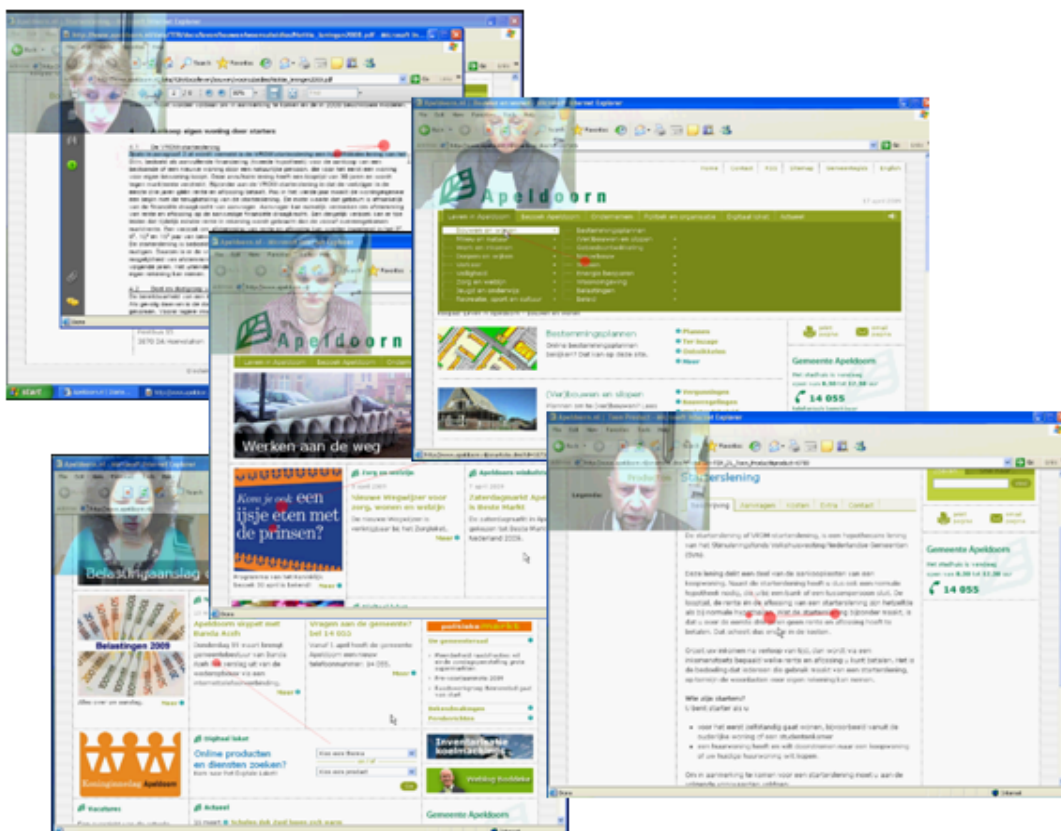


# Zie ik meer dan jij zegt?

Wat voegen de verbalisaties van proefpersonen toe in gebruikersonderzoek waarbij ook het scherm en de oogbewegingen opgenomen worden?



Studente: Ineke Waas, 0313378  
Opleiding: MA Communicatiestudies (Tekst en communicatie)  
Faculteit: Geesteswetenschappen  
Instituut: Universiteit Utrecht  
Begeleider: drs. S.K. Elling  
Tweede lezer: dr. L.R. Lentz  
Datum: 13-7-2010

## **Abstract**

Voor deze Masterscriptie is onderzocht in hoeverre schermobservaties (muis- en oogbewegingen) te gebruiken zijn als zelfstandige probleemopsporende evaluatiemethode. Dit is gedaan door te achterhalen hoeveel problemen proefpersonen tegenkwamen terwijl ze een taak uitvoerden op de website van de gemeente Apeldoorn. Hierbij zijn de aantallen problemen die door observaties gevonden zijn vergeleken met de aantallen problemen die uit de verbalisaties van het hardopdenken naar voren zijn gekomen. Het grootste deel van de problemen is via beide methoden gevonden worden. Ze leveren echter ook beide een aantal unieke problemen op. Bij de verbalisaties zijn dit er meer dan bij de observaties. Via zowel de observaties als de verbalisaties is een verscheidenheid aan soorten problemen gevonden. Problemen met de Inhoud worden echter vaker geverbaliseerd dan geobserveerd. Hierbij maakt het nauwelijks uit welke variant van de hardopdenkmethode gebruikt wordt, er zijn nauwelijks verschillen tussen de condities. Voor de verschillende leeftijdscategorieën, opleidingsniveaus en geslachten zijn wel verschillen gevonden.

# Inhoudsopgave

<b>1 Inleiding</b> .....	<b>5</b>
<b>2 Theoretisch kader</b> .....	<b>6</b>
2.1 Hardopdenkmethode .....	6
2.1.1 Vergelijking concurrent vs retrospectief hardopdenken .....	6
2.1.2 Vergelijking scherm vs oogbewegingen als cue .....	7
2.1.3 Onvolledige verbalisaties .....	8
2.2 Observaties om hardopdenkmethodes te testen .....	8
2.3 Observaties als zelfstandige methode .....	9
2.3.1 Handelingen met de muis als zelfstandige methode .....	9
2.3.2 Oogbewegingen als zelfstandige methode .....	10
<b>3 Onderzoeksvraag en hypothesen</b> .....	<b>15</b>
<b>4 Methode</b> .....	<b>17</b>
4.1 Materiaal .....	17
4.1.1 Website gemeente Apeldoorn (www.apeldoorn.nl) .....	17
4.2 Proefpersonen .....	20
4.3 Analyse filmpjes .....	20
4.3.1 Observaties .....	20
4.3.2 Verbalisaties .....	22
4.3.3 Procedure .....	22
4.4 Vergelijkende analyses .....	26
4.4.1 Criteria identieke problemen .....	26
4.4.2 Alle probleemdetecties .....	26
4.4.3 Alle unieke problemen .....	27
4.4.4 Unieke problemen bij meerdere personen .....	27
4.4.5 Analyses .....	27
<b>5 Resultaten</b> .....	<b>29</b>
5.1 Algemene gegevens .....	29
5.2 Problemen via observaties en/of verbalisaties .....	29
5.2.1 Observatie versus verbalisatie .....	29
5.2.2 Conclusie .....	32
5.3 Verschillen tussen CTA, RTA en RTE .....	32
5.3.1 Observatie versus verbalisatie .....	33
5.3.2 Conclusie .....	37
5.4 Verschillen tussen proefpersonen .....	37
5.4.1 Leeftijd .....	37
5.4.2 Opleiding .....	38
5.4.3 Geslacht .....	39
5.4.4 Conclusie .....	40
<b>6 Conclusies</b> .....	<b>41</b>
6.1 Beantwoording van de deelvragen .....	41
6.1.1 Problemen via observaties en verbalisaties .....	41
6.1.2 Verschillen tussen CTA, RTA en RTE .....	42
6.1.3 Verschillen tussen proefpersonen .....	43
6.2 Antwoord op de hoofdvraag .....	44
<b>7 Discussie</b> .....	<b>45</b>
7.1 Probleempunten .....	45
7.2 Suggesties voor vervolgonderzoek .....	46
<b>Bronvermelding</b> .....	<b>47</b>

**Bijlage I: Taakmodel**  
**Bijlage II: Voorbeeldanalyse**

**Cd-rom**

Databestanden SPSS  
SPSS output van de analyses  
Analyses van de filmpjes  
Alle probleembeschrijvingen

# 1 Inleiding

De usability van websites is tegenwoordig een belangrijk onderwerp. Sinds 2001 wordt er in Nederland elk jaar een award uitgereikt voor de meest gebruikersvriendelijke website van Nederland ([www.usabilityaward.nl](http://www.usabilityaward.nl)). In de eerste plaats worden er door een vakjury een aantal genomineerden geselecteerd uit alle inzendingen. Vervolgens mogen gebruikers elk van deze sites een rapportcijfer geven en wordt op basis daarvan de winnaar uitgeroepen. Maar wanneer is een site nu precies gebruikersvriendelijk en hoe kan een site die het niet is verbeterd worden zodat de gebruikers deze beter gaan waarderen. Hiervoor moet onderzocht worden wat de problemen met de site zijn. Hierbij is het heel nuttig om te kijken welke problemen de daadwerkelijke gebruikers van een website tegenkomen terwijl zij een taak proberen uit te voeren. Om dit te doen wordt vaak gebruik gemaakt van de hardopdenkmethode. Hier zijn echter een aantal nadelen aan verbonden.

Zo leidt hardopdenken tijdens het uitvoeren van een taak (*concurrent thinking aloud*) ertoe dat het werkproces anders is dan normaal (onder andere Ball e.a. 2007, Guan e.a. 2006). Wanneer men eerst in stilte werkt en achteraf hardop nadenkt (*retrospective thinking aloud*) is dit probleem opgelost, maar doet zich een ander probleem voor: men is sterk afhankelijk van het geheugen. Het is dus mogelijk dat men deels vergeten is wat men dacht en hierdoor kan men dingen weglaten of verzinnen (Ball e.a. 2007). Door extra aanknopingspunten te geven, zoals oogbewegingen, kan men helpen zaken te herinneren. Maar nog steeds is het mogelijk om zaken weg te laten.

Het zou ideaal zijn als het mogelijk zou zijn om de gebruiker gewoon te laten werken, zonder dat hij beïnvloed wordt doordat hij hardop moet nadenken, of dat hij achteraf moet verzinnen wat hij ook al weer dacht toen hij bezig was. Dit zou kunnen als men problemen kan opsporen door de gebruiker alleen maar te observeren.

Hiervoor is het belangrijk om te weten of door observeren dezelfde problemen te vinden zijn als door hardopdenken. Er zijn verschillende onderzoeken gedaan waarin de problemen die voortkomen uit de hardopdenkmethode vergeleken worden met die uit observaties. Hierbij kijkt men dan meestal echter alleen naar wat de gebruiker doet, dus waar hij bijvoorbeeld op klikt of wanneer hij scrollt (Van den Haak e.a. 2003, 2004, 2007, 2009, Mueller & Lockerd 2001). Of de onderzoekers kijken voornamelijk naar de eyetracking gegevens, dus waar de gebruiker zijn blik op richt (Cooke & Cuddihy 2005, Guan e.a. 2005). In een aantal scripties van studenten van de Universiteit Utrecht wordt onderzocht of eyetracking als zelfstandige methode te gebruiken is (Vereijken 2007, Van der Burg 2008, De Bode 2008). Om problemen vast te stellen kijken zij enkel naar de oogbewegingen en waar de gebruiker op klikt, niet naar alles wat hij doet. Uit deze scripties is naar voren gekomen dat de oogbewegingen niet alle problemen weergeven die uit de hardopdenkmethode naar voren komen.

Ik wil kijken naar de problemen en hun oorzaken die geconstateerd kunnen worden op basis van de combinatie van schermobservaties en oogbewegingen. Mogelijk is er dan meer overlap tussen de observaties en het hardopdenken. Voor de cursus 'Navigatie op het web' heb ik een onderzoek uitgevoerd waarbij ik heb gekeken of het mogelijk is om problemen en hun oorzaken te achterhalen wanneer er alleen gekeken wordt naar wat de gebruiker doet. Dit bleek inderdaad mogelijk te zijn. Door naar de muisbewegingen en klikken en de oogbewegingen te kijken kon ik bepalen of er sprake was van een probleem en of dit veroorzaakt werd doordat de gebruiker bijvoorbeeld iets niet zag of niet las.

Door deze observaties naast de verbalisaties te leggen wil kijken of hiertussen verschillen zijn en of het mogelijk zou kunnen zijn om het hardopdenken achterwege te laten wanneer men alle observaties van een gebruiker samenvoegt. Op deze manier kan de usability van een website mogelijk onderzocht worden zonder dat de proefpersoon daar hinder van ondervindt.

In het volgende hoofdstuk bespreek ik de hiervoor genoemde methoden en onderzoeken uitgebreider. Daarna beschrijf ik in hoofdstuk 3 mijn onderzoeksvragen en hypotheses. Vervolgens ga ik in op de methode en geef ik in hoofdstuk 5 de resultaten weer. Tot slot beantwoord ik mijn onderzoeksvragen en volgt de discussie.

## 2 Theoretisch kader

In de inleiding heb ik al aangegeven dat er verschillende methoden zijn om problemen op te sporen met websites, zoals de hardopdenkmethode en het observeren van het scherm en de oogbewegingen terwijl de gebruiker aan het werk is. In §2.1 beschrijf ik de hardopdenkmethode. Vervolgens ga ik in §2.2 in op onderzoek waarbij de hardopdenkmethode en observaties gecombineerd worden. Tot slot bespreek ik in §2.3 onderzoeken waarin gekeken wordt in hoeverre observaties als zelfstandige methode gebruikt kunnen worden.

### 2.1 Hardopdenkmethode

Bij de hardopdenkmethode moet de gebruiker hardop vertellen over wat hij denkt (of dacht) tijdens het uitvoeren van een taak. Op deze manier hoopt men inzicht te krijgen in de cognitieve processen die zich afspelen tijdens het werken en zo vast te stellen welke problemen zich voordoen.

'Hardopdenkmethode' is een overkoepelende naam voor verschillende manieren waarop men hardop nadenkt over het uitvoeren van een taak. Over het algemeen wordt er bij usability onderzoek gebruik gemaakt van drie verschillende varianten. De eerste is de *concurrent thinking aloud* (CTA) methode. Hierbij moet de gebruiker tijdens het uitvoeren van een taak vertellen wat hij denkt. De tweede variant is de *retrospective thinking aloud* (RTA) methode. Dan werkt de gebruiker eerst in stilte aan de taak en achteraf kan hij aan de hand van beelden van wat hij deed vertellen wat hij dacht. Deze beelden bevatten dan de verschillende pagina's die hij geopend heeft en de bewegingen en klikken die hij met de muis gemaakt heeft. Bij de derde variant, de *retrospective thinking aloud met eyetracking* (RTE) methode, vertelt de gebruiker ook achteraf wat hij dacht. Hier krijgt hij echter niet alleen de beelden van wat hij deed als cue, maar ook zijn oogbewegingen.

Elk van deze methoden heeft zijn voor- en nadelen. Deze zullen in de volgende paragrafen besproken worden.

