

Multiculturele website-evaluatie

Verschillen tussen individualistische en collectivistische proefpersonen

1. Inleiding

Ontwerpers van websites gericht op internationale doelgroepen moeten rekening houden met cultuurverschillen, dat spreekt vanzelf. Er is dan ook een groeiende interesse in de vakliteratuur voor kwesties die met internationale en interculturele aspecten van websites te maken hebben. Zo is er onderzoek gedaan naar de relatie tussen cultuurkenmerken en internetgebruik (bijvoorbeeld La Ferle, Edwards & Mizuno 2002), de mate waarin culturele achtergronden van invloed zijn op de waardering van websites (bijvoorbeeld O'Keefe e.a. 2000; Simon 2001) en de vraag hoe internationale organisaties op hun websites omgaan met culturele kwesties (bijvoorbeeld Marcus & Gould 2000; Becker 2002; Okazaki & Rivas 2002). Arnold (1998) geeft een overzicht van linguïstische, culturele, juridische en technische complicaties die zich voordoen bij het ontwerpen van websites voor internationale doelgroepen.

Een voor de hand liggende aanpak om een website geschikt te maken voor bezoekers met verschillende culturele achtergronden is die verscheidenheid ook tot zijn recht te laten komen bij het evalueren of testen van de site (Hoft 1995; Nielsen 2000). Voor zo'n evaluatie zijn verschillende methoden voorhanden (zie Schriver 1989; De Jong & Schellens 1995, 1997), en er zijn goede handleidingen voor usability testing beschikbaar (zoals Dumas & Redish 1993; Rubin 1994; Barnum 2002; Schweibenz & Thissen 2003). De Jong & Schellens (2000, 2002) concluderen uit een overzicht van de beschikbare methodologische

Samenvatting

Een website is geëvalueerd met proefpersonen afkomstig uit twee culturen: een collectivistische, contextgevoelige cultuur en een individualistische, weinig contextgevoelige cultuur. Voor de test werd gebruik gemaakt van retrospectieve hardopdenkprotocollen en de plus-minmethode. Uit het onderzoek blijkt dat de plus-minmethode aanzienlijk minder problemen aan het licht brengt bij proefpersonen uit een collectivistische cultuur. Retrospectieve hardopdenkprotocollen blijken minder gevoelig voor culturele invloeden, maar er zijn wel verschillen: proefpersonen uit een collectivistisch cultuur vallen minder vaak uit de rol van gebruiker die ze in de test moeten aannemen en hun commentaar is vaak indirecter geformuleerd dan dat van de proefpersonen uit een individualistische cultuur.

literatuur dat een evaluatie met beoogde gebruikers een effectieve manier is om de bruikbaarheid van documenten en interfaces te beoordelen en te verbeteren. Zowel *in-use* evaluatiemethoden, zoals hardopdenk usability testing, als *non-use* methoden, zoals de plus-en-min-methode, blijken geschikt. Er is echter nooit nagegaan wat de invloed is van de culturele achtergrond van proefpersonen op het verloop en de resultaten van evaluatieonderzoek. Werken hardopdenkprotocollen hetzelfde bij proefpersonen uit Europa en Azië, en zijn de resultaten ook vergelijkbaar? En hoe gedragen beide groepen proefpersonen zich in een plus-en-minonderzoek? Tot dusver is het onderzoek naar de validiteit en de bruikbaarheid van evaluatiemethoden vrijwel uitsluitend gedaan met proefpersonen uit Noord-Amerika en West-Europa. Het is niet duidelijk of de conclusies van dat onderzoek ook opgaan voor proefpersonen uit andere culturen.

Er is wel enig onderzoek beschikbaar naar de invloed van andere persoonsvariabelen op de feedback die verkregen wordt in evaluatieonderzoek. Uit onderzoek van De Jong & Schellens (2001) bleek dat mannelijke en vrouwelijke proefpersonen verschillende commentaren gaven op brochures; ook bleek dat hoger opgeleiden meer problemen in brochures signaleerden dan lager opgeleiden, en meer aandacht hadden voor problemen met de tekststructuur. Van Versseveld (1995) toonde aan dat de betrokkenheid van proefpersonen bij het onderwerp invloed kan hebben op het commentaar dat wordt verkregen op een brochure. In een plus-en-min-test noemden hoog-betrokkenen meer problemen in een brochure, en richtten ze zich vooral meer op volledigheidproblemen (behoefte aan meer informatie), terwijl de laag-betrokkenen juist meer relevantieproblemen noemden. In een experiment met het testen van vragenlijsten vonden Diamantopoulos, Reynolds & Schlegelmilch (1994) dat deelnemers met inhoudelijke voorkennis en kennis van vragenlijsten beter in staat waren om allerlei gebreken in de vraagformulering te ontdekken (zoals ambiguë vragen of ontbrekende antwoordmogelijkheden). Blijkbaar maakt het uit wat voor proefpersonen aan een doelgroepgerichte pretest deelnemen.

Het internationale karakter van het World Wide Web roept de vraag op naar het effect van nationale cultuur als relevant achtergrondkenmerk van proefpersonen in een pretest. Het is immers genoegzaam bekend dat deze variabele invloed heeft op allerlei andere vormen van gedrag (Hall 1977; Hofstede 1980, 1994; Smith & Bond 1998; Trompenaars & Hampden-Turner 1998). De hoofdvraag van ons onderzoek is dan ook: *Beïnvloeden cultuurverschillen tussen proefpersonen de aard van de feedback die verkregen wordt bij de evaluatie van een website?* Om deze vraag te beantwoorden, voerden we een webevaluatie uit met proefpersonen uit West-Europa en uit Azië en Afrika en vergeleken we zowel de resultaten als de ervaringen van de proefpersonen.

Uit het spectrum van mogelijke evaluatiemethoden kozen we er twee, namelijk retrospectieve hardopdenkprotocollen en de plus-en-minmethode. De eerste staat model voor een *in-use* benadering. De proefpersonen voeren taken uit met behulp van een website en hun handelingen worden opgenomen op video. Na afloop bekijken ze de video en proberen ze onder woorden te brengen wat ze gedacht hebben tijdens de uitvoering van de taken (cf. Nielsen 1993). We kozen voor een retrospectieve in plaats van een synchrone hardopdenksessie omdat het onderzoek werd gehouden in het Engels, wat niet de moedertaal was van de proefpersonen. We veronderstellen dat de taak om hardop te denken in een vreemde taal een te zware cognitieve belasting voor de proefpersonen is, waaronder zowel de taakuitvoering als het hardopdenken zou kunnen lijden. Uit eerder onderzoek blijkt dat synchroon en retrospectief hardopdenken goed vergelijkbare resultaten opleveren (Hoc &

Multiculturele website-evaluatie

Leplat 1983; Van den Haak, De Jong & Schellens 2003).

De plus-en-minmethode is een typische *non-use* methode (De Jong 1998). De proefpersonen wordt gevraagd om een document te lezen en plussen en minnen in de marge te zetten op plaatsen waar ze positieve of negatieve leeservaringen hebben. In het tweede deel van een plus-en-minsessie worden de proefpersonen geïnterviewd over de redenen waarom ze plussen en minnen noteerden. Hoewel er enkele pogingen zijn gedaan om deze methode toe te passen voor de evaluatie van websites, hebben we er in dit onderzoek voor gekozen te werken met geprinte versies van enkele webpagina's.

Beide methoden zijn gangbaar in de communicatiepraktijk. Met het oog op mogelijke cultuurverschillen is de combinatie van methoden interessant omdat ze beide een hoge mate van interactie met de proefpersoon met zich meebrengen, maar verschillende eisen stellen aan de proefpersonen. Bij de hardopdenkmethode moeten de proefpersonen zich gedragen als echte gebruikers van de website en inzicht geven in de fouten die ze maken en de twijfels die ze hebben bij het uitvoeren van hun taken. Bij de plus-minmethode vervullen de proefpersonen de rol van beoordelaar en moeten ze hun oordelen over de website geven aan de proefleider. Beide methoden kunnen in bepaalde opzichten bedreigend zijn voor een proefpersoon.

2. Dimensies van cultuurverschillen

Cultuurverschillen worden vaak gekarakteriseerd aan de hand van zogenaamde cultuurdimensies. Dat zijn aspecten van een cultuur die gemeten kunnen worden in relatie tot andere culturen. Een toonaangevende reeks dimensies werd ontwikkeld door Hofstede (1994, 2001). Op basis van een vragenlijst die werd ingevuld door 116.000 personeelsleden van IBM in 50 landen en 20 talen onderscheidde hij de volgende dimensies:

- masculiniteit versus femininiteit,
- hoge versus lage onzekerheidsvermijding,
- grote versus kleine machtsafstand,
- individualisme versus collectivisme,
- lange- versus korte-termijnoriëntatie.

