argumenten
Product	welk product (of dienst) wordt aangeprezen?
Aantal sterke argumenten	hoeveel argumenten zitten er in de sterke set
Aantal zwakke argumenten	hoeveel argumenten zitten er in de zwakke set
Domein: zelfde	komt het in de set voor dat het domein of de dimensie van de genoemde gevolgen in de sterke en zwakke set argumenten hetzelfde is? 0 = nee, 1 = ja
Domein: verschillend	komt het in de set voor dat het domein of de dimensie van de genoemde gevolgen in de sterke en zwakke set argumenten verschillend is? 0 = nee, 1 = ja
Waarschijnlijkheid gemanipuleerd wel/niet	is waarschijnlijkheid gemanipuleerd? 0 = nee, 1 = ja
WA: ja, hoe?	als waarschijnlijkheid gemanipuleerd is, hoe is dat gedaan? Beschrijf in woorden (e.g., door een betrouwbare of niet betrouwbare bron te vermelden)
Wenselijk-minder wenselijk	komt het voor dat gevolgen/aspecten genoemd in zwakke argumenten minder wenselijk (maar nog wel wenselijk) zijn dan wenselijke gevolgen in de sterke argumenten?
Wenselijk - neutraal	komt het voor dat gevolgen/aspecten genoemd in zwakke argumenten neutraal zijn t.o.v. van wel wenselijke gevolgen in de sterke argumenten?
Wenselijk-onwenselijk	komt het voor dat gevolgen/aspecten genoemd in zwakke argumenten onwenselijk zijn t.o.v. wel wenselijke gevolgen in de sterke argumenten?

Appendix B. Frequentietabellen

Tabel 9. Frequenties domein (N = 44)

Domein	Aantal voorkomend	Percentage
Alleen hetzelfde	16	36,4
Alleen verschillend	14	31,8
Beide	13	29,5
Geen van beide	1	2,3

Tabel 10. Frequenties pretest (N = 44)

Pretest	Aantal voorkomend	Percentage
Thoughtlisting	2	4,5
Rating op Likert-schalen	27	61,4
Zelfgemaakte variaties	12	27,3
Anders	3	6,8

Tabel 11. Frequenties argumentatiemaker (N = 44)

Argumentatiemaker	Aantal voorkomend	Percentage
Onderzoekers	32	72,7
Participanten in pretest	8	18,2
Anders	4	9,1

Tabel 12. Frequentie soort waarschijnlijkheidsmanipulatie (N = 44)

Soort	Aantal	Percentage
Geen	23	52,3
Anekdotisch	1	2,3
Statistisch	7	15,9
Causaal	0	0
Autoriteit	8	18,2
Ontbreken evidentie	5	11,4

Tabel 13. Modi van categorieën

Categorieën	Modus
Argumentatiemaker	Onderzoekers
Pretest	Rating op Likert-schalen
Aantal sterke argumenten	3
Aantal zwakke argumenten	3
Domein zelfde	Aanwezig
Domein verschillend	Aanwezig
Waarschijnlijkheidsmanipulatie	Niet aanwezig
Wenselijk - minder wenselijk	Aanwezig
Wenselijk - neutraal	Niet aanwezig
Wenselijk - onwenselijk	Niet aanwezig

Appendix C. Rol pretest en argumentatiemaker

Tabel 14. Rol van pretest "Likertschalen" (N = 27) vergeleken met de rest

<i>Variabele</i>	<i>F</i>	<i>Kans</i>
Aantal sterke argumenten	F (1, 42) = .016	p = .899
Aantal zwakke argumenten	F (1, 42) = .144	p = .707
Domein zelfde	F (1, 42) = .071	p = .792
Domein verschillend	F (1, 42) = 2.414	p = .128
Domein	F (1, 42) = 3.451	p = .070
Waarschijnlijkheidsmanipulatie	F (1, 42) = 3.295	p = .077
Wenselijk - minder wenselijk	F (1, 42) = 1.063	p = .308
Wenselijk - neutraal	F (1, 42) = .465	p = .499
Wenselijk - onwenselijk	F (1, 42) = .148	p = .703
Aantal woorden sterk	F (1, 42) = .000	p = .994
Aantal woorden zwak	F (1, 42) = .002	p = .967

Tabel 15. Rol van pretest "Zelfgemaakt" (N = 12) vergeleken met de rest

<i>Variabele</i>	<i>F</i>	<i>Kans</i>
Aantal sterke argumenten	F (1, 42) = .019	p = .890
Aantal zwakke argumenten	F (1, 42) = .001	p = .981
Domein zelfde	F (1, 42) = .716	p = .402
Domein verschillend	F (1, 42) = .061	p = .806
Domein	F (1, 42) = .401	p = .530
Waarschijnlijkheidsmanipulatie	F (1, 42) = 2.394	p = .129
Wenselijk - minder wenselijk	F (1, 42) = .144	p = .706
Wenselijk - neutraal	F (1, 42) = .109	p = .743
Wenselijk - onwenselijk	F (1, 42) = .716	p = .402
Aantal woorden sterk	F (1, 42) < .001	p = .996
Aantal woorden zwak	F (1, 42) = .153	p = .697

Tabel 16. Rol van argumentatiemakers "Onderzoekers" (N = 32) vergeleken met de rest

<i>Variabele</i>	F	<i>Kans</i>
Aantal sterke argumenten	F (1, 42) < .001	p = 1.000
Aantal zwakke argumenten	F (1, 42) = .001	p = .981
Domein zelfde	F (1, 42) = .017	p = .898
Domein verschillend	F (1, 42) = .061	p = .806
Domein	F (1, 42) = .060	p = .807
Waarschijnlijkheidsmanipulatie	F (1, 42) = .233	p = .632
Wenselijk - minder wenselijk	F (1, 42) = .444	p = .509
Wenselijk - neutraal	F (1, 42) = .157	p = .694
Wenselijk - onwenselijk	F (1, 42) = .716	p = .402
Aantal woorden sterk	F (1, 42) = .002	p = .969
Aantal woorden zwak	F (1, 42) = .002	p = .969

Tabel 17. Rol van argumentatiemakers "Participanten" (N = 8) vergeleken met de rest

<i>Variabele</i>	F	<i>Kans</i>
Aantal sterke argumenten	F (1, 42)	p = .196
Aantal zwakke argumenten	F (1, 42)	p = .184
Domein zelfde	F (1, 42)	p = .711
Domein verschillend	F (1, 42)	p = .943
Domein	F (1, 42)	p = .968
Waarschijnlijkheidsmanipulatie	F (1, 42)	p = .533
Wenselijk - minder wenselijk	F (1, 42)	p = .187
Wenselijk - neutraal	F (1, 42)	p = .253
Wenselijk - onwenselijk	F (1, 42)	p = .232
Aantal woorden sterk	F (1, 42)	p = .699
Aantal woorden zwak	F (1, 42)	p = .689