#### 2.1.1 Vergelijking concurrent vs retrospectief hardopdenken

Veel onderzoeken noemen als belangrijkste nadeel van het concurrent hardopdenken de invloed die het heeft op het uitvoeren van taken. Er zou sprake zijn van 'reactiviteit' (van den Haak e.a. 2007, Ball e.a. 2007). Deze kan zich op twee manieren voordoen. Het is mogelijk dat het hardopdenken een negatieve invloed heeft. Doordat de proefpersoon twee dingen tegelijkertijd moet doen, hardopdenken en taken uitvoeren, wordt hij dubbel belast. Hierdoor kan hij niet alle nodige aandacht besteden aan het uitvoeren van de taak en is hij dus minder succesvol. Het is echter ook mogelijk dat de proefpersoon informatie juist beter verwerkt doordat hij hardop moet vertellen wat hij denkt. Dan heeft de methode een positieve invloed op de taakuitvoering.

Uit onderzoek blijkt dat het afhankelijk is van de manier waarop proefpersonen aangemoedigd worden in hoeverre deze reactiviteit zich voordoet. Hertzum e.a. (2009) hebben in hun onderzoek de klassieke en relaxte manier van hardopdenken vergeleken (beide steeds met een conditie in stilte werken).

De klassieke manier van hardopdenken is gebaseerd op de ideeën van Ericsson & Simon (beschreven in Hertzum e.a. 2009, Boren & Ramey 2000). Zij onderscheiden drie niveaus van verbalisaties. Op het eerste niveau is er sprake van hardop spreken. Men verwoordt gedachten en informatie die al een verbale vorm heeft en waar de aandacht op gericht is. Een voorbeeld is het hardop uitrekenen van een som. Het tweede niveau is het daadwerkelijk hardopdenken. Hierbij moet informatie die niet verbaal is, zoals plaatjes of abstracte concepten, omgezet worden in woorden. Bij verbalisaties op niveau drie denkt men na over wat men al gedaan heeft, men geeft bijvoorbeeld een uitleg of verklaring van gedachten en gedrag. Volgens Ericson & Simon zijn vooral verbalisaties op niveau 1 en 2 voorbeelden van hardopdenken. Iemand die aan het werk is, moet hierbij niet afgeleid worden van wat hij aan het doen is door aanmoedigingen van de onderzoeker om hardop te blijven nadenken. De onderzoeker moet zich daarom beperken tot zinnigjes zoals 'blijf praten'. Boren & Ramey (2000) concluderen echter dat usability experts zich in de praktijk niet strikt aan deze ideeën houden. Deels komt dit door de aard van usability onderzoek. Er kunnen zich onverwachte zaken voordoen, zoals het

vastlopen van een systeem waarbij het noodzakelijk is voor de onderzoeker om in te grijpen en het hardopdenken van de proefpersoon te verstoren. Daarnaast zijn er ook onderzoekers die proefpersonen onderbreken om extra informatie te achterhalen. Ze willen specifiek een toelichting op wat er gebeurt. Bijvoorbeeld door vragen te stellen zoals 'waarom hebt u op deze knop geklikt?' en 'wat zoekt u?'. Hiermee wordt het werkproces onderbroken en daarmee ook het gedachteproces. De onderzoeker vraagt de proefpersoon om te gaan verbaliseren op niveau 3. Dit wordt door Hertzum e.a. (2009) de relaxte manier van hardopdenken genoemd.

Uit het onderzoek van Hertzum e.a. (2009) kwam dat de relaxte manier van hardopdenken meer invloed op het gedrag van de proefpersonen had dan de klassieke manier. Bij de relaxte manier werd er meer tijd besteed aan het scannen van de pagina en werd er meer geklikt en gescrolld dan bij het werken in stilte. Voor de klassieke manier werden geen verschillen gevonden met het in stilte werken. Voor de verschillende aspecten van de mentale werklast werd tussen de klassieke methode en in stilte werken geen algemeen verschil gevonden. In de klassieke methode werd wel op één van de aspecten een verschil gevonden. De mentale inspanning die vereist wordt bij het hardopdenken werd als hoger gezien dan bij het in stilte werken. Bij de relaxte methode waren er meer verschillen met het in stilte werken. De proefpersonen vonden dat mentaal en qua tijd meer van hen gevraagd werd, dat het meer moeite kostte, dat ze slechter presteerden en gefrustreerder raakten. Het maakt dus daadwerkelijk uit welke van de twee soorten aanmoediging gebruikt wordt.

Dit is ook te zien in de onderzoeken van Van den Haak e.a. (2007, 2009) en Ball (2007). Van den Haak e.a. (2007, 2009) vonden bij hun vergelijking van de CTA en RTA conditie bij het werken op een gemeente website geen verschillen voor het gemiddeld aantal geconstateerde problemen, het aantal goed uitgevoerde taken en de tijd die ze hiervoor nodig hadden, noch voor de waardering voor de manier van werken. Het maakte dus niet uit of de proefpersonen in de CTA of RTA conditie werkten. Zij hebben in hun instructie aan de proefpersoon staan: "The facilitator will remind you to keep talking aloud"<sup>1</sup> en gebruiken dus de klassieke manier. Ball e.a. (2007) daarentegen geven aan dat ze vragen stelden aan hun proefpersonen. Zij vonden ook daadwerkelijk bewijs van reactiviteit in hun onderzoek naar online zoekmachines. In de concurrent conditie was men minder succesvol bij het uitvoeren van de taken dan in de retrospectieve condities. Ook werd de concurrent gezien als minder aangenaam, minder snel, kon men minder geconcentreerd werken en werd de aanwezigheid van de onderzoeker negatiever gewaardeerd.

Een voordeel van de CTA methode ten opzichte van het retrospectief hardopdenken is dat wat men aan het doen is in het korte termijn geheugen zit. Bij de retrospectieve methode moet de proefpersoon uit zijn lange termijn geheugen ophalen wat hij gedaan heeft. Het is echter zeer waarschijnlijk dat hij niet alles onthouden heeft. Hierdoor is het mogelijk dat hij zaken vergeet, rationaliseert wat hij ziet en zo reconstrueert wat hij gedaan heeft of dingen verzint (Ball e.a. 2007).

### 2.1.2 Vergelijking scherm vs oogbewegingen als cue

Een probleem bij het gebruiken van de *retrospective thinking aloud* (RTA) methode op websites is dat de proefpersoon maar weinig informatie krijgt om zijn geheugen mee op te frissen (Van den Haak 2007). Vaak staat een tijd lang hetzelfde beeld op het scherm zonder dat er iets gebeurt. Door ook de oogbewegingen te laten zien (RTE) zou de proefpersoon meer geholpen kunnen worden, want hij heeft meer informatie over wat hij gedaan heeft.

Naar het verschil tussen de retrospectieve methoden is nog weinig onderzoek gedaan. Ball e.a. (2009) hebben zich hier al wel mee bezig gehouden. Zij hebben de twee retrospectieve methodes met elkaar en met *concurrent thinking aloud* vergeleken in hun onderzoek naar online zoekmachines. Tussen de retrospectieve methodes werden geen verschillen gevonden, maar in vergelijking met de CTA

---

<sup>1</sup> Er wordt door Van den Haak (2007, 2009) nergens expliciet genoemd op welke manier zij hun proefpersonen precies aanmoedigen om hardop te blijven denken. Ball ea (2007) verwijzen echter naar een ander onderzoek van Van den Haak ea uit 2003 waarin zij hetzelfde onderzoek hebben uitgevoerd als in 2007 en 2009 maar dan toegepast op online bibliotheekcatalogussen. Volgens Ball ea werd hierbij de strikte (klassieke) vorm van hardopdenken gebruikt. Van den Haak ea geven in hun onderzoeken uit 2007 en 2009 aan dat ze precies dezelfde werkwijze hebben aangehouden. Het is dus aannemelijk dat ze de klassieke manier hebben gebruikt.

methode wel. In de RTE conditie werden namelijk meer problemen gevonden dan in de CTA conditie, dit had vooral te maken met problemen met de begrijpelijkheid en de 'feedback' van de zoekmachine. Tussen de CTA en RTA conditie was echter geen verschil. In haar scriptie heeft Haas (2009) ook onderzocht of er een verschil was tussen de retrospectieve condities. Hieruit bleek dat er bij de RTE methode minder geverbaliseerd werd dan bij de RTA methode en dat proefpersonen soms zelfs afgeleid werden door de oogbewegingen. Verder waren er weinig verschillen. Zij heeft zich echter niet bezig gehouden met de problemen die uit de verbalisaties naar voren kwamen, maar puur naar het aantal en de soort verbalisaties gekeken.

### **2.1.3 Onvolledige verbalisaties**

Naast deze methodegebonden problemen zijn er ook nog een aantal zaken die voor hardopdenken in het algemeen gelden. Ball e.a. (2007) wijzen er bijvoorbeeld op dat verbalisaties onvolledig kunnen zijn. Dit kan bewust of onbewust gedaan worden door de proefpersoon. Het is mogelijk dat hij bepaalde zaken niet kan vertellen doordat hij zich er niet van bewust is. Hij kan echter ook zaken bewust weglaten. Bijvoorbeeld omdat hij de onderzoeker een plezier wil doen, hij wil hem bijvoorbeeld niet voor het hoofd stoten door problemen te noemen als hij weet dat de onderzoeker belangen heeft bij het programma dat getest wordt. Een andere mogelijkheid is dat hij zaken weglaat om een positief beeld van zichzelf over te brengen, hij wil bijvoorbeeld niet als dom overkomen (Van den Haak e.a. 2003). Hij kan ook denken dat sommige informatie zo vanzelfsprekend is dat hij denkt dat het niet genoemd hoeft te worden (Ball e.a. 2007). Een ander probleem is dat mensen sneller denken dan ze kunnen spreken (Ball e.a. 2007, Nielsen e.a. 2002), ze kunnen dus niet alles zeggen wat ze denken.

Nielsen e.a. (2002) hebben een aantal studenten de hardopdenkmethode laten evalueren. Zij noemden een deel van de bovenstaande punten, maar voegden daar nog aan toe dat sommige gedachten te complex zijn om onder woorden te brengen en dat hardopdenken onnatuurlijk is.

## **2.2 Observaties om hardopdenkmethodes te testen**

In verschillende onderzoeken worden de opnames van het scherm en de oogbewegingen niet (alleen) gebruikt om de proefpersoon te helpen bij het hardopdenken, maar ook om de onderzoeker een extra bron van informatie te geven. Hieronder beschrijf ik kort drie onderzoeken waar de onderzoekers observaties gebruiken om de hardopdenkmethode te testen.

Guan e.a. (2006) gebruiken de opnames van oogbewegingen bijvoorbeeld om te onderzoeken of de RTA methode valide is. Ze hebben een scherm in vlakken verdeeld (*areas of interest*) en gekeken in welke volgorde de proefpersonen hiernaar kijken. Vervolgens hebben ze de verbalisaties getranscribeerd en gekeken of de punten waarover men sprak en de volgorde hiervan overeen kwamen met wat de oogbewegingen lieten zien. Dit bleek inderdaad het geval te zijn. Het grootste deel van wat de proefpersonen zeiden kwam overeen met wat ze volgens de oogbewegingen gedaan hadden.

Hertzum e.a. (2009) hebben in hun vergelijking van de klassieke en relaxte hardopdenkmethode opnames gemaakt van het scherm en de oogbewegingen. Met de oogbewegingen hebben ze gekeken waar de proefpersonen naar kijken en hoelang. Daarnaast hebben ze bepaald welk soort visueel gedrag men vertoonde. Dit kon gefocust zijn, dan kijkt te proefpersoon naar een beperkt deel van de site. Daarnaast kon er sprake zijn van verspreid gedrag, in dat geval verspringt de blik van de proefpersoon over verschillende onderdelen op het scherm. Ze hebben behalve oogbewegingen ook de handbewegingen geregistreerd, de klikken met de muis, het scrollen en het schrijven. Hiermee hebben ze gekeken of het gedrag van de proefpersonen verschilde tussen de relaxte en klassieke manier van hardopdenken. Dit was inderdaad het geval, de relaxte manier leidde wel tot reactiviteit terwijl dit bij de klassieke manier niet zo was.

In de onderzoeken van Van den Haak e.a. (2003, 2004, 2007, 2009) zijn observaties naast verbalisaties gebruikt om problemen vast te stellen. Op basis van het aantal problemen dat door observaties, verbalisaties of beide vastgesteld kon worden hebben ze drie hardopdenkmethodes



vergeleken (CTA, RTA en Constructieve interactie). Bij de onderzoeken naar online bibliotheekcatalogussen (2003, 2004) vonden ze dat in de CTA conditie meer problemen geobserveerd werden dan in de RTA conditie. In de RTA conditie werden juist meer problemen geverbaliseerd. Bij de evaluatie van gemeentelijke websites (2007, 2009) waren er echter geen verschillen tussen de twee condities.

Bij alle in deze paragraaf besproken onderzoeken is de hardopdenkmethode de methode die onderzocht wordt. Er zijn echter ook studies die zich richten op de mogelijkheid om observaties zelfstandig te gebruiken. Deze bespreek ik in de volgende paragraaf.

## 2.3 Observaties als zelfstandige methode

Binnen observaties vallen zowel de handelingen op het scherm als de oogbewegingen van de proefpersoon. Er zijn een aantal onderzoeken uitgevoerd waarbij gekeken wordt in hoeverre de bewegingen met de muis te gebruiken zijn als zelfstandige methode, hier ga ik in §2.3.1 op in. De laatste jaren wordt vooral gekeken naar het nut van oogbewegingen. Onderzoeken hierover bespreek ik in §2.3.2.

### 2.3.1 Handelingen met de muis als zelfstandige methode

Met hun onderzoek wilden Mueller & Lockerd (2001) aantonen dat het niet alleen nuttig is om de klikken van de muis te registreren, maar ook te kijken naar andere muisbewegingen. Zij hebben een groep proefpersonen taken laten uitvoeren op verschillende websites. Hieruit hebben ze aantal trends voor de muisbewegingen gehaald. Eén daarvan, die van nut kan zijn bij usability onderzoek, is dat de locatie van de muis aan kan geven dat men twijfelt tussen bepaalde links of tekst. Bij twee van de taken moesten de proefpersonen een cd of dvd uitkiezen die ze zouden willen hebben. Op basis van de muisbewegingen is bepaald wat de tweede keus van de proefpersoon zou kunnen zijn (een andere cd of dvd waar de muis op is blijven staan). Dit kwam in 65% en 75% van de gevallen overeen met wat de proefpersonen achteraf als tweede keus noemden. De overige trends waren dat men de muis gebruikt om de onderdelen van een menu te markeren bij het doorkijken ervan. Verder waren er een aantal personen die zonder twijfel de muis op een link zetten, dit zien Mueller & Lockerd (2001) als een teken van bekendheid met de taak. Tot slot vonden ze dat de witte vlakken van een pagina worden gebruikt om de muiscursor op te plaatsen wanneer men deze niet nodig heeft.