Vergelijkend onderzoek heeft aangetoond dat de dimensie *individualisme-collectivisme* de belangrijkste is als het gaat om verschillen tussen culturen (Ting-Toomey 1998). In individualistische culturen zijn de banden tussen mensen relatief los: mensen worden geacht vooral voor zichzelf en de meest directe familieleden te zorgen. In collectivistische culturen leven mensen van geboorte tot dood in hechte groepen, die een levenslange beschermde omgeving vormen en daarvoor in ruil een groot beroep doen op loyaliteit. De Verenigde Staten en Zweden zijn sterk individualistische landen; de Arabische landen en Indonesië scoren juist laag op de individualisme-index (IDV).

Verskillende auteurs stellen dat deze dimensie sterk is gerelateerd aan de manier waarop mensen met elkaar communiceren. Dit sluit aan op een ander onderscheid van Hall (1977), tussen contextgevoelige (*high-context*) en weinig contextgevoelige (*low-context*) culturen (Hofstede 2001; Ting-Toomey 1998). In weinig contextgevoelige culturen dient communicatie expliciet, direct en eenduidig te zijn. In contextgevoelige culturen ligt veel informatie besloten in de context, of is deze geïnternaliseerd in de personen die met elkaar

communiceren; er wordt weinig expliciet gemaakt in communicatieve boodschappen. Duitsland, Scandinavië, de Verenigde Staten en Zwitserland bezitten een relatief weinig contextgevoelige cultuur, terwijl Aziatische en Mediterrane landen juist gekenmerkt worden door een contextgevoelige cultuur. Contextgevoelige culturen corresponderen met collectivistische culturen, terwijl weinig contextgevoelige culturen doorgaans individualistisch zijn. In individualistische culturen moeten veel vanzelfsprekende zaken expliciet gezegd worden (Hofstede, 2001 p. 212).

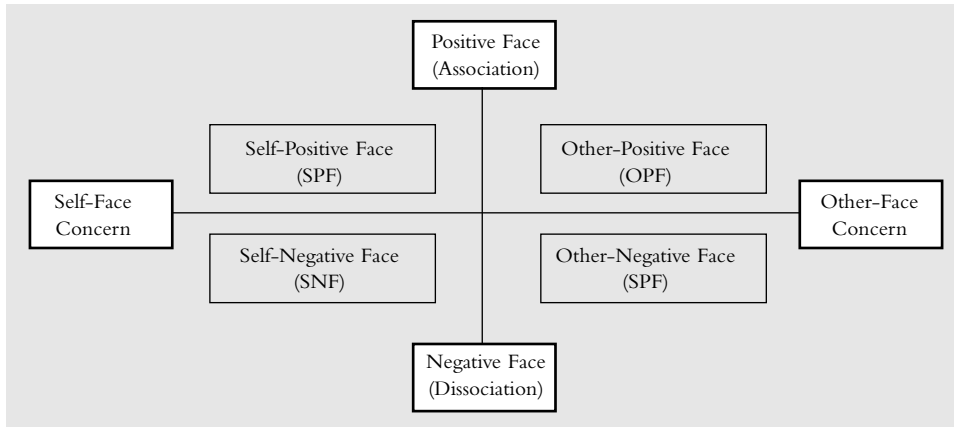
Het verschil tussen de West-Europese en de Aziatisch en Afrikaanse proefpersonen in ons onderzoek kan goed gekarakteriseerd worden door een combinatie van Hofstedes dimensie individualisme-collectivisme en Halls onderscheid naar contextgevoeligheid. Aziatische en Afrikaanse proefpersonen bevinden zich dan aan de collectivistische en contextgevoelige zijde van het spectrum, West-Europese proefpersonen aan de individualistische en weinig contextgevoelige zijde. Vanzelfsprekend gaat het hierbij om verschillen tussen groepen, en doen we geen voorspellingen over het gedrag van individuen. Bovendien moeten we in het oog houden dat culturen meestal verschillen op meer dimensies tegelijk. Onze karakterisering van de twee groepen is dus een (bedoelde) simplificatie van de werkelijkheid, die, zoals we hieronder zullen betogen, een vruchtbare basis is om hypothesen te ontwikkelen voor ons onderzoek.

3. *Cultuur en beleefdheid: hypothesen voor het onderzoek*

Ting-Toomey (1998) verbond theorieën over cultuurverschillen met de beleefdheidstheorie van Brown & Levinson (1990). Centraal in deze beleefdheidstheorie staat de behoefte van mensen om hun 'gezicht te behouden'. Mensen willen door anderen gewaardeerd worden (positive face) en ze willen niet door anderen gedwongen worden tot gedrag dat ze niet wensen (negative face). Brown & Levinson hebben een typologie van gezichtsbedreigende handelingen ontwikkeld en onderzoek gedaan naar de wijze waarop die voorkomen. Voorbeelden van positief-gezichtsbedreigende handelingen zijn afkeuren en kritiek geven: de spreker (S) laat merken dat hij of zij de verlangens, eigendommen of persoonlijke eigenschappen van de hoorder (H) niet op prijs stelt. Negatief-gezichtsbedreigende handelingen zijn bijvoorbeeld opdrachten: S geeft te kennen dat hij wil dat H iets doet. Maar ook bijvoorbeeld een aanbod kan een negatief-gezichtsbedreigende handeling zijn: S verplicht H om het aanbod te accepteren of af te slaan, en wellicht om in de toekomst een wederdienst te bewijzen. Hoewel er culturele verschillen zijn in de manier waarop inhoud wordt gegeven aan het begrip 'face', is de erkenning van het belang ervan voor mensen universeel.

Aan de hand van een onderscheid tussen 'self-face concern' en 'other-face concern' en van het hierboven genoemde onderscheid tussen positive en negative face ontwikkelde Ting-Toomey (1998) een model met vier kwadranten (zie figuur 1). Mensen die gericht zijn op 'self-positive face' (SPF), zoeken in hun communicatie aansluiting bij en waardering van anderen. Mensen die gericht zijn op 'other-positive face' (OPF), houden in hun communicatie vooral rekening met de behoefte aan aansluiting en waardering van anderen. Mensen die gericht zijn op 'self-negative face' (SNF) proberen in hun communicatie vooral hun eigen vrijheid van handelen en autonomie te beschermen. Mensen die gericht zijn op 'other-negative face' (ONF) tonen in hun communicatie veel respect voor de handlingsvrijheid van anderen.

Multiculturele website-evaluatie



Figuur 1. Vier kwadranten van facework (Ting-Toomey 1998, p.218)

Volgens Ting-Toomey verschillen individualistische, weinig contextgevoelige culturen in veel opzichten van collectivistische, contextgevoelige culturen. Voortbouwend op figuur 1 ontwikkelde ze een set theoretische aannames over verschillen in ‘facework’ tussen de twee culturen (zie tabel 1). Deze aannames vormden de basis voor de hypothesen die we voor ons onderzoek hebben geformuleerd. Hieronder zullen we de zeven hypothesen beschrijven en toelichten.

Tabel 1. Facework in individualistische, weinig contextgevoelige en in collectivistische, contextgevoelige culturen (gebaseerd op Ting-Toomey, 1998, p.230)

Elementen van ‘face’	Individualistische culturen	Collectivistische culturen
Identiteit	Nadruk op ‘ik’	Nadruk op ‘wij
Primaire zorg voor	Self-face	Other-face
Behoeft aan	Negative face	Positive face
Strategieën gericht op Stijl	Self-positieve en self-negatieve face	Other-positieve en other-negatieve face
Taalhandelingen	Directe taalhandelingen	Indirecte taalhandelingen
Nonverbaal gedrag	Directe emotionele uitdrukkingen	Indirecte emotionele uitdrukkingen

De eerste vraag die we ons stellen, is in hoeverre culturele achtergrond van invloed is op de resultaten van de pretest. Bij de plus-en-minmethode is deze invloed te verwachten. Proefpersonen die in een plus-en-minmethode problemen naar voren willen brengen, moeten openlijk en direct kritiek op de website geven. Dit past niet bij de neiging tot indirecte communicatie en bij de ‘other-face concern’ die in collectivistische culturen dominant zijn. Bij de hardopdenkprotocollen is voorhands geen verschil te verwachten. Veel van de problemen die in hardopdenkonderzoek aan het licht komen, zijn direct gerelateerd aan knelpunten die zich voordoen tijdens de taakuitvoering. Er is geen reden om te veronderstellen dat één van beide groepen proefpersonen meer knelpunten zal ervaren dan de andere. We hebben voor ons onderzoek namelijk twee groepen proefpersonen geworven met een gelijkwaardig opleidingsniveau (zie paragraaf 4.2).