Chen e.a. (2001) hebben onderzocht of de muisbewegingen ook echt laten zien waar proefpersonen naar kijken. De proefpersonen mochten vrij surfen op een aantal pagina's. Deze pagina's waren in regio's verdeeld (knop, menu, titel, figuur, banner, nergens). Vervolgens is gekeken hoeveel regio's bezocht worden met de ogen en met de muis. Dit is met elkaar vergeleken en daaruit bleek dat 84% van de regio's die bezocht werd door de muis ook bekeken zijn. Verder werd 88% van de regio's die niet bekeken werden ook niet met de muis bezocht. Ook vonden ze dat de ogen de muis volgen wanneer deze binnen een regio verplaatst en wanneer deze naar een andere regio gaat. Chen e.a. (2001) gaan er echter vanuit dat de muis ook daadwerkelijk over het scherm bewogen wordt. Uit een ander onderzoek blijkt echter dat dit niet altijd zo is.

Het onderzoek van Rodden e.a. (2008) houdt zich bezig met de patronen waarmee oogbewegingen en muisbewegingen gecoördineerd worden bij het scannen van een pagina met online zoekresultaten. Zij lieten een groep proefpersonen een aantal taken uitvoeren in de Google zoekmachine. Alleen kijkend naar de bewegingen van de muis deden zich twee typen gebruik voor. Aan de ene kant was dit wat zij incidenteel gebruik noemen. Hierbij werd er pas iets gedaan met de muis op het moment dat de proefpersoon besloten had wat hij wilde gaan doen. Dit was het geval bij "many of the visualisations<sup>2</sup>" (Rodden e.a. 2008, 2999). Aan de andere kant was er sprake van actief gebruik van de muis. Hierbij werden drie verschillende patronen gevonden:

- Met de muis horizontaal de ogen volgen over of net onder de tekst die gelezen wordt.

---

<sup>2</sup> Rodden ea (2008) geven geen exacte gegevens of percentages. Ook niet in een artikel waarnaar ze verwijzen (Rodden, K., & Fu, X. (2007) Exploring how mouse movements relate to eye movements on web search results pages. In: *Proc. ACM SIGIR 2007 Workshop on Web Information Seeking and Interaction*, 29-32.)

- Met de muis verticaal de ogen volgen, zodat deze ongeveer gelijk staat met het deel dat op dat moment gelezen wordt.
- Het markeren van een bepaald resultaat. De muis wordt dan op een bepaald punt geplaatst (waarschijnlijk het punt dat het meest interessant lijkt), uit de oogbewegingen blijkt dan dat men nog verder kijkt.

Doordat veel van de proefpersonen in dit onderzoek Rodden e.a. (2008) niet actief met de muis bewegen, blijkt dat door het analyseren van muisbewegingen niet altijd achterhaald kan worden waar een proefpersoon mee bezig is. Wanneer ook de oogbewegingen mee genomen worden, is het mogelijk om te zien wat er gebeurt als de muis niet gebruikt wordt. De laatste jaren zijn er dan ook een aantal onderzoeken die bekijken wat de toegevoegde waarde van oogbewegingen is.

### 2.3.2 Oogbewegingen als zelfstandige methode

In dit hoofdstuk heb ik het al meerdere keren gehad over oogbewegingen en de manier waarop deze gebruikt zijn in andere onderzoeken. In deze paragraaf beschrijf ik een aantal onderzoeken waarin wordt onderzocht wat de toegevoegde waarde is van oogbewegingen, of in hoeverre deze zelfstandig te gebruiken zijn. Voordat ik doe, bespreek ik eerst kort de manier waarop de gegevens verzameld worden bij onderzoek waarbij men oogbewegingen gebruikt.

#### Eyetracking

Bij oogbewegingsonderzoek gaat men uit van de 'eye-mind' hypothese van Just & Carpenter (1976 – in Cooke 2005). Deze houdt in dat waar men naar kijkt overeenkomt met waar men op dat moment aan denkt. Op deze manier is het dus mogelijk om na te gaan waar een proefpersoon aandacht voor heeft terwijl hij aan het werk is met bijvoorbeeld een website.

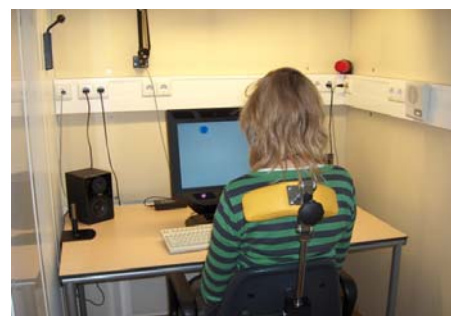
#### *De eyetracker*

De oogbewegingen worden opgenomen door middel van een *eyetracker*. Er zijn twee verschillende soorten eyetrackers, *head-mounted systems* en *remote systems*. Bij de *head-mounted eyetracker* heeft de proefpersoon een soort helm op zijn hoofd waarbij de reflectie van de pupil vanuit verschillende hoeken opgenomen kan worden. Doordat de eyetracker op het hoofd zit en onderscheid kan maken tussen oogbewegingen en hoofdbewegingen, is het mogelijk om het hele hoofd te bewegen. Hierdoor is het mogelijk de oogbewegingen op te nemen terwijl een proefpersoon bijvoorbeeld een auto of vliegtuig bestuurt (Cooke 2005).



*Figuur 2.1: Head-mounted eyetracker.<sup>3</sup>*

De *remote eyetracker* meet de positie van het oog in relatie tot een object dat stil staat, zoals een computerscherm (Cooke 2005). De proefpersoon krijgt hierbij niets op zijn hoofd, maar moet voor het computerscherm gaan zitten. Onderaan het scherm zit dan een camera die de oogbewegingen opvangt. Cooke (2005) noemt als nadeel van dit type systeem dat de proefpersoon zijn hoofd bijna stil moet houden. Maar ze geeft al aan dat de technische ontwikkelingen ervoor zorgen dat er al meer beweging mogelijk is. Lentz e.a. (2006) wijzen zelfs op een Tobii systeem (dat ook voor dit onderzoek gebruikt is) waarbij de proefpersonen "alle vrijheid hebben om te bewegen" en waarbij "proefpersonen nauwelijks meer in de gaten hebben dat hun oogbewegingen geregistreerd worden". Bovendien hebben ze de mogelijkheid om weg te lopen en als ze weer terugkomen meteen verder te werken. Als nadeel van dit systeem noemen ze dat het minder nauwkeurig is dan *head-mounted* systemen.



*Figuur 2.2: Remote eyetracker (gebruikt in dit onderzoek). Het witte puntje is één van de camera's die de oogbewegingen opnemen.<sup>4</sup>*

<sup>3</sup> Bron figuur 2.1: [http://mindwaretech.com/product\\_detail.asp?ItemID=1529](http://mindwaretech.com/product_detail.asp?ItemID=1529)

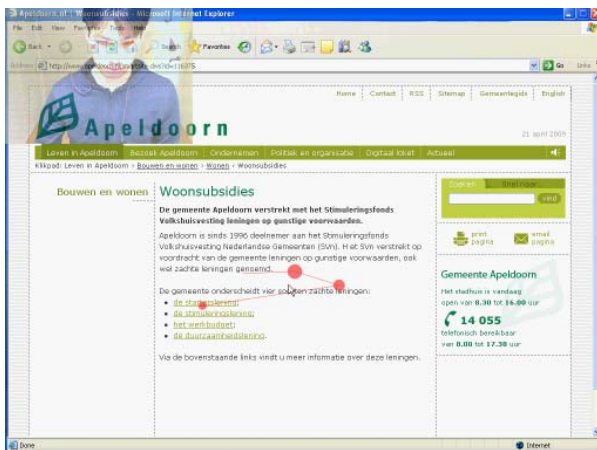
<sup>4</sup> Bron figuur 2.2: <http://www.let.uu.nl/uilots/lab/facilities.php?page=fac-etlab>

## ***Eyetracking gegevens***

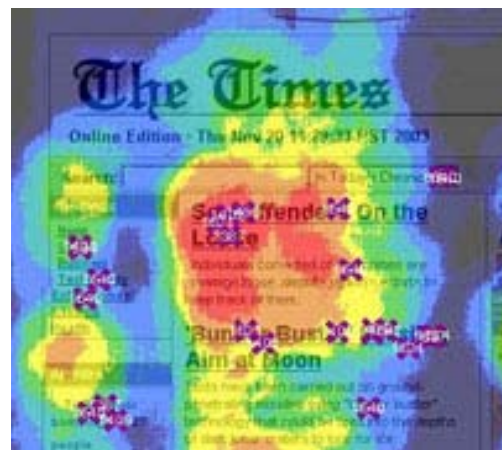
De oogbeweging gegevens kunnen gevisualiseerd worden als fixaties en saccades (Cooke 2005). Hierbij zijn fixaties bepaalde punten waar het oog op gericht is. Deze zijn op het scherm te zien als stippen (zie figuur 2.3). Hoe langer er naar één punt wordt gekeken hoe groter de stip wordt. De sprong van de ene fixatie naar de volgende wordt een saccade genoemd, deze is te zien als een streep tussen twee fixaties (zie figuur 2.3). Tijdens een saccade ziet men niets, het zicht is dan uitgeschakeld (Williams e.a. 2005).

Wanneer al deze fixaties en saccades achter elkaar geplakt worden vormen deze samen een *scanpad*. Dit kan in zijn geheel in een keer in beeld gebracht worden, dat kan echter heel onoverzichtelijk worden. Het is ook mogelijk om de oogbewegingen af te laten spelen, dus dynamisch weer te geven. (Manhartsberger & Zellhofer 2005). Dan is steeds een deel van het scanpad in beeld, zoals in figuur 2.3. Aan een scanpad is bijvoorbeeld te zien wanneer men scant en leest. Scannen is te herkennen aan fixaties die verspreid zijn over de pagina en die niet de tekst volgen. Bij het lezen zijn er meerdere korte fixaties te zien die de tekst volgen (Ehmke & Wilson 2007), dus op dezelfde regel blijven en aan het einde naar het begin van de volgende regel springen.

Naast het scanpad is het mogelijk te visualiseren waar alle proefpersonen samen het meest of het minst naar kijken. Dit wordt dan gerepresenteerd in een *hot spot map* of *heatmap* (Manhartsberger & Zellhofer 2005), bijvoorbeeld zoals in figuur 2.4.<sup>5</sup>



Figuur 2.3: Visualisaties van oogbewegingen, fixaties en saccades.



Figuur 2.4: Visualisaties van oogbewegingen, hot spot map/heatmap.<sup>5</sup>

## **Eyetracking als probleemopsporende methode**

### ***Bepaalde design aspecten onderzoeken***

Manhartsberger & Zellhofer (2005) bespreken een aantal problemen die met de hulp van oogbewegingen opgespoord kunnen worden. Zo kan er vastgesteld worden welke informatie of delen van een website gebruikers daadwerkelijk lezen en wat ze alleen scannen. Een voorbeeld dat ze hierbij noemen is het invullen van een formulier waarboven een toelichtende tekst stond over de lengte van een wachtwoord dat men nodig had. Veel personen kregen een foutmelding, omdat ze die tekst niet lezen. Op basis van dit onderzoek is de toelichting verplaatst naar het deel van het formulier waar dit wachtwoord ingevuld moet worden.

Aan de hand van oogbewegingen kan men ook zien of de naam van links goed is, als veel gebruikers een bepaalde belangrijke link niet aanklikken terwijl ze er wel regelmatig op fixeren is er iets mis met de naam en kan deze beter aangepast worden. Een andere mogelijkheid is vast te stellen waar gebruikers bepaalde zaken verwachten. Door hen opdracht te geven op een website te bepalen welke site het is en daarbij te kijken naar de oogbewegingen konden Manhartsberger & Zellhofer achterhalen dat men hiervoor vooral links bovenaan kijkt.

Ook kunnen problemen met *affordance* opgespoord worden. Affordance betekent dat in één oogopslag duidelijk is hoe iets gebruikt kan worden. Zo moet een navigatie menu er bijvoorbeeld

<sup>5</sup> Bron figuur 2.4: <http://www.poynterextra.org/eyetrack2004/about.htm>

uitzien als een navigatiemenu, anders hebben gebruikers niet door dat ze hiermee de verschillende onderdelen van een site (of in het voorbeeld van Manhartsberger & Zellhofer van een wizard) kunnen bereiken. Tot slot wijzen ze op de mogelijkheid om verkeerde eye catchers te ontdekken. Deze leiden de aandacht van de gebruiker af van onderdelen die belangrijk zijn om een taak uit te voeren op een site. Bijvoorbeeld op een site waar iets aangeschaft kan worden moet de knop om daadwerkelijk te kopen wel duidelijk zijn anders zou dit er toe kunnen leiden dat men uiteindelijk niets koopt.

Bij het vinden van deze problemen richten de onderzoekers zich steeds op één aspect van de website en bekijken hoe de gebruiker daarmee omgaat. Gaat dit goed dan is het in orde, gaat het fout dan is er sprake van een probleem. Ehmke & Wilson (2007) hebben daarentegen geprobeerd bepaalde patronen in oogbewegingen te verbinden aan problemen.

### ***Patronen van oogbewegingen en problemen***

In hun onderzoek hebben Ehmke & Wilson (2007) 19 proefpersonen bepaalde informatie laten zoeken op twee websites, één met nieuws (BBC) en één met reisinformatie (met de trein). Hierbij moesten ze hardopdenken (concurrent of retrospectief met oogbewegingen) en werden hun oogbewegingen opgenomen. Op basis van het commentaar dat de proefpersonen gaven en de observaties van wat ze deden zijn problemen vastgesteld. Deze zijn samengevoegd tot een aantal unieke problemen. Vervolgens zijn eerst algemene oogbewegingspatronen vastgesteld en deze zijn gerelateerd aan de set met problemen. Deze methode was niet echt succesvol dus hebben ze een tweede methode toegepast. Hierbij werd het oogbewegingspatroon beschreven voor elk van de unieke problemen. Vervolgens werd gekeken of de verschillende personen die eenzelfde probleem vonden hetzelfde patroon in oogbewegingen lieten zien. Op basis hiervan hebben ze een aantal patronen gevonden die overeenkomen met problemen.