- H1 De plus-en-minmethode zal minder problemen aan het licht brengen bij proefpersonen uit collectivistische culturen dan bij proefpersonen uit individualistische culturen.
- H2 De retrospectieve hardopdenkprotocollen zullen evenveel problemen aan het licht brengen bij proefpersonen uit collectivistische culturen als bij proefpersonen uit individualistische culturen.

De culturele oriëntatie kan daarnaast een verschil in de appreciatie van de testmethode met zich meebrengen. Beide onderzochte methoden hebben consequenties voor de 'face' van de proefpersoon. We veronderstellen daarbij dat er een relatie is tussen het type 'facework' dat dominant is in beide culturen en de ervaringen van de proefpersonen tijdens het evaluatieonderzoek. In collectivistische culturen is er met name sprake van 'other-face concern' en een behoefte aan 'positive face'. Beide worden bedreigd door een opzet met de plus-en-minmethode, waarbij proefpersonen immers kritiek moeten uiten op een website. Individualistische culturen zijn met name gericht op een 'self-face concern' en op 'negative face'. De 'self-face concern' wordt bedreigd door een test met retrospectieve hardopdenkprotocollen, omdat daarin duidelijk wordt wat de proefpersonen allemaal fout doen tijdens de taakuitvoering (hoewel ze natuurlijk kunnen besluiten om de schuld van deze problemen te zoeken bij de website en niet bij zichzelf; zie hypothese 5). De behoefte aan 'negative face' wordt bedreigd doordat de proefpersonen in de onderzoekssetting worden gedwongen om de website op een bepaalde manier, aan de hand van een set taken, te gebruiken (hoewel ze natuurlijk kunnen besluiten om hun rol als proefpersoon ruimer op te vatten; zie hypothese 6).

- H3 De proefpersonen uit collectivistische culturen zullen positiever oordelen over de retrospectieve hardopdenkprotocollen dan de proefpersonen uit individualistische culturen.
- H4 De proefpersonen uit individualistische culturen zullen positiever oordelen over de plus-en-minmethode dan de proefpersonen uit collectivistische culturen.

Een derde aspect dat vanuit cultureel oogpunt interessant is, betreft het verschijnsel 'blaming' (Schriver, 1997): de mate waarin proefpersonen zichzelf, het product of de testsituatie de schuld geven wanneer zij tegen problemen aanlopen tijdens een usability test. De vraag of proefpersonen in ons onderzoek zichzelf (intern) dan wel de website of de testsituatie (extern) de schuld geven van problemen kan cultuurafhankelijk zijn. Vanwege hun 'other-face concern' kunnen proefpersonen uit collectivistische culturen meer geneigd zijn om zelf de schuld op zich te nemen; vanwege hun 'self-face concern' zouden proefpersonen uit individualistische culturen juist meer geneigd kunnen zijn om de schuld bij externe factoren te zoeken.

- H5 Proefpersonen uit collectivistische culturen zullen in het hardopdenkonderzoek meer geneigd zijn om de schuld voor problemen tijdens de taakuitvoering bij zichzelf te zoeken dan proefpersonen uit individualistische culturen.

Multiculturele website-evaluatie

Een vierde mogelijke invloed van cultuurverschillen betreft de rol van proefpersonen in de test. In hardopdenkonderzoek moeten proefpersonen zich gedragen als gebruikers die een beperkt aantal, niet door henzelf gekozen, taken uitvoeren en met name aandacht besteden aan direct aan die taken gerelateerde problemen. Deze rol kan in strijd zijn met de behoefte aan 'negative face' die bij individualistisch proefpersonen sterker aanwezig is dan bij collectivistische proefpersonen. Als reactie op deze beperkingen zouden ze gedurende het proces andere rollen kunnen aannemen dan de strikte gebruikersrol die in het onderzoek opgesloten ligt. Deze rollen kunnen in hun hardopdenkcommentaar naar voren komen. We onderscheiden de volgende rollen: de rol van proefpersoon (ingaan op de eigen ervaringen tijdens de test en de eigen prestaties), de rol van internetgebruiker (vertellen over het eigen gebruik van het World Wide Web in normale omstandigheden) en de rol van reviewer (oordelen geven over de website).

- H6 In het retrospectieve hardopdenkonderzoek zullen proefpersonen uit individualistische culturen minder geneigd zijn om zich te houden aan de door de taken geïmpliceerde gebruikersrol dan proefpersonen uit collectivistische culturen.

De laatste hypothese heeft betrekking op de manier waarop de proefpersonen zich in de test uitdrukken. Mensen uit collectivistische culturen hebben naar verwachting een voorkeur voor indirecte taalhandelingen bij het leveren van commentaar, omdat ze meer gericht zijn op 'other-face concern' en 'positive face'. We hebben dit overigens alleen onderzocht in de retrospectieve hardopdenkprotocollen. De plus-en-minresultaten bevatten zoveel commentaar en zoveel combinaties van directe en indirecte uitingen dat een betrouwbare scoring van de taalhandelingen niet goed mogelijk was.

- H7 Proefpersonen uit collectivistische culturen zullen meer geneigd zijn om indirecte en eufemistische formuleringen te kiezen voor de kritiek die ze hebben op de website dan proefpersonen uit individualistische culturen.

4. Methode

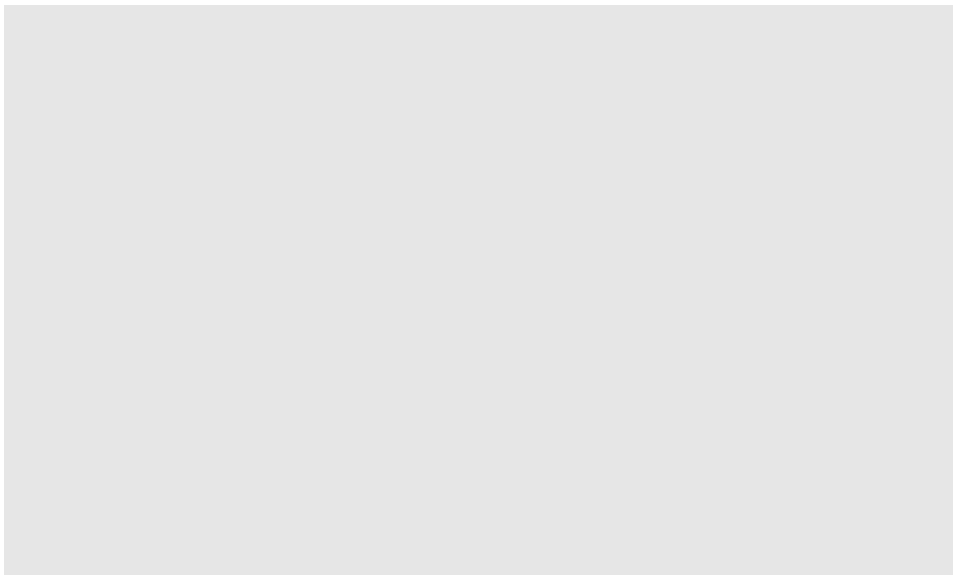
Om de bovenstaande hypothesen te toetsen, hebben we een website geëvalueerd met een combinatie van retrospectieve hardopdenkprotocollen en de plus-en-minmethode. In het onderzoek werden twee groepen proefpersonen betrokken: assistenten in opleiding (aio's) van Aziatische of Afrikaanse afkomst en West-Europese aio's. Hieronder bespreken we de website die we hebben gebruikt, de proefpersonen en de procedure. Vervolgens gaan we in op de afhankelijke variabelen in het onderzoek.

4.1 Onderzochte website: *Web of Science*. Voor ons onderzoek waren we op zoek naar een website die voldeed aan de volgende criteria:

- Een instructieve functie: voor het retrospectieve hardopdenkonderzoek was het wenselijk dat proefpersonen een aantal realistische taken aan de hand van de website konden uitvoeren.
- Substantiële tekstuele inhoud: voor de plus-en-minmethode was het wenselijk dat de website tekstuele informatie bevatte die de proefpersonen konden lezen en becommentariëren.

- Geen cultuurspecifieke inhoud: de website moest gericht zijn op gebruikersgroepen uit verschillende culturen. Het zou even waarschijnlijk moeten zijn dat Aziatische/Afrikaanse en West-Europese proefpersonen de website zouden bezoeken en gebruiken.

Op grond van deze criteria hebben we de *Web of Science* database gekozen als onderzoeksobject. *Web of Science* wordt gepubliceerd en bijgehouden door het Institute for Scientific Information (ISI). De kern van de database bestaat uit citatie-indexen. Wetenschappers kunnen er opzoeken hoe vaak en in welke artikelen er naar een bepaald artikel is verwezen. *Web of Science* omvat een groot aantal wetenschappelijke tijdschriften in de alfa-, bèta- en gammawetenschappen. Voor aio's geldt *Web of Science* als een belangrijke informatiebron: het is één van de manieren om systematisch wetenschappelijke literatuur te zoeken. Ter illustratie bevat figuur 2 het 'full search' scherm van de website.



Figuur 2. De 'full search' pagina van het *Web of Science*

4.2 Proefpersonen. We hebben voor het onderzoek twee homogene groepen proefpersonen geworven, die alleen verschilden op hun nationale herkomst. We vroegen mannelijke aio's in de technische wetenschappen aan de Universiteit Twente om deel te nemen. Deze steekproef bleek om meerdere redenen geschikt voor ons onderzoek:

- De proefpersonen behoorden allemaal tot de doelgroep van *Web of Science*.
- Het opleidingsniveau van de proefpersonen was in beide groepen gelijk. Alle proefpersonen hadden hun mastersdiploma gehaald en werkten aan hun promotieonderzoek. Een gelijkwaardig opleidingsniveau is belangrijk, omdat eerder onderzoek van De Jong & Schellens (2001) uitwees dat er een relatie is tussen opleiding en de aard van de feedback in een pretest.
- De proefpersonen uit Azië en Afrika waren op het moment van onderzoek slechts één tot twee jaar in Nederland, en kunnen daarom nog worden beschouwd als representanten van de cultuur van hun land van herkomst.

Multiculturele website-evaluatie

- Voor alle proefpersonen was Engels niet hun moedertaal. Vanwege de sterk uiteenlopende nationaliteiten van de proefpersonen hadden we besloten om het gehele onderzoek, in beide groepen in het Engels af te nemen. De taalvaardigheid van de proefpersonen kan immers invloed hebben op hun verbalisaties in de retrospectieve hardopdenkprotocollen en op het soort en de hoeveelheid feedback die ze geven bij de plus-en-minmethode (de website zelf was ook in het Engels). Gezien de vooropleiding van de deelnemers en de Engelstalige praktijk aan de Universiteit Twente, mogen we aannemen dat de beheersing van het Engels in beide groepen min of meer gelijk was.

Deze selectiecriteria maken het goed mogelijk om eventuele culturele verschillen op het spoor te komen, maar er is natuurlijk ook een keerzijde, in termen van generaliseerbaarheid: het onderzoek beperkt zich tot hoger opgeleide en mannelijke proefpersonen. Toekomstig onderzoek zou zich mede moeten richten op proefpersonen met lagere opleidingsniveaus en op vrouwelijke proefpersonen.

Proefpersonen die voldeden aan de eerdergenoemde criteria werd mondeling gevraagd of ze bereid waren om mee te werken aan ons onderzoek. In totaal wilden 38 aio's meedoen: 20 proefpersonen uit een individualistische cultuur (allen uit Nederland) en 18 uit collectivistische culturen (uit India, Indonesië, China, Turkije en Soedan).

4.3 Procedure. Het onderzoek vond plaats in individuele sessies in een onderzoeksruijme aan de Universiteit Twente. Het onderzoek bestond uit vier delen. Eerst werd de proefpersonen gevraagd om zeven taken met *Web of Science* uit te voeren. Hun taakuitvoering werd opgenomen met behulp van het programma HyperCam (<http://www.hyperionics.com>). Omdat het soort taken mogelijk van invloed is op de resultaten van een usability test (zie Sienot 1997; Van Waes 2000), kregen de proefpersonen in totaal zeven taken, verdeeld over twee categorieën: zoektaken en toepassingstaken. Om de eventuele invloed van voorkennis op de taakuitvoering uit te sluiten, besloten we om geen taken op te nemen die aansloten op de technisch-wetenschappelijke literatuur, maar in plaats daarvan dicht bij ons eigen vakgebied te blijven. Hieronder staat een overzicht van alle (in het Nederlands vertaalde) taken.

1. Hoeveel artikelen zijn er beschikbaar over het onderwerp 'communication theory'? Zorg ervoor dat u artikelen vindt waarin beide woorden opeenvolgend worden gebruikt.
2. Zoek op hoeveel artikelen er beschikbaar zijn over 'web evaluation'. Deze keer hoeven de twee woorden niet opeenvolgend te zijn.
3. Bewaar het laatste zoekresultaat op de A-drive van deze computer.
4. Hoeveel artikelen geschreven door Jan H. Spyridakis zijn er beschikbaar in de database?
5. Staat er in de database een tijdschrift met de titel *IEEE Transactions on Knowledge and Data Engineering*?
6. Hoeveel tijdschriften staan er in de Science Citation Index Expanded?
7. Hoe vaak werd het artikel van P.J. Schellens in *Technical Communication* geciteerd in andere artikelen?

Na de taakuitvoering werd de proefpersonen gevraagd om de schermopnamen te bekijken en hardop te denken in het Engels over de manier waarop de taken zijn uitgevoerd. De proefpersonen mochten het afspelen even stoppen wanneer ze dat wilden. Hun verbalisaties werden op cassette opgenomen. Van Someren, Barnard & Sandberg (1994) stellen dat retrospectieve data niet altijd waarheidsgetrouw zijn, vooral als er de nodige tijd zit tussen de taakuitvoering en de uitleg achteraf. Om dergelijke vertekeningen tegen te gaan, moesten de proefpersonen meteen na de taakuitvoering doorgaan met het bekijken en commentariëren van de schermopnamen.

Vervolgens werd de proefpersonen gevraagd om enkele uitgeprinte helppagina's van *Web of Science* met behulp van de plus-en-minmethode te evalueren. In de onderzochte pagina's werd uitleg gegeven over het doel van *Web of Science*, over de tijdschriften die in de database zijn opgenomen, en over zoekmogelijkheden op de site. Eerst zetten de proefpersonen plussen en minnen in de kantlijn voor passages die ze positief dan wel negatief waardeerden. Benadrukt werd dat de proefpersonen zelf mochten uitmaken om welke redenen ze plussen en minnen plaatsten en dat ze de teksteenheden voor een plus of min ook zelf konden kiezen (variërend van individuele woorden tot een hele pagina). Nadat de proefpersonen klaar waren met het lezen en plussen en minnen zetten, werd in individuele interviews getracht om de redenen voor elke plus en min te achterhalen. Dit interview werd wederom op cassette opgenomen.

Tot slot vulden de proefpersonen een vragenlijst in, waarmee we extra gegevens verzamelden over (1) hun oordelen over de twee evaluatiemethoden, (2) de mate waarin zij zichzelf, de website of de testsituatie de schuld gaven van de problemen die ze in het retrospectieve hardopdenkonderzoek waren tegengekomen, en (3) hun plaats op het continuüm collectivistisme-individualisme. Voor dat laatste gebruikten we, bij gebrek aan beter, Hofstede's IDV-index, hoewel er, mede door Hofstede zelf, vraagtekens gezet zijn bij de validiteit van dit instrument als middel om individuele culturele oriëntaties te meten (Hofstede, 2001, p.497).

4.4 Afhankelijke variabelen in het onderzoek. Om de opbrengst van de beide evaluatiemethoden te onderzoeken (H1 en H2), werd het gemiddelde aantal geconstateerde hardopdenk- en plus-en-minproblemen per proefpersoon berekend. Bij de plus-en-minmethode werd iedere negatieve opmerking van de proefpersonen gecodeerd als probleem. Bij de retrospectieve hardopdenkprotocollen werden alle afwijkingen van het optimale handlingsverloop als probleem gecodeerd, evenals de opmerkingen die proefpersonen maakten om hun twijfel, verrassing, afkeuring en dergelijke kenbaar te maken.

De oordelen van de proefpersonen over de twee methoden (H3 en H4) werden onderzocht door drie sets vragen, alle op vijfpunts Likert schalen. Een eerste set van vijf vragen richtte zich op de ervaringen van de proefpersonen in het retrospectieve hardopdenkonderzoek (bijvoorbeeld: 'Ik voelde me ongemakkelijk bij het uitvoeren van de taken'). Een tweede set van vier vragen had betrekking op de ervaringen in het plus-en-minonderzoek (bijvoorbeeld: 'Ik vond het niet prettig om de website met deze methode te evalueren'). Een derde set van vier vragen betrof een vergelijking van de twee evaluatiemethoden (bijvoorbeeld: 'Door *Web of Science* op papier te evalueren kon ik betere aanwijzingen voor verbeteringen geven dan door de taken uit te voeren').