Ehmke & Wilson (2007) baseren de problemen in hun onderzoek op wat de proefpersonen vertellen. De hardopdenkmethode brengt echter niet alles wat men denkt aan het licht (zie §2.1). Dit blijkt ook uit het onderzoek van Cooke & Cuddihy (2005).

### ***Eyetracking in vergelijking met de problemen uit de hardopdenkmethode***

Cooke & Cuddihy (2005) lieten tien proefpersonen vier verschillende taken uitvoeren op een informatieve website. Tijdens het werken moesten zij hardop nadenken en werden hun oogbewegingen en handelingen op het scherm opgenomen. Uit analyses van de transcripten waarin zowel de verbalisaties als de handelingen en oogbewegingen zijn opgenomen bleek dat de oogbewegingen informatie opleveren die anders verloren zouden gaan. Zo laten oogbewegingen bijvoorbeeld zien als iemand meerdere keren naar een bepaalde link kijkt, terwijl ze dit meestal slechts één keer noemen. Daarnaast lieten oogbewegingen zien dat proefpersonen soms nog even naar andere links keken voordat ze daadwerkelijk op de link klikten waar hun muis al op stond. Ook dit wordt zelden genoemd bij het hardopdenken.

In een aantal scripties van studenten van de Universiteit Utrecht is onderzocht of eyetracking als zelfstandige methode te gebruiken is bij het opsporen van problemen.

Vereijken heeft in 2008 een onderzoek uitgevoerd waarin ze gekeken heeft of de problemen die met eyetracking gevonden worden overeenkomen met de resultaten die via andere methoden gevonden worden. Daarnaast heeft ze onderzocht of bepaalde typen problemen op verschillende manieren tot uiting komen in de oogbewegingen. Tot slot heeft ze zich afgevraagd bij welke typen problemen eyetracking als zelfstandige methode gebruikt kan worden. Om dit te doen heeft ze van 19 (voornamelijk) studenten de oogbewegingen opgenomen terwijl ze aan twee moeilijke en twee makkelijke taken werkten op de website van de gemeente Woerden. Hierbij heeft ze gekeken hoelang de proefpersonen nodig hadden om de taken uit te voeren, hoeveel antwoorden goed en fout waren en hoe snel ze de juiste link vonden en dit vergeleken voor de moeilijke en makkelijke taken. Daarnaast heeft ze gekeken hoelang men naar delen keek die relevante en irrelevante tekst bevatten en deze met elkaar vergeleken. Tot slot heeft ze het kijkgedrag getypeerd als lezen, scannen en kijken naar één woord. Wat het verschil tussen lezen en scannen is, heb ik al kort beschreven bij het stukje over Eyetracking gegevens. Het kijken naar één woord bestaat uit een langere fixatie op één punt.

Op basis van het kijkgedrag heeft Vereijken (2008) gekeken of ze patronen kon ontdekken die op verschillende problemen konden wijzen. Deze heeft ze inderdaad gevonden. Deze worden getoetst in de scripties van twee andere studenten (De Bode 2009, Van der Burg 2009), hier ga ik later verder op in.

Om de problemen die gevonden zijn met eyetracking te vergelijken met andere methodes heeft Vereijken (2008) gebruik gemaakt van eerdere onderzoeken die op dezelfde website zijn uitgevoerd. De andere methodes die meegenomen werden zijn de hardopdenkmethode, vragenlijsten en muisobservaties. Voor de hardopdenkmethode heeft ze zich daarbij vooral gebaseerd op de scriptie van Van Rijswijk (2007). Zij heeft zich echter niet zozeer gericht op het opsporen van problemen door middel van de hardopdenkmethode, maar heeft enkel onderscheid gemaakt tussen positieven en negatieve uitingen. Het resultaat van de analyses van Vereijken (2008) was dat problemen met de bereikbaarheid vaak deels met andere methoden op te sporen zijn, maar dat vooral eyetracking gedetailleerdere informatie oplevert. Ook bij inhoudelijke problemen zijn de eyetracking gegevens gedetailleerder dan de hardopdenkmethode, terwijl via muisobservaties nauwelijks inhoudelijke problemen worden opgespoord. Problemen met het design zijn met alle vier de methoden op te sporen, maar oogbewegingen bevestigen dat er sprake is van een probleem. In haar onderzoek is één probleem gevonden dat alleen door eyetracking op te merken was, namelijk dat een bepaalde alinea over het hoofd werd gezien. Geen enkel probleem werd wel via de andere methoden gevonden, maar niet met eyetracking. Uit haar resultaten concludeert Vereijken (2008) dat oogbewegingen voor het opsporen van bereikbaarheids- en vormgevingsproblemen zelfstandig gebruikt kunnen worden. Met betrekking tot inhoud kan er wel vastgesteld worden dat er problemen zijn, maar is het lastig vast te stellen om welk specifiek probleem het gaat.

Van der Burg (2009) en De Bode (2009) hebben zich in hun scripties ook gericht op de vraag of eyetracking als zelfstandige methode te gebruiken is. Zij hebben ieder 15 proefpersonen op de website van een gemeente (Amersfoort en Haarlem) taken uit laten voeren en na afloop hardop laten denken terwijl ze hun oogbewegingen terugzagen (RTE). Om te kijken welke problemen met eyetracking op te sporen zijn hebben ze hypothesen vastgesteld op basis van de patronen die Vereijken (2008) gevonden heeft, hieronder bespreek ik die kort en in hoeverre zij hier inderdaad bewijs van vonden:

- Als een persoon langdurig op een link fixeert is er sprake van een usability probleem. Dit hebben zowel Van der Burg als De Bode gevonden.
- Als een persoon langdurig op een woord fixeert is er sprake van een usability probleem. Van der Burg en De Bode hebben beide gevonden dat dit soms inderdaad het geval is. Zo kunnen problemen met de begrijpelijkheid van de inhoud opgespoord worden. Hierbij moet er echter wel rekening mee gehouden worden dat de fixaties op de tekst waar het antwoord op de taak in staat langer zijn zonder dat dit betekent dat er sprake is van een probleem. Het gaat dan eerder om het zeker weten van het antwoord of de fixaties zijn langer doordat de proefpersoon aan het voorlezen is.
- Wanneer proefpersonen wel op een goede link fixeren, maar er niet op klikken, is er sprake van een usability probleem. Dit wordt door beiden gevonden en levert eigenlijk alleen problemen op in de dimensie labels (bereikbaarheid).
- Wanneer proefpersonen niet op een cruciale link fixeren is er sprake van een usability probleem. Ook dit wordt door Van der Burg en De Bode gevonden.
- Als proefpersonen herhaaldelijk (minimaal 4x) tijdens één bezoek naar dezelfde link kijken, is er sprake van een usability probleem. Beide vinden dat er door een proefpersoon heel vaak meerdere keren op dezelfde link wordt gefixeerd, maar tussen proefpersonen verschilt het veel welke links dit zijn. Hierom stellen ze dat dit geen nuttige indicatie van een probleem is en dat de grens van vier mogelijk verhoogd moet worden.
- Als proefpersonen wel de doelpagina bezoeken, maar niet het goede antwoord op de taak geven, wijst dit op een usability probleem. Dit wordt door beiden gevonden.
- Als proefpersonen regressies naar een specifiek stukje tekst vertonen, is er sprake van een usability probleem. Regressies zijn een deel van het normaal lezen dus niet altijd een probleem. De Bode geeft hierbij bovendien aan dat het lastig was om te bepalen wanneer iets een regressie was. Dit door het kleine lettertype op de website en doordat de oogbewegingen niet altijd even nauwkeurig waren. Van der Burg heeft alleen regressies waarbij hele regels

teruggekeken worden meegenomen als usability probleem, omdat terugkijken naar woorden onwillekeurig lijkt en het teruglezen veroorzaakt wordt door scrollen.

Uit hun onderzoek concluderen Van der Burg (2009) en De Bode (2009) dat eyetracking in beperkte mate zelfstandig te gebruiken is. De methode lijkt vooral bepaalde soorten problemen te kunnen opsporen. Vooral voor de problemen met het bereiken van de informatie lijkt de methode nuttig. Hierbij dan nog vooral bij problemen met labels en de structuur van de site. Ook problemen met de begrijpelijkheid kan men met eyetracking opsporen. Andere soorten problemen worden door Van der Burg (2009) en De Bode (2009) vooral via de hardopdenkmethode gevonden. Eyetracking levert echter wel meer problemen op dan de hardopdenkmethode.

### 3 Onderzoeksvraag en hypothesen

Vereijken (2008), De Bode (2009) en Van der Burg (2009) hebben alle drie in hun scripties onderzocht in hoeverre eyetracking als zelfstandige methode te gebruiken is. Zij kijken hierbij naar de oogbewegingen en de links waarop de proefpersoon met de muis klikt. Het onderzoek van Mueller & Lockerd (2001) heeft echter aangetoond dat ook het bekijken van andere bewegingen van de muis nuttig kan zijn. Deze wil ik dan ook in een onderzoek expliciet meenemen en alle observaties gezamenlijk gebruiken om tot identificatie van problemen op een website te komen. De hoofdvraag van mijn onderzoek is dan ook:

*In hoeverre zijn schermobservaties (muis- en oogbewegingen) te gebruiken als zelfstandige probleemopsporende evaluatiemethode?*

Om deze vraag te beantwoorden stel ik de volgende drie deelvragen:

1. *Welke problemen kan men aanwijzen door middel van schermobservaties (eyetracking data en het klikpad/muisbewegingen)?*
2. *Welke problemen kan men opsporen door hardopdenken?*
3. *In hoeverre zijn de problemen die opgespoord worden door observaties en de hardopdenkmethode hetzelfde of verschillend?*

Op basis van de onderzoeken van Vereijken (2009), Van der Burg (2009) en De Bode (2009) is de verwachting dat met de observaties meer problemen opgespoord worden dan door de hardopdenkmethode, maar dat dit vooral het geval is voor de problemen die te maken hebben met het bereiken van de informatie en de begrijpelijkheid van de tekst. Bij de hardopdenkmethode zouden de problemen in meer verschillende probleemcategorieën vallen. Dit komt deels doordat de gegevens van de observaties gedetailleerder zijn terwijl de verbalisaties van een abstracter niveau zijn (Cooke & Cuddihy 2005).

In §2.1 zijn verschillende varianten besproken van de hardopdenkmethode. Deze hebben ieder zo hun beperkingen. Zo kan concurrent hardopdenken meer invloed hebben op de taakuitvoering, terwijl retrospectief hardopdenken onderhevig is aan het geheugen. Door middel van het tonen van oogbewegingen wordt geprobeerd het geheugen te steunen, maar hiernaar is nog niet veel onderzoek uitgevoerd (Ball e.a. 2007, Haas 2009). Om deze verschillende invloeden mee te nemen stel ik ook de volgende deelvraag:

4. *Verschilt het per conditie van de hardopdenkmethode welke problemen er gevonden worden en hoe dit in verhouding staat met de geobserveerde problemen?*

In het onderzoek van Ball e.a. (2007) werden in de retrospectieve conditie met oogbewegingen (RTE) meer problemen gevonden dan in de concurrent (CTA) conditie. Voor de retrospectieve conditie zonder oogbewegingen in vergelijking met de RTE en CTA condities werden geen verschillen gevonden. Ook Van den Haak e.a. (2007, 2009) vonden geen verschil tussen het gemiddeld aantal problemen in de CTA en RTA conditie. Wanneer er echter gekeken werd naar het aantal verschillende problemen, zonder het aantal keer dat een probleem voorkwam mee te nemen, werden er in de CTA conditie minder problemen gevonden dan in de RTA conditie. Zij hebben daarnaast ook gekeken naar de verhouding tussen geverbaliseerde en geobserveerde problemen die in de twee condities gevonden werden. In de onderzoeken op een gemeentelijke website (2007, 2009) werden echter geen verschillen gevonden in de manier waarop problemen gedetecteerd werden bij de verschillende hardopdenkmethodes.

Van der Burg (2009) heeft bij de beschrijving van het scangedrag vier typen onderscheiden. Eén daarvan, waarbij de website bijna helemaal gelezen wordt, zag ze vooral bij oudere gebruikers. Ook Ehmke & Wilson (2007) vonden verschillen tussen de proefpersonen die meegewerkt hadden aan hun onderzoek. Ik wil dan ook een aantal verschillen tussen proefpersonen meenemen om te onderzoeken of die van invloed zijn op de problemen die gevonden worden. Daarom stel ik tot slot de volgende deelvraag:

5. *Zijn proefpersoonkenmerken (opleiding, leeftijd, geslacht) van invloed op de soort en hoeveelheid problemen die gevonden worden?*

Hier zijn nog maar weinig eerdere onderzoeken naar uitgevoerd voor de evaluatie van websites. Elling, Lentz & De Jong (niet gepubliceerd) hebben in hun onderzoek naar een nieuwe tool om websites te beoordelen (Infocus) onderzocht of er verschillen waren tussen de proefpersonen. Hierbij verwachtten zij dat jongeren meer ervaring hebben met online tools en daardoor meer op hun gemak zijn met nieuwe tools. Bij hoger opgeleiden verwachten ze dat zij beter feedback kunnen geven omdat zij intensiever bezig zijn geweest met hoe men moet lezen, beoordelen en documenten kan becommentariëren. In hun onderzoek werd er voor de verschillende opleidingsniveaus inderdaad een verschil gevonden tussen het aantal commentaren dat gegeven werd. De hoger opgeleiden gaven meer negatieve commentaren dan de mensen met een midden opleiding en de lager opgeleiden gaven de minste commentaren. Bovendien waren de commentaren van de hoger opgeleiden duidelijker dan van de twee andere groepen. Voor leeftijd werden geen verschillen gevonden.