De mate waarin de proefpersonen zichzelf, de website of de testsituatie de schuld gaven van de problemen die ze waren tegengekomen (H5), werd onderzocht met een set van

Multiculturele website-evaluatie

negen vragen (weer op vijf-punts Likert schalen). De proefpersonen moesten aangeven in hoeverre ze mogelijke verklaringen voor hun gebruikersproblemen onderschreven. Drie vragen hadden betrekking op de eigen vaardigheden (bijvoorbeeld: 'doordat ik niet goed heb gelezen'), drie op de kwaliteit van *Web of Science* (bijvoorbeeld: 'door een gebrekkige gebruikersvriendelijkheid van de website') en drie op de testsituatie (bijvoorbeeld: 'doordat de onderzoeker over mijn schouder meekeek').

De vraag in hoeverre de proefpersonen zich in het retrospectieve hardopdenkonderzoek hielden aan de opgelegde gebruikersrol (H6) werd onderzocht door middel van een analyse van de hardopdenkprotocollen. Iedere afwijking van de typische gebruikersrol werd gemarkeerd en geduid in termen van de drie alternatieve rollen die we eerder onderscheidde (proefpersoon, internetgebruiker en reviewer). In de analyse keken we naar het totale aantal afwijkingen van de opgelegde gebruikersrol, maar ook naar de rollen afzonderlijk.

De directheid in de formuleringen van de commentaren (H7) werd onderzocht aan de hand van een lijst met letterlijke uitingen van proefpersonen uit de retrospectieve hardopdenkprotocollen, die vervolgens werd voorgelegd aan 12 studenten Toegepaste Communicatiewetenschap, met de opdracht om aan elke uiting een directheidsscore op een vijf-puntsschaal te geven. De twaalf studenten vormden samen een betrouwbaar beoordelingsinstrument om directe en indirecte uitingen van elkaar te onderscheiden (Cronbach's alfa = .82). De analyse beperkte zich tot de proefpersonen die in hun hardopdenkprotocollen commentaar hadden gegeven op de website. De directheidsscores per uiting werden vergeleken met de culturele achtergrond van de proefpersonen als onafhankelijke variabele.

5. Resultaten

De retrospectieve hardopdenkprotocollen en de plus-en-minmethode hebben beide veel relevante problemen in de *Web of Science* website aan het licht gebracht. Zo werd in het hardopdenkonderzoek gemiddeld 33% van de taken zonder succes uitgevoerd. Per taak varieerde dit percentage van 3% (voor de derde opdracht: het bewaren van zoekresultaten op diskette) tot 54% (voor de tweede opdracht: het uitvoeren van een zoekopdracht). Veel problemen hadden te maken met het invoeren van onderwerps- en auteursgegevens op het zoekscherm. Daarnaast bleken twee eigenaardigheden van de interface zeer contra-intuïtief voor vrijwel alle gebruikers: (1) na het gebruik van de "back"-toets van de browser moet de nieuwe webpagina vaak opnieuw geladen worden, en (2) wie na het invoeren van zoektermen de enter-toets op het toetsenbord indrukt, wist daarmee alle net ingevoerde zoekgegevens en belandt vervolgens op de homepage van *Web of Science*. Met name dat laatste zal regelmatige *Web of Science* bezoekers vermoedelijk bekend voorkomen.

Bij de verwerking van de data is één van de collectivistische proefpersonen uiteindelijk buiten de analyse gehouden, omdat hij zich bij beide methoden onttrok aan de rol die hem was toebedeeld. Ondanks de instructies weigerde hij de zeven taken uit het hardopdenkonderzoek uit te voeren en gaf hij bij de plus-en-minmethode geen enkel specifiek commentaar op de voorgelegde helppagina's. In plaats daarvan verkende hij de website en de helppagina en uitte hij zijn bewondering ervoor ("It's an exhaustive list and it's definitely very handy to the layman who does not know how and what kinds of words could be en-

tered for a particular search. Indeed very good”). Deze handelwijze kan eventueel worden gezien als een extreme variant op een collectivistisch houding.

Hieronder zullen we de resultaten van de overige 37 proefpersonen bespreken aan de hand van de zeven hypothesen die we hadden opgesteld. Voordat we daarmee beginnen, gaan we in op twee relevante achtergrondkenmerken van de proefpersonen: hun scores op de IDV-index en hun eerdere ervaringen met *Web of Science*.

5.1 Achtergrondkenmerken van de proefpersonen. Een eerste vraag met betrekking tot de achtergrondkenmerken van de proefpersonen is of er, naast hun land van herkomst, een onafhankelijke bevestiging kan worden gevonden voor de veronderstelde verschillen in culturele oriëntatie. Aan het einde van de onderzoekssessies kregen de proefpersonen daartoe de vier vragen uit Hofstede's IDV-index voorgelegd. De IDV-index van de twee groepen bleek niet significant te verschillen (t-toets, $t = .554$, $df = 35$, $p = .583$). De Nederlandse proefpersonen scoorden precies zoals verwacht, maar de Aziatische en Afrikaanse proefpersonen scoorden individualistischer dan op grond van hun nationaliteit werd verwacht.

Een mogelijke verklaring betreft het type proefpersonen dat aan ons onderzoek heeft deelgenomen. Door selectie- en/of assimilatieprocessen zouden onze proefpersonen uit Azië en Afrika minder collectivistische kenmerken kunnen hebben dan we op voorhand hadden aangenomen. De invloed van selectieprocessen is waarschijnlijk, omdat al deze proefpersonen de ingrijpende beslissing hadden genomen om voor een wetenschappelijke carrière tijdelijk te emigreren naar een ver en onbekend land. Dergelijk avontuurlijk gedrag sluit beter aan op een individualistische oriëntatie dan op een collectivistische. Assimilatieprocessen zijn ook mogelijk, omdat alle proefpersonen inmiddels enige tijd in Nederland woonden en zich wellicht hebben aangepast aan Nederlandse normen.

Wellicht belangrijker is dat de IDV-index geen bewezen valide maat is om cultuurverschillen te meten. Het gaat om vier items die vreemd genoeg geen inhoudelijke relatie hebben met typisch individualistische of collectivistische kenmerken, en dus hoogstens als predictor kunnen fungeren. Hofstede (2001, p.497) is zelf ook niet onverdeeld optimistisch over de betrouwbaarheid en de validiteit van de index, met name voor het gebruik ervan om individuele verschillen in kaart te brengen. En in het algemeen is er twijfel aan de mogelijkheden om met behulp van vragenlijsten culturele verschillen aan te tonen (zie Peng, Nisbett & Wong 1997). Op grond van deze overwegingen blijven we de proefpersonen beschouwen als representanten van hun nationale culturen, ondanks de uitslagen van de IDV-scores. Een aanvullend argument-achteraf zullen we nog geven in paragraaf 6.2.

Een tweede relevante vraag over de twee groepen proefpersonen betreft hun eerdere ervaringen met *Web of Science*: in hoeverre zijn de beide groepen proefpersonen vergelijkbaar op dit punt? De meerderheid van de proefpersonen had *Web of Science* al eens eerder gebruikt. Er bleek hierbij echter een bijna-significant verschil te zijn tussen de individualistische en de collectivistische proefpersonen: er waren relatief minder collectivistische proefpersonen met ervaring met *Web of Science* (58% tegenover 90%, Fisher's exact test, $p = .052$). De intensiteit waarmee *Web of Science* in de laatste drie maanden was gebruikt, bleek vrijwel gelijk bij de twee groepen proefpersonen (1,8 tegenover 1,9 keer, t-toets, $t = .103$, $df = 26$, $p = .918$). Omdat voorkennis van invloed zou kunnen zijn op de feedback die de proefpersonen geven (Diamantopoulos, Reynolds & Schlegelmilch, 1994), hebben we besloten om de eerdere ervaring van de proefpersonen als extra (dichotome) variabele mee te nemen in onze analyses van de ontdekte problemen.

Multiculturele website-evaluatie

5.2 Aantal ontdekte problemen. Bij de plus-en-minmethode verwachtten we dat de collectivistische proefpersonen minder problemen zouden noemen dan de individualistische proefpersonen (H1). Bij de retrospectieve hardopdenkprotocollen verwachtten we geen verschillen tussen de beide groepen (H2). Zoals te zien is in tabel 2, werden beide hypothesen bevestigd in dit onderzoek. De η^2 bij de plus-en-minresultaten duidt op een substantieel verschil tussen de twee groepen proefpersonen. De eerdere ervaring met *Web of Science*, die we als extra variabele in de analyses hadden meegenomen, bleek bij beide methoden geen effect te hebben op het aantal ontdekte problemen. Er was ook geen sprake van een interactie-effect.

Tabel 2. Gemiddeld aantal problemen per proefpersoon in *Web of Science*

	Individualistisch	Collectivistisch	Significantie
Lezersproblemen in het plus-en-minonderzoek	4.8	1.8	$F(1,33)=8.97, p<.01, \eta^2=.21$
Gebruikersproblemen in de de retrospectieve hardopdenkprotocollen	7.2	9.5	n.s.

Naast de specifieke problemen leverden de retrospectieve hardopdenkprotocollen ook een indicatie op van het overall succes van de proefpersonen. We gaven al eerder aan dat er relatief veel taken waren die zonder succes werden uitgevoerd. Bij een vergelijking van het aantal met succes afgeronde taken hebben we geen significante verschillen gevonden. Zowel de culturele achtergrond van de proefpersonen als hun eerdere ervaring met de website had geen invloed op het succes in de taakuitvoering.

5.3 Oordelen over de twee evaluatiemethoden. We verwachtten dat de collectivistische proefpersonen positiever zouden oordelen over de retrospectieve hardopdenkprotocollen dan de individualistische proefpersonen (H3), en dat de individualistische proefpersonen juist positiever zouden zijn over de plus-en-minmethode (H4). Beide hypothesen werden in het onderzoek niet bevestigd. De vijf vragen over de retrospectieve hardopdenkprotocollen en de vier vragen over de plus-en-minmethode vormden geen betrouwbare schalen en moesten dus apart worden geanalyseerd. Op alle vragen vonden we geen significante verschillen tussen de twee groepen proefpersonen. Over het algemeen oordeelden de proefpersonen positief over beide evaluatiemethoden.

De vier vragen waarin de proefpersonen de retrospectieve hardopdenkprotocollen en de plus-en-minmethode moesten vergelijken, vormden twee adequate schalen. De ene was gericht op de ervaringen als proefpersoon (Cronbach's alfa = .60); de andere op de inschattingen van het belang van de zelf gegeven feedback (Cronbach's alfa = .67). Ook voor deze schalen vonden we geen significante verschillen tussen de twee groepen proefpersonen, maar bij de inschatting van het belang van de gegeven feedback vonden we wel een opmerkelijke, niet verwachte tendens: de collectivistische proefpersonen neigden naar een positiever oordeel over de plus-en-minmethode dan de individualistische proefpersonen (t-toets, $t=1.975, df=35, p=.056$).

5.4 Schuldvraag bij gebruikersproblemen. We verwachtten dat collectivistische proefpersonen meer dan de individualistische proefpersonen geneigd zouden zijn om de schuld van

gebruikersproblemen bij zichzelf te zoeken (H5). Dit is onderzocht aan de hand van negen vragen in de vragenlijst aan het einde van elke sessie. Drie vragen stonden voor interne schuldtoekenning, zes vragen voor externe schuldtoekenning (waarvan drie betrekking hadden op de kwaliteit van de website en drie op de testsituatie). De vragen over externe schuldtoekenning vormden twee adequate schalen (Cronbach's alfa = .64 voor de kwaliteit van de website, en .61 voor de testsituatie); de vragen over interne schuldtoekenning niet. Zoals te zien is in tabel 3, hebben we geen significante verschillen gevonden tussen de beide groepen proefpersonen over de schuldvraag bij gebruikersproblemen. Anders dan in het onderzoek van Schriver (1997) bleken de proefpersonen in ons onderzoek niet geneigd om vooral zichzelf de schuld te geven van de problemen waar ze tegenaan liepen. De gunstige oordelen van de proefpersonen over de testsituatie maakt duidelijk dat de opzet van het retrospectieve hardopdenkonderzoek in hun ogen geslaagd was: de taken waren realistisch en de testsituatie was niet storend.

Tabel 3. Waar ligt de schuld voor de gebruikersproblemen in het hardopdenkonderzoek?

	Individualistisch	Collectivistisch	Significantie
Extern: de onnatuurlijke testsituatie (gemiddelde van drie items)	4.4	4.4	n.s.
Extern: de kwaliteit van het <i>Web of Science</i> (gemiddelde van drie items)	3.2	3.3	n.s.
Intern: gebrek aan ervaring met databases	4.4	4.5	n.s.
Intern: gebrek aan ervaring met het <i>Web of Science</i>	2.7	2.5	n.s.
Intern: niet goed gelezen	3.0	3.3	n.s.

N.B: Scores op een vijfpuntsschaal (1 = mee eens – 5 = mee oneens)

5.5 Proefpersoonrollen in het hardopdenkonderzoek. De volgende hypothese richtte zich op het gedrag van de proefpersonen tijdens het retrospectieve hardopdenkonderzoek. We verwachtten dat de collectivistische proefpersonen minder geneigd zouden zijn om af te wijken van de (door de methode veronderstelde) typische gebruikersrol dan de individualistische proefpersonen (H6). Daarbij onderscheidde we drie mogelijke andere rollen die de proefpersonen zouden kunnen aannemen: proefpersoon, internetgebruiker en reviewer. Hieronder hebben we uit de protocollen een aantal voorbeelden van elke rol opgenomen.

Uitingen vanuit de rol van proefpersoon:

I was doing a few checks. That's why I didn't complete all the questions.

I learned something from this today. I should pay more attention.

It's a stupid way of doing it. But it's the only way I know. I'm getting annoyed with myself.

Don't look at this. This is so stupid. I even typed it wrong!

Uitingen vanuit de rol van internetgebruiker:

Normally, I never use save.

Normally, I would ask someone else whether they know how to do it.

I always press back on the browser, not on the page.

That's what I usually do. In my field of research ...

Uitingen vanuit de rol van reviewer:

I think this shouldn't be so strict.

That's not very good that you have to use a list. How difficult can it be to implement...

That was the main problem: enter.

It's not one of my favorite databases.

De resultaten zijn te vinden in tabel 4. Wanneer gekeken wordt naar alle afwijkingen van de standaard-gebruikersrol, wordt de hypothese bevestigd: de collectivistische proefpersonen hielden zich meer aan de gebruikersrol dan de individualistische proefpersonen. De gevonden waarde voor Cohen's *d* duidt op een gemiddeld tot groot effect. Kijken we naar de afzonderlijke niet-gebruikersrollen, dan is er geen sprake van significante verschillen. Op grond van dit resultaat mag verwacht worden dat individualistische proefpersonen aan een usability test een grotere diversiteit aan problemen naar voren zullen brengen dan collectivistische proefpersonen. Collectivistische proefpersonen houden zich meer aan de gebruikersrol die voortvloeit uit de taken die ze krijgen.

Table 4. Gemiddeld aantal afwijkingen van de gebruikersrol in de hardopdenkprotocollen

	Individualistisch	Collectivistisch	Significantie
Proefpersoon	3.1	2.2	n.s.
Internetgebruiker	2.1	1.1	n.s.
Reviewer	1.4	0.7	n.s.
Totaal	6.6	4.0	t-toets (eenzijdig), $t=1.818$, $df=35$, $p<.05$, Cohen's $d=0.61$

5.6 Directe en indirecte formuleringen bij het geven van commentaar. We verwachtten dat proefpersonen uit collectivistische landen vaker dan de individualistische proefpersonen zouden kiezen voor indirecte en eufemistische formuleringen (H7). Een eerste opmerkelijke bevinding betreft het aantal commentaren dat door de twee groepen proefpersonen is gemaakt. Geheel in lijn met hypothese 6 was het overgrote deel van de commentaren in de retrospectieve hardopdenkprotocollen afkomstig van individualistische proefpersonen (70% tegenover 30%).

Hieronder zijn voorbeelden van directe en indirecte commentaren opgenomen (samen met de gemiddelde scores op de schaal direct-indirect (1-5), die door twaalf studenten Toegepaste Communicatiewetenschap zijn toegekend). Er bleek sprake van een significant verschil tussen de beide groepen, in de richting van onze hypothese. Commentaren afkomstig van collectivistische proefpersonen hadden een gemiddelde score op de schaal direct-indirect van 2,8; commentaren van individualistische proefpersonen eindigden met een gemid-

delde score van 2,2. Het verschil is statistisch significant (t-toets, $t=2.507$, $df=37$, $p<.05$) en correspondeert met een groot effect (Cohen's $d = .89$).