Ook Haas (2009) heeft in haar scriptie onderzocht of er verschillen waren tussen proefpersonen. Zij heeft alleen naar opleiding gekeken. Hiervoor werden maar weinig verschillen gevonden. De verbalisaties van de hoger opgeleiden waren van hogere kwaliteit dan die van lager opgeleiden. Daarnaast klikten ze minder links aan en kregen ze minder hulp. De taakuitvoering was echter niet sneller of beter en waren er geen verschillen voor het aantal en soort verbalisaties.



## 4 Methode

In dit hoofdstuk bespreek ik de opzet van het onderzoek. In §4.1 komt het materiaal dat de basis vormt voor het onderzoek aan bod. Dit zijn filmpjes van proefpersonen die op een website hebben gewerkt die eerder zijn verzameld. Vervolgens ga ik in op de proefpersonen die hebben meegewerkt aan het onderzoek. In §4.3 bespreek ik hoe die filmpjes geanalyseerd zijn en tot slot in §4.4 wat er met deze gegevens gedaan is.

### 4.1 Materiaal

In dit onderzoek is materiaal gebruikt dat verzameld is voor het promotieonderzoek van drs. S.K. Elling. Zij heeft in de periode van februari tot mei 2009 proefpersonen taken laten uitvoeren op de websites van drie Nederlandse gemeentes, Apeldoorn, Dordrecht en Nijmegen. De proefpersonen kregen steeds eerst de taak te lezen op het scherm. Als ze deze gelezen hadden, werd de website geopend en konden ze de informatie op gaan zoeken. De taak hadden ze ook op papier naast zich liggen zodat ze terug konden kijken. De proefpersonen konden en mochten in principe alles doen wat ze wilden, behalve de zoekmachine gebruiken. Wanneer ze de antwoorden op de taak hadden gegeven werd de volgende taak in beeld gebracht en ging het weer hetzelfde in zijn werk.

Hierbij werden hun oogbewegingen opgenomen met een remote eyetracker (Tobii). Daarnaast zijn ook de muisbewegingen en klikken geregistreerd. Bovendien zijn er met een webcam opnames gemaakt van de proefpersoon. Afhankelijk van de conditie moesten ze tijdens of na de taakuitvoering hardopdenken. Bij de aansporingen is Elling uitgegaan van de klassieke methode. Ze gebruikte zinnen als 'wat dacht u hier?' 'Blijft u hardopdenken' en 'Wat dacht u hier verder?'. Wanneer een proefpersoon de juiste informatie na ongeveer 5 minuten nog niet gevonden had, hielp Elling en stuurde ze hem naar de pagina met de relevante informatie.

In deze scriptie gebruik ik alleen de filmpjes van de website van Apeldoorn. In totaal hebben 60 proefpersonen op de website van de gemeente Apeldoorn gewerkt, 20 in elke conditie (CTA, RTA en RTE). Hierop hebben ze 3 taken uitgevoerd. Ik gebruik hier alleen de eerste taak. Dit omdat ze dan nog fris op de site beginnen en geen voorkennis hebben kunnen opdoen over hoe de site in elkaar zit. De taak die de proefpersonen moesten uitvoeren was:

#### Taak 1

U woont in een huurwoning in Apeldoorn en hebt van uw verhuurder het aanbod gekregen om de woning waarin u woont te kopen voor €160.000,-. U zou graag op dit aanbod ingaan, maar de prijs is eigenlijk te hoog voor u. U hebt wel eens gehoord dat gemeenten financiële steun geven aan mensen zoals u, die voor het eerst een woning kopen. U gaat op de website zoeken of de gemeente Apeldoorn een regeling heeft om nieuwe kopers te steunen en of u voor deze regeling in aanmerking kunt komen.

- 1a. Hoe heet de regeling waarmee de gemeente Apeldoorn financiële steun geeft aan mensen die voor het eerst een woning kopen?
- 1b. Komt u als koper in aanmerking voor deze steun? Waarom wel/niet?
- 1c. Hoe zou u een aanvraag kunnen indienen voor deze steun?

#### 4.1.1 Website gemeente Apeldoorn ([www.apeldoorn.nl](http://www.apeldoorn.nl))

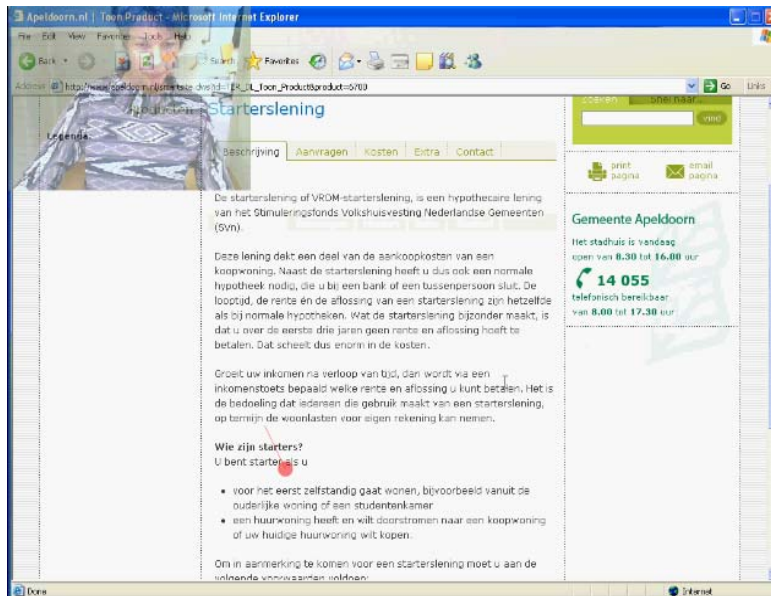
Op de website van de gemeente Apeldoorn kan de informatie om de taak uit te voeren op een aantal manieren bereikt worden. Daarnaast zijn er verschillende versies van de website. Deze bespreek ik in deze paragraaf.

#### **Routes**

Men kan via het onderwerp *Leven in Apeldoorn* door het menu heen naar de pagina *Woonsubsidies* gaan (zie figuur 4.1), daar moet men dan klikken op de link *Starterslening*. Dan komt men uit op de pagina die te zien is in figuur 4.2. Daar kan men het antwoord op vraag 1a en 1c vinden. Om vraag 1b te beantwoorden moeten de proefpersoon op de link klikken, dan opent er een pdf bestand waarin alle vier de soorten leningen die de gemeente aanbiedt beschreven worden (figuur 4.3).

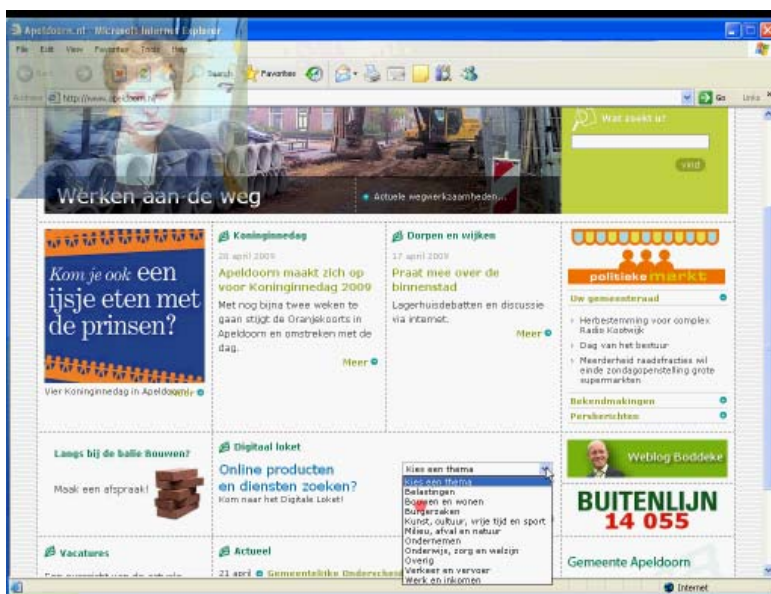


dan moet men in het submenu een manier van zoeken kiezen, op thema of op alfabet. Dan opent er een pagina waar ze verschillende thema's zien of de letters van het alfabet. Bij de thema's moet men dan voor *Bouwen en wonen* kiezen, dan krijgt men een lijst met producten en daar staat de *Starterslening* bij. Bij het zoeken op alfabet moeten de proefpersonen op de 'S' klikken en daar staat dan de *Starterslening*. Deze pagina *Starterslening* is echter een andere pagina dan men via *Leven in Apeldoorn* verkrijgt (zie figuur 4.4). Hier kan men het antwoord op vraag 1a en 1b vinden. Voor vraag 1c moeten de proefpersonen op de tab *Aanvragen* klikken. Hier staat precies hetzelfde stukje tekst als onderaan op figuur 4.2.



Figuur 4.4: Pagina *Starterslening* via route *Digitaal loket*.

Ook op de startpagina zelf staat een mogelijkheid om in het *Digitaal loket* te zoeken. In figuur 4.5 is deze te zien. Hier kan men op thema en/of product kiezen en zo ook de pagina *Starterslening* vinden.



Figuur 4.5: Zoeken in *Digitaal loket* op de startpagina.

Van de 60 proefpersonen heeft uiteindelijk 78,3% de informatie gezocht op de pagina *Starterslening* die bereikt wordt via *Leven in Apeldoorn* (figuur 4.1-4.3). De overige 21,6% van de proefpersonen heeft de antwoorden op de vragen uit de taak gegeven op basis van de pagina via het *Digitaal loket* (figuur 4.4). Deels is deze verhouding te verklaren doordat Elling de proefpersonen die hulp nodig hebben naar de pagina *Starterslening* via *Leven in Apeldoorn* leidt.

### **Versies**

Tijdens het onderzoek zijn ook een aantal verschillende versies van de website online geweest, het verschil zat hem hierbij in de nieuwsberichten die op de site stonden. Twee hiervan leken heel erg de aandacht te trekken. Dit waren de berichten 'Vragen aan de gemeente' en 'Nieuwe wegwijzer zorg, wonen en welzijn'. Na analyse bleek echter dat er slechts één significant verschil was tussen de verschillende versies, bij de versie met het bericht 'Nieuwe wegwijzer zorg, wonen en welzijn' deden zich meer problemen voor met labels dan in alle andere versies ( $F(22,6, 4,2) = 5,39, p = .007$ ). Omdat dit het enige verschil was heb ik bij de verdere analyses geen onderscheid gemaakt tussen de versies.

## **4.2 Proefpersonen**

De personen die aan dit onderzoek hebben meegewerkt variëren in leeftijd, geslacht en opleiding. Dit omdat de websites van gemeenten voor een zeer grote groep mensen toegankelijk moeten zijn. Het is belangrijk dat zij allemaal de informatie kunnen vinden die zij nodig hebben. Bovendien is het op deze manier mogelijk om te onderzoeken of er verschillen bestaan tussen mensen in hun manier van werken op een website en de manier waarop zij verbaliseren.

Om deze gemengde groep te verkrijgen heeft Elling gebruik gemaakt van een wervingsbureau. Aan hen heeft ze gevraagd om personen met ervaring met het internet die zonder bril goed kunnen lezen van een scherm. Daarnaast moest er spreiding zijn op leeftijdscategorie, geslacht en opleidingsniveau. Voor de website van Apeldoorn staat de verdeling van de proefpersonen in tabel 4.1.

**Tabel 4.1: Proefpersonen website Apeldoorn**

	<b>CTA</b>	<b>RTA</b>	<b>RTE</b>	<b>Totaal</b>
<b>Totaal</b>	20	20	20	60
<b>Geslacht</b>				
Man	10	11	11	32
Vrouw	10	9	9	28
<b>Leeftijd</b>				
18-29	6	5	5	16
30-39	4	3	5	12
40-54	5	6	4	15
55+	5	6	6	17
<b>Opleiding*</b>				
Laag	6	6	6	18
Midden	7	7	7	21
Hoog	7	7	7	21

\* Laag is basisschool, lbo, vmbo, mavo, mulo. Midden is havo, vwo, mbo. Hoog is hbo, wo.

## **4.3 Analyse filmpjes**

Het doel van mijn analyses van de filmpjes was het achterhalen van de problemen die de proefpersonen tegenkwamen. Hierbij heb ik onderscheid gemaakt tussen de problemen die te vinden zijn door de proefpersonen te observeren en de problemen die aan het licht komen door wat ze zeggen.

### **4.3.1 Observaties**

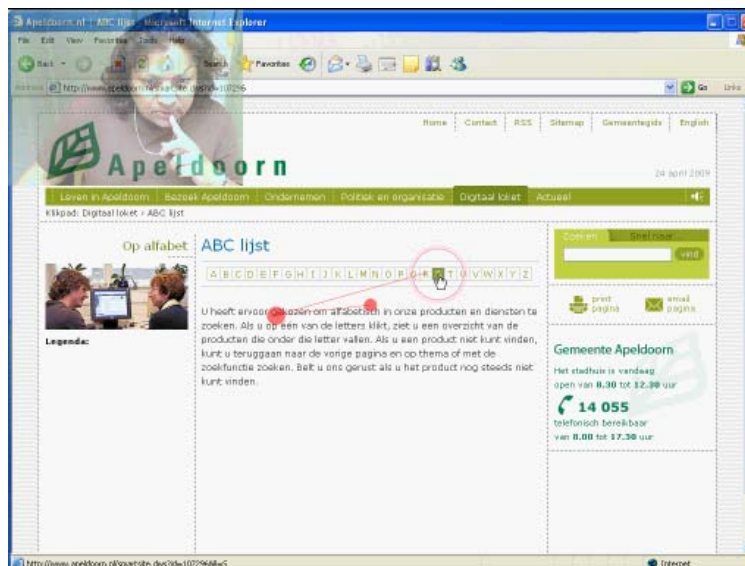
De observaties bestaan uit alle observaties van het scherm, de antwoorden op de vragen en wat er verder om de proefpersoon heen gebeurt.

#### **Scherf**

De observaties van het scherm zijn zowel het klikpad en de muisbewegingen als de oogbewegingen, oftewel alles wat op het scherm te zien is. Dat een persoon klikt, is in de opnames te zien als een cirkel met een dubbele rand. Wanneer er met links geklikt wordt is deze cirkel rood, wanneer men met rechts klikt groen. De pagina die opent na zo'n klik komt in beeld zoals dat op een normale website gebeurt. De muis ziet men gewoon zoals deze normaal op het scherm te zien is. Wanneer deze bewogen wordt is het een pijltje. Als deze ergens op komt te staan verandert het pijltje in een



handje. Zo kan ik zien dat een bepaalde link met de muis geselecteerd wordt. Figuur 4.6 laat een voorbeeld zien van hoe het eruit ziet als men ergens op klikt.