Directe formuleringen van commentaar:

Stupid thing (score 1,1).

That's very annoying (score 1,2).

Doesn't work properly. Annoying (score 1,3).

Of course, it still didn't work (score 1,4)

Indirecte formuleringen van commentaar:

There's something strange ... I guess it's not important (3,5)

It's not the most convenient way ... (3,3)

I think this shouldn't be so strict (3,2)

It's not one of my favorite databases (3,2)

6. Conclusies en discussie

In deze paragraaf zullen we eerst conclusies trekken over de effecten die culturele verschillen op de dimensie individualisme–collectivisme hebben op de feedback die verzameld wordt in een pretest. Daarna bespreken we het opvallende verschil in ons onderzoek tussen het feitelijke gedrag van proefpersonen en hun antwoorden op opiniërende vragen. Tot slot gaan we, op grond van onze ervaringen, in op de problematiek van intercultureel onderzoek.

6.1 Culturele invloeden op website-evaluatie. De belangrijkste conclusie uit ons onderzoek is dat de culturele achtergrond van de proefpersonen inderdaad een factor is die van invloed kan zijn op de feedback die in een website-evaluatie wordt verzameld. Hoewel de IDV-index geen verschillen liet zien tussen de twee groepen in ons onderzoek, bleken de proefpersonen die op grond van hun nationaliteit een collectivistische respectievelijk individualistische cultuur vertegenwoordigden, zich in een aantal gevallen overeenkomstig de verwachtingen te gedragen. Het advies dat in veel webdesign-literatuur wordt gegeven om internationale websites te evalueren met gebruikersgroepen uit verschillende landen blijft waardevol, maar daarbij moet de kanttekening worden gemaakt dat de waarde van de evaluatiemethoden die in dergelijk onderzoek gebruikt worden, eveneens cultuurafhankelijk kan zijn.

Dat geldt met name voor de plus-en-minmethode, die bij proefpersonen uit individualistische, weinig contextgevoelige culturen meer commentaren uitlokt dan bij proefpersonen uit collectivistische, contextgevoelige culturen. Opvallend genoeg kwam dit duidelijke verschil niet tot uiting in de oordelen die de collectivistische proefpersonen gaven over de plus-en-minmethode: ze waren zelf verhoudingsgewijs optimistisch over de waarde van hun plus-en-mincommentaar. In het tot dusverre beschikbare onderzoek, waarin de plus-en-minmethode steeds als waardevol pretestinstrument naar voren komt, werd altijd gewerkt met West-Europese of Amerikaanse proefpersonen (zie De Jong 1998). Op grond van de

Multiculturele website-evaluatie

resultaten van het hier gepresenteerde onderzoek lijkt de conclusie gerechtvaardigd dat de methode wellicht minder geschikt is in collectivistische, contextgevoelige culturen. Overigens is in ons onderzoek alleen de kwantiteit van het commentaar gemeten; het is niet uit te sluiten dat de vergelijking anders uitvalt als ook de kwaliteit van het plus-en-mincommentaar wordt meegenomen.

De retrospectieve hardopdenkprotocollen lijken minder te worden beïnvloed door culturele verschillen tussen proefpersonen, doordat de fouten die proefpersonen maken de ruggengraat van de methode vormen en er op voorhand geen redenen zijn om te veronderstellen dat proefpersonen uit de ene cultuur meer fouten maken dan proefpersonen uit de andere cultuur. Toch vonden we ook hier twee effecten van de culturele achtergrond van de proefpersonen. Ten eerste waren de proefpersonen uit de collectivistische, contextgevoelige culturen meer geneigd om dicht bij de in het onderzoek opgelegde gebruikersrol te blijven. Waar de proefpersonen uit de individualistische, weinig contextgevoelige culturen regelmatig commentaren gaven vanuit andere perspectieven (kritiek geven op de website vanuit de rol van reviewer, commentaar geven op de testsituatie of reflecteren op de dingen die ze normaal gesproken met het internet of met databases zouden doen), waren de proefpersonen uit de collectivistische, contextgevoelige culturen meer geneigd om in het hele hardopdenkonderzoek de aangereikte gebruikersrol aan te houden. Men zou kunnen zeggen dat ze zich als de ideale proefpersonen voor een usability test gedroegen, maar ook het commentaar van proefpersonen die even buiten de gebruikersrol traden, bevatte vaak waardevolle suggesties voor de webdesigner. Ten tweede bleken de proefpersonen uit de collectivistische, contextgevoelige culturen geneigd om hun commentaren minder direct te formuleren dan de individualistische proefpersonen. Dit zou in de praktijk consequenties kunnen hebben in de revisiefase, waarin de onderzoeker(s) en webdesigner(s) moeten inschatten hoe ernstig de ontdekte problemen zijn.

Op een wat abstracter niveau blijkt het onderscheid tussen collectivistische, contextgevoelige en individualistische, weinig contextgevoelige culturen, zoals geoperationaliseerd door Ting-Toomey (1998), een vruchtbare benadering om voorspellingen te doen over verschillen tussen West-Europese en Aziatische of Afrikaanse proefpersonen. Niet alle hypothesen werden in ons onderzoek bevestigd, maar er was duidelijke ondersteuning voor vier van de hypothesen die we op grond van Ting-Toomey hadden opgesteld.

6.2 Verschillen tussen gedrag en zelfrapportage. Een opmerkelijke discrepantie in onze data betreft het verschil tussen gedrag en zelfrapportage. De drie hypothesen die in ons onderzoek niet werden bevestigd (namelijk over de waardering van de twee gebruikte evaluatiemethoden en de schuldvraag bij gebruikersproblemen) zijn getoetst door middel van zelfrapportage. De drie significante verschillen die we wél hebben gevonden (het aantal plus-en-minproblemen, de afwijkingen van de opgelegde gebruikersrol en de directheid van de commentaren) waren gebaseerd op het feitelijke gedrag van de proefpersonen. Dit verschil kan verklaard worden aan de hand van Hofstede's (1994) 'ui-diagram' van cultuurkenmerken, dat veronderstelt dat de culturele oriëntatie van mensen gelaagd is (met waarden als de kern en symbolen als de buitenste laag). Het gedrag van de proefpersonen lijkt een meer fundamentele laag van hun culturele oriëntatie te vertegenwoordigen dan de antwoorden die ze geven in een vragenlijst. De antwoorden in de vragenlijst kunnen vertekend zijn door sociale wenselijkheid. En sociale wenselijkheid blijkt ook weer een factor die gerelateerd is aan culturele kenmerken: onderzoek van Middleton & Jones (2000) maakt

duidelijk dat de invloed van sociale wenselijkheid groter is in collectivistische culturen dan in individualistische culturen. Zo geredeneerd is te verwachten dat de invloed van de culturele achtergrond op het gedrag en met name de attitude van proefpersonen in een pre-test groter is wanneer die proefpersonen worden onderzocht in hun land van herkomst, zonder dat selectie- en assimilatieprocessen invloed hebben gehad.

Deze invloed van sociale wenselijkheid zou ook kunnen gelden voor de scores op de IDV-index, die immers ook berusten op zelfrapportage. Dit is een reden te meer om aan het ontbreken van verschillen op deze index niet te veel consequenties te verbinden.

6.3 Intercultureel communicatieonderzoek. Onze ervaringen in dit onderzoek roepen twijfel op over twee populaire benaderingen in intercultureel onderzoek. De eerste betreft het inschakelen van immigranten als getrouwe representanten van de cultuur van hun land van herkomst. Als gevolg van selectie- en assimilatieprocessen kunnen deze immigranten behoorlijk afwijken van de culturele oriëntatie in hun oorspronkelijke land. Deze afwijkingen zullen het eerst merkbaar zijn in vragenlijstonderzoek. Natuurlijk hangt veel af van het doel van het onderzoek: als het niet gaat om wereldwijde culturele verschillen maar om verschillen binnen de nationale context, zoals in het onderzoek van Lentz & Hulst (2000), is er natuurlijk geen probleem. De tweede twijfel betreft het gebruik van vragenlijsten, zoals de IDV-index, om culturele verschillen te meten. Dergelijke instrumenten richten zich primair op de buitenste lagen van cultuur en dringen niet gemakkelijk door tot de waarden en gewoonten die voor een belangrijk deel bepalend zijn voor het gedrag.