Figuur 4.6: Voorbeeld van schermobservaties (pagina Digitaal loket > Op alfabet)

De oogbewegingen zijn te zien als rode stippen op het scherm met ertussen strepen. De strepen zijn saccades en zijn momenten waarop men niets ziet. De stippen zijn de fixaties. Deze worden groter naarmate men langer naar één punt kijkt. Er is in de literatuur nog geen duidelijkheid over wanneer men nu precies iets ziet. Daarom heb ik in principe elke fixatie op een element geïnterpreteerd als zien. Alleen wanneer de fixatie zich voordoet op het moment dat men naar beneden scrollt heb ik dit niet gedaan. Dan gaat dat punt namelijk zo snel voorbij dat het niet waarschijnlijk is dat het daadwerkelijk gezien wordt.

Bij zien heb ik onderscheid gemaakt tussen lezen en scannen. Lezen is te zien doordat de proefpersoon verschillende fixaties achter elkaar maakt op ongeveer dezelfde hoogte. Bij scannen zijn de fixaties en saccades langer dan bij lezen. Daarnaast zijn de bewegingen grilliger, men gaat meer onregelmatig over de tekst of over de pagina (Ehmke & Wilson 2007).

In een aantal usability onderzoeken worden in de observaties het aantal muisklikken, fixaties en saccades en de duur van deze fixaties en saccades geanalyseerd (onder andere Vereijken 2008, Hertzum e.a. 2009). In dit onderzoek wil ik echter de oogbewegingen met de muisbewegingen combineren. Daarvoor is het nodig om te kijken naar het geheel van wat er gebeurt en niet naar de oogbewegingen of muisbewegingen apart. De analyses zijn dus meer kwalitatief dan kwantitatief van aard.

### **Antwoorden**

De antwoorden op de vragen uit de taak heb ik ook meegenomen als observatie, omdat deze aangeven of de taak goed uitgevoerd is. Hierbij heb ik me wel beperkt tot de uitkomst van wat ze zeiden, of het uiteindelijke antwoord goed of fout is. Dus bijvoorbeeld, bij de eerste vraag *Hoe heet de regeling waarmee de gemeente Apeldoorn financiële steun geeft aan mensen die voor het eerst een woning kopen?*, heb ik gekeken of ze '(VROM)starterslening' antwoorden of niet. Bij vraag 1b (voorwaarden om in aanmerking te komen) en vraag 1c (aanvragen) heb ik de antwoorden opgedeeld in een aantal aspecten en gekeken welke hiervan naar voren kwamen bij het antwoord. Bij vraag 1c zijn dit bijvoorbeeld dat men (1) een aanmeldformulier moet hebben, (2) dit kan verkrijgen bij de makelaar of tussenpersoon en (3) moet opsturen naar het Stimuleringsfonds Volkshuisvesting. Voor elk van deze punten heb ik dan aangegeven of dit in het antwoord zat. Het commentaar dat daaromheen gegeven wordt, zoals bijvoorbeeld dat ze het niet zeker weten heb ik niet meegenomen.

### ***Overig***

Met wat er omheen gebeurt bedoel ik bijvoorbeeld dat ze aan het praten zijn, waardoor ze zich waarschijnlijk minder goed kunnen concentreren op wat ze aan het doen zijn. Of bijvoorbeeld dat er tegen hen gepraat wordt. Een aantal proefpersonen draaien zich deels om naar de onderzoeksleidster als ze iets tegen haar zeggen of er tegen hen gepraat wordt, dan is er op het scherm dus niets te zien. Anderen blijven echter naar het scherm kijken. Dan kan het lijken dat ze heel lang een pagina scannen, maar achteraf blijkt dan dat ze in gesprek zijn en dus nog niet echt aan het zoeken zijn.

Ook de hulp die de onderzoeksleider geeft, heb ik opgenomen als een observatie. Alles wat er gebeurt tijdens het geven van die hulp, dus wat er aangeklikt wordt, wanneer en hoelang men er dan over doet om op de goede pagina te komen, heb ik niet meegenomen. Hier wordt de proefpersoon gestuurd en is het niet zijn eigen manier van werken die geobserveerd wordt.

### **4.3.2 Verbalisaties**

De verbalisaties bestaan uit alles wat de proefpersonen zeggen. Voor de proefpersonen in de *concurrent thinking aloud* (CTA) conditie is dit wat ze vertellen tijdens het uitvoeren van de taak. Bij de retrospectieve condities gaat het dan om wat ze achteraf vertellen. Wat ze gezegd hebben tijdens de taakuitvoering heb ik hierbij in principe niet meegenomen, hoe duidelijk ze op dat moment hun problemen dan ook verwoordden. Mijn doel is te achterhalen wat ze achteraf nog verbaliseren, dus de gedachten tijdens de taakuitvoering zijn bij hen niet van belang.

Hierbij stuitte ik wel op een probleem. In de CTA conditie vertellen de proefpersonen hun antwoord tijdens de verbalisaties, maar bij het retrospectief hardopdenken hoeven ze de antwoorden niet meer te geven. Een onjuist antwoord is echter wel een indicatie van een probleem. Daarom heb ik in de retrospectieve condities, naast alles wat bij het terugzien van hun taakuitvoering gezegd werd, ook de (onjuiste) antwoorden meegenomen.

### **4.3.3 Procedure**

Het doel van de analyses is het identificeren van problemen die de proefpersonen tegenkomen bij het werken op de website. In deze paragraaf beschrijf ik hoe ik hierbij te werk ben gegaan.

#### **Problemen identificeren (observaties)**

##### ***Taakmodel***

Om de filmpjes te analyseren heb ik van tevoren een taakmodel opgesteld. Bij het samenstellen ervan heb ik de indicaties van usability problemen meegenomen die Van der Burg (2009) en De Bode (2009) getoetst hebben. Zij vonden bijvoorbeeld dat er sprake is van een probleem wanneer er niet gefixeerd wordt op een link of wel gefixeerd maar niet geklikt. Daarom heb ik voor elke pagina bepaald wat ze moeten zien en waar ze op moeten klikken. Ook het bereiken van de doelpagina, maar een onjuist antwoord geven is een indicatie van een probleem. Het is dan mogelijk dat ze de juiste informatie niet gelezen hebben of niet goed begrepen. Daarom heb ik ook steeds aangegeven wat ze precies moeten lezen om de vragen te kunnen beantwoorden.

Zoals beschreven in §4.1.1 zijn er op de website van Apeldoorn een aantal manieren om bij de informatie over de Starterslening te komen. In het taakmodel zijn al deze routes volledig uitgewerkt. In Bijlage I staat het volledige taakmodel, op de volgende pagina staat hieruit een fragment waarin beschreven staat wat men moet doen op de pagina *Leven in Apeldoorn > ... > Starterslening*.

**Tabel 4.2: Fragment uit taakmodel**

<b>7. Leven in Apeldoorn - ... - Starterslening</b>
Zien + lezen één van de zinnen over waarvoor de Starterslening is:
<i>De starterslening dekt een deel van de aankoopkosten van een koopwoning voor starters.</i>
Zien
Lezen
<i>Deze lening dekt een deel van de aankoopkosten van een koopwoning.</i>
Zien
Lezen
Zien: link naar het pdf document
Lezen: link naar het pdf document
Muis op: link naar het pdf document
Aanklikken: link naar het pdf document
Zien + lezen informatie voor vraag 1c:
<i>Aanmeldformulier</i>
Zien
Lezen
<i>Vul dan het speciale aanvraagformulier - dat bij uw makelaar of tussenpersoon verkrijgbaar is - in en stuur deze op naar: Stimuleringsfonds Volkshuisvesting ...</i>
Zien
Lezen
Beantwoorden vraag 1a:
(VROM)Starterslening
Beantwoorden vraag 1c:
Via een aanvraagformulier
dit moet men ophalen bij makelaar of tussenpersoon
en opsturen naar Stimuleringsfonds Volkshuisvesting.

### **Identificeren**

Het taakmodel heb ik gebruikt om vast te stellen wanneer een proefpersoon een probleem heeft. Hierbij heb ik een aantal indicaties van problemen vastgesteld, deze zijn:

1. Er wordt afgeweken van het taakmodel.
  - a. Een stap uit het model wordt niet genomen
  - b. Er wordt iets anders gedaan voordat de volgende stap uit het model genomen wordt.
2. Er wordt lang gedaan over het nemen van de volgende stap. Hierbij neem ik als ondergrens 5 seconden. In het explorerende onderzoek bleek deze grens realistisch te zijn.
3. Het antwoord op een vraag is onjuist of onvolledig.

Vervolgens heb ik gekeken of er daadwerkelijk sprake zou zijn van een probleem. Bij het onjuist of onvolledig beantwoorden van een vraag is dit altijd het geval. Wanneer een stap niet genomen wordt of pas laat is dit niet noodzakelijk.

Zo wordt er ondanks het uitgebreide taak model door een enkele proefpersoon een andere route genomen waarlangs de informatie toch bereikt kan worden. Eén proefpersoon gebruikt bijvoorbeeld de sitemap. Ze klikt daar echter bijna meteen de link *Woonsubsidies* aan. Ze heeft dus geen probleem met het vinden van de informatie.

Soms is het juist dankzij het uitgebreide taakmodel dat er geen sprake is van een probleem. Er zijn gevallen waarbij de proefpersoon in het menu het onderdeel *Leven in Apeldoorn* wel ziet, maar toch naar de informatie gaat zoeken via *Digitaal loket*. Dat hij niet gaat zoeken in het submenu van *Leven in Apeldoorn* is dan geen probleem, hij gaat simpelweg via een andere route verder.

Ook als het even duurt voor de volgende stap genomen wordt, hoeft er niet noodzakelijk sprake te zijn van een probleem. Een voorbeeld hiervan is dat een proefpersoon bijvoorbeeld de hele pagina bekijkt voordat hij daadwerkelijk iets gaat doen. Dit is in principe geen probleem, hij is zich aan het

oriënteren. Pas wanneer hij meerdere keren achter elkaar de hele pagina overkijkt zal dit erop wijzen dat hij niets kan vinden. Dus dan is er wel sprake van een probleem.

Een ander voorbeeld is dat een proefpersoon in het menu bijvoorbeeld al wel de link *Woonsubsidies* heeft gezien en zijn muis er al op heeft gezet, maar eerst nog de rest van het menu gaat bekijken. Wanneer hij daarna weer terug komt bij *Woonsubsidies* zie ik dit niet als probleem, hij is gewoon aan het nakijken of er nog andere relevante links in het menu staan. Wanneer hij niet terug komt op de pagina *Woonsubsidies* of eerst op een andere link klikt, is er wel sprake van een probleem. Op deze manier heb ik voor elke probleemindicatie bepaald of er daadwerkelijk sprake was van een probleem door de handelingen met de muis en de oogbewegingen goed te bekijken.

Om te voorkomen dat mijn observaties teveel gestuurd werden door de verbalisaties heb ik de CTA filmpjes eerst geanalyseerd zonder geluid. Op basis daarvan heb ik dan eerst de problemen geïdentificeerd die ik kon zien. Vervolgens heb ik het filmpje bekeken met het geluid erbij. Op die manier kon ik vaststellen wat de antwoorden op de vragen waren en wanneer ze aan het praten waren. In de retrospectieve condities hoefden de proefpersonen niets te zeggen tenzij ze antwoord gaven. Daarom heb ik die filmpjes wel meteen met geluid bekeken. Af en toe bleken zij echter ook commentaar te geven tijdens de taakuitvoering. Deze verbalisaties heb ik dan zoveel mogelijk genegeerd.

Wat de proefpersonen doen op pagina's die buiten het taakmodel vallen, heb ik niet meegenomen zodat de analyses niet te breed worden. Dus bijvoorbeeld begripsproblemen op een niet relevante pagina komen in dit onderzoek niet aan de orde. De enige uitzondering hierop is wanneer men vanaf een andere pagina met het menu ging werken. Het menu is zo belangrijk bij het vinden van de juiste informatie dat ik in het taakmodel een apart onderdeel 'Navigatie-1 Andere pagina' heb opgenomen.

Op de bijgevoegde Cd-rom staan de bestanden met alle analyses, deze zijn niet uitgeprint omdat deze meerdere pagina's per proefpersoon omvatten. In bijlage II staat ter illustratie de uitgeprinte versie van de analyse van één van de proefpersonen.

### **Problemen identificeren (verbalisaties)**

Nadat ik de problemen had vastgesteld die uit de observaties te halen waren, ben ik aan de slag gegaan met de verbalisaties. Van een groot deel van de filmpjes waren al transcripten beschikbaar, de overigen heb ik zelf uitgetypt. Hierbij ging het me vooral om wat ze zeiden. Het was niet van belang hoe lang ze tussendoor stil waren of wat hun intonatie was. Het ging om de informatie waaruit ik af kon leiden welke problemen ze tegenkwamen. Om dit te kunnen bepalen heb ik in het bestand waar ook de analyses van de observaties stonden per pagina erachter gezet wat er gezegd is. Dus als een proefpersoon meerdere keren teruggekomen is op de startpagina, dan staat alles wat ze daar zeiden achter het deel met observaties bij de startpagina. Dit is te zien in de voorbeeldanalyse in bijlage II.

Vervolgens heb ik problemen geïdentificeerd op basis van wat ze zeiden. Hierbij heb ik niet alleen problemen die ze zelf noemden meegenomen, zoals 'ik kan het niet vinden' of 'dit snap ik niet', maar ook problemen die uit hun woorden af te leiden zijn. Eén van de proefpersonen zegt bijvoorbeeld: "De kosten [gemompel] €55,-, er zijn wel kosten aan verbonden"<sup>6</sup>. Uit deze verbalisatie blijkt dat ze denkt dat er kosten aan de starterslening verbonden zijn. Dit staat echter niet in de tekst over de Starterslening, maar een stuk verder bij de Duurzaamheidslening. Hieruit leid ik dan af dat ze niet goed begrepen heeft dat dit niet hoort bij de informatie over de Starterslening en dat er dus sprake is van een probleem. Een ander voorbeeld van problemen die ik heb afgeleid zijn de onjuiste antwoorden. De proefpersonen hebben zelf niet door dat ze een fout of onvolledig antwoord geven, maar het is wel een belangrijk probleem.