Hoewel we op grond van ons onderzoek een aantal duidelijke conclusies hebben kunnen trekken over de invloed van culturele verschillen op website-evaluatie, hebben onze ervaringen ons ook bewust gemaakt van de geweldige complexiteit van intercultureel onderzoek. Culturele verschillen kunnen op veel verschillende manieren invloed hebben, zowel op de communicatie als op het onderzoek zelf. Een vergelijking met de bekende Russische matruschkapoppen ligt voor de hand: als de eerste pop geopend wordt, verschijnt er een tweede, identiek maar een beetje kleiner; als de tweede pop geopend wordt, komt de derde te voorschijn, etcetera. Ons interculturele onderzoek behelsde het openen van de tweede matruschkapop, maar terwijl we daarmee bezig waren, werden we ons al bewust van de derde en de vierde. Zoals we eerder stelden, spelen interculturele verschillen mogelijk ook weer een rol bij de manier waarop de proefpersonen vragenlijsten invullen (sociale wenselijkheid). En weer een ander onderzoek dat we recentelijk op het spoor kwamen, laat zien dat de culturele overeenkomst tussen proefleider en proefpersonen mogelijk ook van belang is. In het onderzoek van Vatrappu (2002) bleken usability interviews vruchtbaarder als de onderzoeker en de proefpersoon een gedeelde culturele achtergrond hadden. In ons onderzoek werden alle sessies 'gewoon' door dezelfde Nederlandse onderzoekster geleid.

We hopen met dit onderzoek te hebben aangetoond dat multiculturele aspecten relevant zijn voor de opzet en uitvoering van een website-evaluatie. Naast de bovengenoemde methodologische complicaties schuilt er in dit type onderzoek nog een ander probleem: het gevaar dat de resultaten leiden tot een stereotiep beeld van de groepen proefpersonen en hun culturele oriëntatie. We benadrukken dat we het hier hebben over verschillen tussen groepen, niet tussen individuen. Ondanks alle mogelijke bezwaren zijn we van mening dat intercultureel onderzoek naar website-evaluatie in de huidige context van internationalisering essentieel is voor een beter begrip van de waarde en beperkingen van gangbare evaluatiemethoden.

- Arnold, M. (1998).** Building a truly World Wide Web: A review of the essentials of international communication. *Technical Communication*, 45, 197-206.
- Barnum, C.M. (2002).** *Usability testing and research*. New York: Longman.
- Becker, S.A. (2002).** An exploratory study on Web usability and the internationalization of US e-business. *Journal of Electronic Commerce Research*, 3, 265-278.
- Brown, P. & S.C. Levinson (1990).** *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Diamantopoulos, A., N. Reynolds & B. Schlegelmilch (1994).** Pretesting in questionnaire design: The impact of participant characteristics on error detection. *Journal of Marketing Research*, 36, 295-311.
- Dumas, J.C. & J.S. Redish (1993).** *A practical guide to usability testing*. Norwood, NJ: Ablex.
- Haak, M. van den, M. de Jong & P.J. Schellens (2003).** Hardopdenkprotocollen als pretestmethode: Synchroon en retrospectief hardopdenken vergeleken. *Tijdschrift voor Taalbeheersing*, 25, 236-252
- Hall, E. T. (1977).** *Beyond culture*. Garden City, NY: Anchor Press/Doubleday.
- Hoc, J.M. & J. Leplat (1983).** Evaluation of different modalities of verbalisation in a sorting task. *International Journal of Man-Machine Studies*, 18, 283-306.
- Hofstede, G. (1994).** *Culture and organizations: Software of the mind*. London: Harper Collins.
- Hofstede, G. (2001).** *Culture's consequences: Comparing values, behaviours, institutions, and organizations across nations*. Second edition. Beverly Hills, CA: Sage.
- Hoft, N. (1995).** *International technical communication: How to export information about high technology*. New York: John Wiley.
- Jong, M. de (1998).** *Reader feedback in text design. Validity of the plus-minus method for the pretesting of public information brochures*. Amsterdam: Rodopi.
- Jong, M. de & P.J. Schellens (1995).** *Met het oog op de lezer. Pretestmethoden voor schriftelijk voorlichtingsmateriaal*. Amsterdam: Thesis.
- Jong, M. de & P.J. Schellens (1997).** Reader-focused text evaluation: An overview of goals and methods. *Journal of Business and Technical Communication*, 11, 402-432.
- Jong, M. de & P.J. Schellens (2000).** Toward a document evaluation methodology: What does research tell us about the validity and reliability of methods? *IEEE Transactions on Professional Communication*, 43, 242-260.
- Jong, M. de & P.J. Schellens (2001).** Readers' background characteristics and their feedback on documents: The influence of gender and educational level on evaluation results. *Journal of Technical Writing & Communication*, 31, 267-281.
- Jong, M. de & P.J. Schellens (2002).** Tekstevaluatie. Onderzoek naar de validiteit van probleemopsporende methoden. *Tijdschrift voor Taalbeheersing*, 24, 146-166.
- La Ferle, C., S.M. Edwards & Y. Mizuno (2002).** Internet diffusion in Japan: Cultural considerations. *Journal of Advertising Research*, 42, 2, 65-79.
- Lentz, L. & J. Hulst (2000).** Babel in document design: the evaluation of multilingual texts. *IEEE Transactions on Professional Communication*, 43, 313-322.
- Marcus, A. & E.W. Gould (2000).** Cultural dimensions and global web user-interface design: What? So what? Now what? *Proceedings of the 6th Conference on Human Factors and the Web*. http://www.tri.sbc.com/hfweb/marcus/hfweb00_marcus.html.
- Middleton, K.L., & J.L. Jones (2000).** Socially desirable response sets: the impact of country culture. *Psychology & Marketing*, 17, 149-163.
- Nielsen, J. (1993).** *Usability engineering*. Boston, MA: Academic Press.
- Nielsen, J. (2000).** *Designing Web usability: The practice of simplicity*. Indianapolis, IN: New Riders.

- Okazaki, S. & J.A. Rivas (2002).** A content analysis of multinationals' Web communication strategies: Cross-cultural research framework and pre-testing. *Internet Research*, 12, 380-390.
- O'Keefe, R.M., e.a. (2000).** From the user interface to the consumer interface: Results from a global experiment. *International Journal of Human-Computer Studies*, 53, 611-628.
- Peng, K., R.E. Nisbett & N.Y.C. Wong (1997).** Validity problems comparing values across cultures and possible solutions. *Psychological Methods*, 2, 329-344.
- Rubin, J. (1994).** *Handbook of usability testing. How to plan, design, and conduct effective tests*. New York: John Wiley.
- Schriver, K.A. (1989).** Evaluating text quality: The continuum from text-focused to reader-focused methods. *IEEE Transactions on Professional Communication*, 32, 238-255.
- Schriver, K.A. (1997).** *Dynamics in document design. Creating text for readers*. New York: John Wiley.
- Schweibenz, W. & F. Thissen (2003).** *Qualität im Web. Benutzer-freundliche Webseiten durch Usability Evaluation*. Berlin: Springer.
- Sienot, M. (1997).** Pretesting Web sites. A comparison between the plus-minus method and the think aloud method for the World Wide Web. *Journal of Business and Technical Communication*, 11, 469-482.
- Simon, J.S. (2001).** The impact of culture and gender on Web sites. *Data Base for Advances in Information Systems*, 32, 18-37.
- Smith, P.B. & M.H. Bond (1998).** *Social psychology across cultures*. Second edition. London: Prentice Hall.
- Someren, M.W van, Y.F. Barnard & J.A.C. Sandberg (1994).** *The think aloud method. A practical guide to modeling cognitive processes*. London: Academic Press.
- Ting-Toomey, S. (1998).** Intercultural conflicts styles. A face-negotiation theory. In: Y.Y. Kim & W.B. Gudykunst (eds.), *Theories in intercultural communication* (pp. 213-235). Newbury Park, CA: Sage.
- Trompenaars, F. & C. Hampden-Turner (1998).** *Riding the waves of culture. Understanding cultural diversity in global business*. New York: McGraw-Hill.
- Vatrapu, R. (2002).** Culture and international usability testing: The effects of culture in interviews. Master's thesis, Virginia Polytechnic Institute and State University. http://scholar.lib.vt.edu/theses/available/etd-09132002-083026/unrestricted/Vatrapu_Thesis.pdf.
- Versseveld, C. van (1995).** Betrokkenheid bij voorlichting. De invloed van betrokkenheid op pretest-commentaar. *Tekst[blad]*, 1 (3), 40-44.
- Waes, L. van (2000).** Thinking aloud as a method for testing the usability of Web sites: The influence of task variation on the evaluation of hypertext. *IEEE Transactions on Professional Communication*, 43, 279-291.
- Zahedi, F., W.V. van Pelt & J. Song (2001).** A conceptual framework for international Web design. *IEEE Transactions on Professional Communication*, 44, 83-103.