Voor elke proefpersoon heb ik zo een aantal verbalisaties geselecteerd die op een probleem wijzen. Hierbij heb ik in veel gevallen echter wel meerdere verbalisaties bij elkaar genomen. Veel proefpersonen herhalen bijvoorbeeld meerdere keren dat ze de informatie niet kunnen vinden. Dit

---

<sup>6</sup> Proefpersoon 2 tijdens bekijken pdf



blijft het zelfde probleem, ze zijn aan het zoeken maar kunnen niets vinden. In deze gevallen heb ik het als één probleem geïnterpreteerd.

De verbalisaties waaruit blijkt dat ze niet snappen wat ze moeten doen, dus problemen hebben met de opdracht zelf heb ik niet meegenomen, omdat het me alleen gaat om de problemen met de website.

## Probleembeschrijvingen

In het bestand met analyses heb ik steeds een probleembeschrijving opgenomen. Hornbæk (2010) stelt dat het identificeren van identieke problemen geholpen kan worden door een goede, uitgebreide probleembeschrijving. Waaraan zo'n probleembeschrijving dan precies moet voldoen wordt echter niet aangegeven.

In dit onderzoek neem ik verschillende punten op in de beschrijving. Ten eerste de probleemindicatie, waardoor heb ik het probleem opgemerkt, en wat houdt het probleem precies in. Vervolgens heb ik de oorzaak van het probleem proberen te achterhalen. Problemen moeten opgelost worden, het is daarom van belang de oorzaak te weten. Dit is de reden dat ik de problemen hierop in categorieën verdeeld heb. Deze categorieën worden hieronder toegelicht. Tot slot heb ik gekeken welke andere problemen hierdoor ontstaan. Als een proefpersoon bijvoorbeeld een bepaalde tekst niet heeft gezien en gelezen kan hij ook de vragen niet goed beantwoorden. Het onjuist beantwoorden van de vraag heb ik dan niet genoemd als apart probleem.

## Probleemcategorieën

Voor de probleemcategorieën heb ik me gebaseerd op de categorieën in de Web Evaluation Questionnaire (WEQ) die Elling, Lentz en De Jong (2007) ontwikkeld hebben. De hoofdcategorieën zijn Navigatie, Inhoud en Design. Deze bestaan elk uit een aantal subcategorieën. Deze staan in tabel 4.3 genoemd met een korte beschrijving erachter van wat ik eronder heb verstaan.

**Tabel 4.3: Categorieën van oorzaken van problemen met definitie**

<b>Inhoud</b>	Alles wat met de inhoud op de pagina's te maken heeft.
Begrijpelijkheid	De informatie is niet duidelijk genoeg.
Volledigheid	De informatie is niet volledig, de proefpersoon mist informatie.
Relevantie	De informatie wordt gezien als niet nuttig, of als niet relevant. Wanneer men informatie duidelijk wel gelezen heeft, maar bijvoorbeeld slechts één van de voorwaarden bij vraag 1b noemt heb ik dit geïnterpreteerd als: ze zien de voorwaarden niet als nuttig voor het antwoord.
Overig	Alles wat wel met de inhoud van de pagina's te maken heeft, maar niet in één van de andere subcategorieën past.
<b>Navigatie</b>	Alles wat te maken heeft met het bereiken van de informatie
Menulabel/hyperlink	Het label is niet duidelijk genoeg, want men klikt op de verkeerde link of niet op de relevante link.
Gebruikersgemak	De proefpersoon kan niet makkelijk met de website werken. Hij snapt bijvoorbeeld niet hoe hij iets moet doen
Structuur	Het is de proefpersoon niet duidelijk waar hij de informatie kan vinden en/of hij weet niet waar hij zich bevindt.
Snelheid	De website doet niet snel genoeg wat de proefpersoon wil.
Overig	Alles wat wel met de navigatie op de site te maken heeft, maar niet in één van de andere subcategorieën past.
<b>Design</b>	Alles wat met het uiterlijk van de site te maken heeft.
Ontwerp	Problemen met het ontwerp van de site, kleurgebruik, afbeeldingen, achtergronden, frames en dergelijke.
Lay-out	Problemen met de opmaak van een pagina. Dus de lay-out van tekst, kopjes en dergelijke.
Overig	Alles wat wel met het uiterlijk van de site te maken heeft, maar niet in de andere subcategorieën past.
<b>Overig</b>	Problemen die niet in één van de andere categorieën onder te brengen zijn en problemen die wel gesignaleerd worden, maar die niet verklaard kunnen worden doordat bijvoorbeeld oogbewegingen wegvallen.

In de WEO (Elling e.a. 2007) worden de (sub)categorieën Overig niet gebruikt. Ik heb deze echter wel toegevoegd om de problemen ook wanneer ik niet echt weet wat de oorzaak is in te kunnen delen. Ook mijn categorie Design komt niet helemaal overeen met die in de WEO. In de WEO is deze niet uitgesplitst. Dat heb ik wel gedaan, omdat de opmaak van de tekst er in mijn ogen vaak de oorzaak van lijkt te zijn dat informatie niet gevonden kan worden. Dit heeft echter niets te maken met het algemene uiterlijk van de website.

Bij de problemen die ik afgeleid heb uit onjuiste of onvolledige antwoorden was het vooral bij de verbalisaties moeilijk om te achterhalen wat de oorzaak was. Hier werd meestal niets over gezegd, omdat de proefpersoon zelf niet wist dat het antwoord niet juist was. In deze gevallen heb ik de oorzaak overgenomen die ik uit de observaties gevonden had. Dit om te voorkomen dat de problemen bij de verbalisaties bij de categorie Overig ingedeeld worden terwijl het bij de observaties wel mogelijk is om een specifieke categorie toe te wijzen.

## 4.4 Vergelijkende analyses

Na het vaststellen van de problemen die elke proefpersoon tegenkwam tijdens het uitvoeren van de taak uit zowel de observaties als verbalisaties ben ik deze problemen met elkaar gaan vergelijken. Dit om vast te stellen welke problemen hetzelfde waren en welke verschillend.

### 4.4.1 Criteria identieke problemen

Om identieke problemen vast te stellen heb ik een aantal criteria gebruikt. Deze zijn:

1. De problemen doen zich op dezelfde plaats op de website voor.
2. De problemen behoren tot dezelfde categorie en hebben ook echt dezelfde oorzaak.
3. De problemen hebben dezelfde 'voorgeschiedenis'.

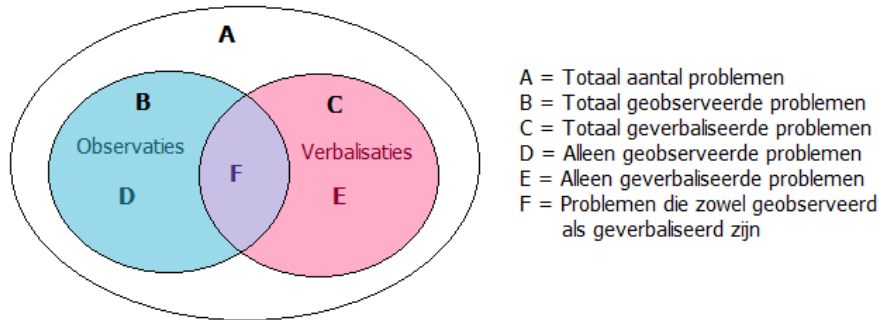
Voor het eerste criterium geldt over het algemeen dat het om dezelfde pagina moet gaan. Het menu staat echter op elke pagina en is steeds hetzelfde. Daarom heb ik dit als apart onderdeel van de site gezien. Voor problemen met het menu maakt het dus niet uit op welke pagina dat menu wordt gebruikt. Wel heb ik daarbij rekening gehouden met het tweede criterium. Zo is het mogelijk dat bijvoorbeeld de link *Woonsubsidies* niet aangeklikt wordt. Wanneer de proefpersoon deze link wel gezien heeft is er sprake van een andere oorzaak dan wanneer hij die link niet gezien heeft. Dit zijn dus geen identieke problemen.

Het derde criterium heeft te maken met het moment waarop het probleem zich voordoet. Zo zijn er bijvoorbeeld een paar personen die al op de pagina *Starterslening* geweest zijn. Daarna gaan ze verder zoeken in het menu en klikken daar op de link *Woningmarktinformatie*. Nu is er blijkbaar iets mis met de informatie op de pagina *Starterslening*. Wanneer ze echter op een eerder moment op de link *Woningmarktinformatie* klikken, ligt het waarschijnlijk vooral aan het label ervan. Dat lijkt hen dan de plaats waar ze de relevante informatie kunnen vinden.

### 4.4.2 Alle probleemdetecties

In de eerste plaats heb ik voor elke proefpersoon gekeken welke van de geobserveerde problemen ook uit de verbalisaties naar voren kwamen en welke uniek waren voor elk van de methodes. Hierbij was het redelijk simpel om te bepalen of het om hetzelfde probleem ging in de verbalisaties en observaties. Dit bleek meestal wel uit het moment waarop een probleem genoemd werd of uit de opmerking zelf.

In een databestand heb ik per proefpersoon aangegeven hoeveel problemen men in totaal tegengekomen is, hoeveel hiervan geverbaliseerd en geobserveerd zijn, hoeveel overlap er is en hoeveel problemen alleen geverbaliseerd of geobserveerd zijn. Voor een visuele weergave hiervan zie figuur 4.7.



Figuur 4.7: Visualisatie van data

In dit bestand staan dus de problemen per proefpersoon. De overlappende problemen zijn dus ook voor elke proefpersoon apart opgenomen. Dit betekent dat de problemen die bij meerdere personen gevonden zijn ook meerdere keren geteld zijn. In het vervolg noem ik deze set de probleemset met het totaal aantal probleemdetecties.

#### 4.4.3 Alle unieke problemen

Na het bepalen van de unieke problemen per proefpersoon heb ik gekeken naar de overlap tussen problemen in het algemeen. Hiervoor heb ik alle probleembeschrijvingen in een apart bestand gezet en de in §4.4.1 beschreven criteria toegepast. Dit leidde er over het algemeen toe dat het vooral bij geobserveerde problemen redelijk eenvoudig was om vast te stellen welke problemen hetzelfde waren. Voor de verbalisaties die nog niet verbonden waren met een observatie was dit lastiger, omdat geïnterpreteerd moest worden wat er werd gezegd. Vaak bleek dit toch net iets anders te zijn dan een andere verbalisatie en heb ik het als een uniek probleem in de analyses meegenomen.

Bij elk probleem uit de set met unieke problemen heb ik bekeken hoe vaak het voorkomt, door hoeveel personen het genoemd wordt, hoe vaak het alleen uit de observaties afgeleid kan worden, hoe vaak het alleen uit verbalisaties afgeleid kan worden en hoe vaak het zowel via observaties als verbalisaties gevonden is. De visualisatie uit figuur 4.7 is ook bij dit databestand van toepassing, nu is echter elk probleem slechts één keer opgenomen. Wanneer meerdere proefpersonen hetzelfde probleem tegen zijn gekomen, is dit dus als één probleem geteld.

Voor de overlappende problemen (F) moet bovendien opgemerkt worden dat deze alle problemen bevat die gevonden zijn via verbalisaties en observaties. Niet alleen de problemen die door verschillende proefpersonen via zowel de observaties als de verbalisaties gevonden zijn, zijn hierin opgenomen. Ook wanneer een probleem bij de ene proefpersoon alleen geobserveerd wordt en bij een ander alleen uit de verbalisaties naar voren komt heb ik het probleem bij deze groep ingedeeld, omdat het probleem met beide methoden gevonden kan worden.

#### 4.4.4 Unieke problemen bij meerdere personen

Uit het databestand met de unieke problemen heb ik nog een databestand afgeleid. Namelijk een bestand dat alleen de unieke problemen bevat die bij meerdere proefpersonen gevonden zijn. Dit heb ik gedaan omdat De Bode (2009) en Van der Burg (2009) ervan uit gaan dat er pas echt sprake is van een probleem als minsten twee personen het gevonden hebben.

#### 4.4.5 Analyses

Het doel van de analyses is onderzoeken of er meer problemen via observatie of verbalisatie gevonden zijn. Dit heb ik gedaan door voor elk van de databestanden te bepalen of er een verschil bestaat tussen het aantal problemen dat alleen via observatie en het aantal problemen dat alleen via verbalisatie opgespoord is. Terugkijkend naar figuur 4.7 betekent dit dat ik steeds de problemen in groep D en E met elkaar vergelijk. Deze groepen laten namelijk zien of er problemen zijn die enkel met één van beide methoden gevonden kunnen worden.

Bij de interpretatie van de resultaten ga ik ervan uit dat er daadwerkelijk een verschil is tussen de observaties en verbalisaties wanneer in alle drie de probleemsets (§4.4.2-§4.4.4) dezelfde trend te zien is en deze minstens één keer significant is. Dus in alle drie de probleemsets moet er bijvoorbeeld sprake van zijn dat er gemiddeld of procentueel meer problemen geobserveerd dan geverbaliseerd zijn. Is in één van de probleemsets echter het aantal geverbaliseerde problemen hoger dan is er geen eenduidig beeld en is er dus niet duidelijk meer opgespoord via observatie dan verbalisatie.

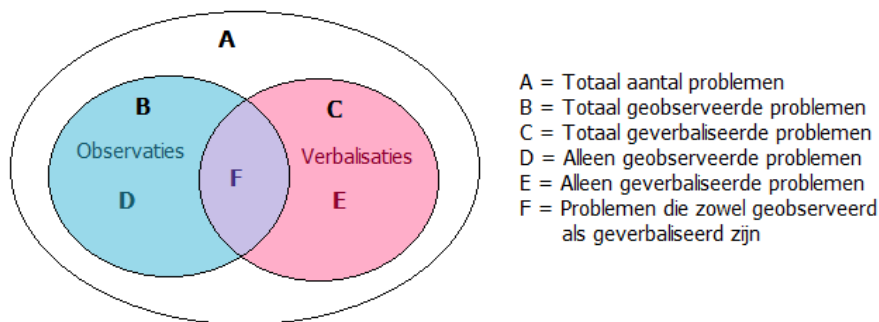
## 5 Resultaten

In dit hoofdstuk bespreek ik de resultaten van de analyses. §5.1 behandelt de algemene gegevens. Vervolgens wordt ingegaan op de verschillende deelvragen. In §5.2 bespreek ik de resultaten met betrekking tot de problemen die opgespoord zijn met schermobservaties en de verbalisaties en in hoeverre deze verschillen of overeenkomen. Of er verschillen zijn tussen de drie varianten van de hardopdenkmethode onderzoek ik in §5.3 en in §5.4 bekijk ik de verschillen tussen groepen proefpersonen.

### 5.1 Algemene gegevens

Gemiddeld deden de proefpersonen er ruim 6 minuten (374,9 seconden, std 117,3) over om de taak uit te voeren. Het varieerde echter van een minimum van 2:22 minuten tot een maximum van 10:24 minuten.

In totaal zijn er bij alle proefpersonen samen 429 problemen gevonden. Dit bevat alle problemen uit de groepen D, E en F uit figuur 5.1. Gemiddeld zijn dit 7,15 (std 3,0) problemen per proefpersoon. Er zijn echter problemen die bij meerdere proefpersonen gevonden zijn. Daarom zijn de criteria uit §4.1.1 toegepast. Hierna zijn er 196 unieke problemen overgebleven. Van den Burg (2009) en De Bode (2009) hebben in hun onderzoeken alleen problemen meegenomen die bij meer dan één persoon gevonden zijn. Wanneer dit ook bij de data in dit onderzoek gedaan wordt blijven er nog 68 unieke problemen over.



Figuur 5.1: Visualisatie van data

### 5.2 Problemen via observaties en/of verbalisaties

Op basis van de analyses in deze paragraaf wordt antwoord gegeven op de vragen welke problemen gevonden worden aan de hand van observaties en verbalisaties en in hoeverre dit overeenkomt of verschillend is.

Zoals vermeld zijn er in totaal 429 problemen gevonden voor de verbalisaties en observaties samen voor alle proefpersonen. In tabel 5.1 staat het gemiddeld aantal problemen per categorie uitgesplitst voor de problemen die alleen via observatie (D in figuur 5.1), alleen via verbalisatie (E) en die via beide methoden (F) gevonden zijn. In tabel 5.2 en tabel 5.3 is hetzelfde gedaan voor de 196 unieke problemen en de 68 unieke problemen die bij meerdere personen gevonden zijn. In tabel 5.2 zijn dus de problemen die bij meerdere personen gevonden zijn slechts één keer geteld en in tabel 5.3 zijn alleen de problemen meegenomen die bij meer dan één persoon gevonden zijn.

Uit deze tabellen blijkt dat in alle probleemcategorieën door middel van zowel observaties als verbalisaties problemen gevonden zijn.

#### 5.2.1 Observatie versus verbalisatie

In alle drie de tabellen is het aandeel van problemen die door beide methoden gevonden worden het grootst. Wanneer gekeken wordt naar het totaal aantal problemen (tabel 5.1) is het 45,2%, voor de

unieke problemen (tabel 5.2) 48,5% en voor de problemen die bij meerdere personen gevonden worden zelfs 73,5%. Daarnaast zijn er echter ook problemen die enkel via observaties of verbalisaties opgespoord zijn. Het verschil hiertussen is alleen bij de unieke problemen (tabel 5.2) significant ( $t(195)=-4,25$ ,  $p<.001$ ).

### ***Inhoud***

Bij de problemen met de inhoud van de website is het aandeel problemen dat opgespoord is via verbalisatie het grootst wanneer gekeken wordt naar het totaal aantal problemen (tabel 5.1) en de unieke problemen (tabel 5.2). Bij de unieke problemen die door meer dan één persoon gevonden worden maakt het voor het grootste deel van de problemen niet uit welke methode gebruikt wordt.

Ongeacht de probleemset waarnaar gekeken wordt, zijn er in de categorie Inhoud meer problemen gevonden via verbalisatie dan door observatie. Bij de subcategorie Begrijpelijkheid is dit ook het geval. Ook de categorie Volledigheid laat dezelfde trend zien, maar hier is het verschil alleen significant wanneer gekeken wordt naar het totaal aantal problemen en de unieke problemen, niet bij de unieke problemen die bij meerdere personen gevonden worden.

De subcategorie Relevantie is een apart geval, deze laat alleen in de probleemset waarbij gekeken wordt naar het totaal aantal problemen (tabel 5.1) een significant verschil zien ( $t(59)=2,40$ ,  $p=.02$ ). Hierbij worden echter meer problemen geobserveerd dan geverbaliseerd, dit in tegenstelling tot het algemene beeld bij Inhoud.

### ***Navigatie***

Bij de categorie Navigatie worden de meeste problemen opgespoord door middel van beide methoden (53,0%, 56,1% en 80,6%). Gekeken naar alle problemen met het bereiken van de informatie zijn er geen significante verschillen tussen het aantal problemen dat alleen via observatie of verbalisatie gevonden wordt. Bij een aantal subcategorieën bestaat hiertussen wel een verschil.

In de subcategorie Gebruikersgemak zijn meer problemen gevonden via observaties dan verbalisaties, in de set met alle probleemdetecties is dit verschil daadwerkelijk significant ( $t(59)=2,01$ ,  $p=.05$ ). Ook in de categorie Snelheid is het aandeel van geobserveerde problemen groter. Het verschil met de verbalisaties is zelfs in alle drie de probleemsets significant (Totaal:  $t(59)=4,28$ ,  $p<.001$ . Uniek:  $t(11)=3,02$ ,  $p=.01$ . Meerdere personen:  $t(4)=4,00$ ,  $p=.02$ ).

De problemen met Labels worden daarentegen vaker opgespoord via de verbalisaties dan via de observaties. Bij de set met unieke problemen (tabel 5.2) is dit verschil significant ( $t(61)=2,68$ ,  $p=.01$ ).

### ***Design***

Voor de probleemcategorie Design is het niet te zeggen op welke manier het grootste deel van de problemen opgespoord wordt. Wanneer gekeken wordt naar het totaal aantal problemen (tabel 5.1) wordt het grootste deel alleen geobserveerd (36%). Bij de unieke problemen (tabel 5.2) wordt echter 57,9% geverbaliseerd en de unieke problemen die bij meerdere proefpersonen gevonden zijn, zijn voornamelijk via beide methodes gevonden (58,8%).

Ook bij het vergelijken van de het aantal problemen dat geobserveerd of geverbaliseerd wordt komt er geen eenduidig beeld naar voren. Enkel bij de unieke problemen (tabel 5.2) zijn er significant meer problemen opgespoord via verbalisatie dan observatie ( $t(37)=-3,42$ ,  $p=.002$ ). In de twee andere probleemsets zijn er het gemiddeld aantal en percentage problemen in via observatie juist hoger.

Ook wanneer er naar de subcategorieën bij Design gekeken wordt (Ontwerp en Lay-out) is er alleen bij de unieke problemen (tabel 5.2) een significant verschil. Er zijn meer unieke problemen met de lay-out opgespoord door middel van de verbalisaties dan via de observaties ( $t(25)=-4,24$ ,  $p<.001$ ). In dit geval laten de andere probleemsets echter wel dezelfde trend zien.

Tabel 5.1: Gemiddeld aantal problemen per proefpersoon geobserveerd en/of geverbaliseerd, in totaal en per type probleem. Met resultaat gepaarde t-toets voor observatie versus verbalisatie.

	Totaal (N=60)	Observatie (N=60)	Beide (N=60)	Verbalisatie (N=60)	T(59)
<b>Totaal</b>	<b>7,15 (3,0)</b>	<b>1,9 (1,4)</b>	<b>3,2 (1,4)</b>	<b>2,0 (1,7)</b>	-0,20, p=.85
<b>Inhoud</b>	<b>2,0 (1,2)</b>	<b>0,4 (0,5)</b>	<b>0,8 (0,7)</b>	<b>0,9 (0,9)</b>	-3,87, p<.001*
Begrijpelijkheid	0,6 (0,8)	0,1 (0,4)	0,1 (0,3)	0,4 (0,6)	-3,18, p=.002*
Relevantie	0,8 (0,7)	0,2 (0,4)	0,6 (0,6)	0,1 (0,2)	2,40, p=.02*
Volledigheid	0,6 (0,6)	0,1 (0,3)	0,1 (0,3)	0,4 (0,6)	-3,96, p<.001*
<b>Navigatie</b>	<b>3,1 (2,2)</b>	<b>0,9 (1,1)</b>	<b>1,6 (1,4)</b>	<b>0,6 (0,9)</b>	1,93, p=.06
Gebruikersgemak	0,4 (0,6)	0,2 (0,5)	0,1 (0,4)	0,1 (0,3)	2,01, p=.05*
Structuur	0,6 (0,7)	0,1 (0,2)	0,4 (0,5)	0,2 (0,4)	-1,63, p=.11
Label	1,6 (1,3)	0,2 (0,5)	1,1 (1,0)	0,3 (0,6)	-0,96, p=.34
Snelheid	0,5 (0,7)	0,4 (0,6)	0,1 (0,2)	0,0 (0,2)	4,28, p<.001*
Overig	0,1 (0,2)	0,0 (0,1)	0,0 (0,1)	0,0 (0,1)	0, p=1
<b>Design</b>	<b>1,9 (1,4)</b>	<b>0,7 (0,7)</b>	<b>0,7 (0,8)</b>	<b>0,6 (0,9)</b>	0,94, p=.35
Ontwerp	0,4 (0,6)	0,3 (0,5)	0,0 (0,2)	0,1 (0,4)	1,48, p=.15
Lay-out	1,5 (1,3)	0,4 (0,6)	0,6 (0,8)	0,4 (0,8)	0, p=1
<b>Overig</b>	<b>0,2 (0,6)</b>	<b>0,0 (0,1)</b>	<b>0,1 (0,5)</b>	<b>0,0 (0,1)</b>	0, p=1

Tabel 5.2: Aantal en percentage unieke problemen geobserveerd en/of geverbaliseerd, in totaal en per type probleem. Met resultaat van gepaarde t-toets voor observatie versus verbalisatie<sup>7</sup>.

	Totaal N	Observatie N %	Beide N %	Verbalisatie N %	T
<b>Totaal</b>	<b>196</b>	<b>30 15,3</b>	<b>95 48,5</b>	<b>71 36,2</b>	t(195)=-4,25, p<.001*
<b>Inhoud</b>	<b>43</b>	<b>2 4,7</b>	<b>19 44,2</b>	<b>22 51,2</b>	t(42)=-5,16, p<.001*
Begrijpelijkheid	22	1 4,5	8 36,4	13 59,1	t(21)=-4,29, p<.001*
Relevantie	10	1 10,0	6 60,0	3 30,0	t(9)=-1,00, p=.34
Volledigheid	11	0 0,0	5 45,5	6 54,5	t(10)=-3,46, p=.006*
<b>Navigatie</b>	<b>107</b>	<b>21 19,6</b>	<b>60 56,1</b>	<b>26 24,3</b>	t(106)=-0,73, p=.47
Gebruikersgemak	14	6 42,9	5 35,7	3 21,4	t(13)=1,00, p=.34
Structuur	16	1 6,3	11 68,8	4 25,0	t(15)=-1,38, p=.19
Label	62	5 8,1	40 64,5	17 27,4	t(61)=-2,68, p=.01*
Snelheid	12	8 66,7	3 25,0	1 8,3	t(11)=3,02, p=.01*
Overig	3	1 33,3	1 33,3	1 33,3	t(2)=0, p=1
<b>Design</b>	<b>38</b>	<b>6 15,8</b>	<b>10 26,3</b>	<b>22 57,9</b>	t(37)=-3,42, p=.002*
Ontwerp	12	4 33,3	2 16,7	6 50,0	t(11)=-0,62, p=.55
Lay-out	26	2 7,7	8 30,8	16 61,5	t(25)=-4,24, p<.001*
<b>Overig</b>	<b>8</b>	<b>1 12,5</b>	<b>6 75,0</b>	<b>1 12,5</b>	t(7)=0, p=1

Tabel 5.3: Unieke problemen die bij meer dan 1 persoon gevonden zijn per categorie met hoeveelheid geobserveerd en/of geverbaliseerd<sup>6</sup>.

	Totaal N	Observatie N %	Beide N %	Verbalisatie N %	T
<b>Totaal</b>	<b>68</b>	<b>9 13,2</b>	<b>50 73,5</b>	<b>9 13,2</b>	t(67)=0, p=1
<b>Inhoud</b>	<b>18</b>	<b>0 0,0</b>	<b>13 72,2</b>	<b>5 27,8</b>	t(17)=-2,56, p=.02*
Begrijpelijkheid	9	0 0,0	5 55,6	4 44,4	t(8)=-2,53, p=.04*
Relevantie	4	0 0,0	4 100,0	0 0,0	nvt
Volledigheid	5	0 0,0	4 80,0	1 20,0	t(40)=-1, p=.37
<b>Navigatie</b>	<b>31</b>	<b>5 16,1</b>	<b>25 80,6</b>	<b>1 3,2</b>	t(30)=1,68, p=.10
Gebruikersgemak	4	0 0,0	4 100,0	0 0,0	nvt
Structuur	7	0 0,0	7 100,0	0 0,0	nvt
Label	15	1 6,7	13 86,7	1 6,7	t(14)=0, p=1
Snelheid	5	4 80,0	1 20,0	0 0,0	t(4)=4,00, p=.02*
Overig	0	0 0,0	0 0,0	0 0,0	nvt
<b>Design</b>	<b>17</b>	<b>4 23,5</b>	<b>10 58,8</b>	<b>3 17,6</b>	t(16)=0,37, p=.72
Ontwerp	7	4 57,1	2 28,6	1 14,3	t(6)=1,44, p=.20
Lay-out	10	0 0,0	8 80,0	2 20,0	t(9)=-1,5, p=.17
<b>Overig</b>	<b>2</b>	<b>0 0,0</b>	<b>2 100,0</b>	<b>0 0,0</b>	nvt

<sup>7</sup> Gemiddeldes zijn gelijk aan de percentages.

## 5.2.2 Conclusie

Uit de voorgaande analyses blijkt dat er zowel via observatie als verbalisatie en een combinatie van beide in alle categorieën problemen gevonden zijn. Via beide methoden kunnen dus verschillende soorten problemen opgespoord worden.

Over het algemeen is er geen verschil tussen het aantal problemen dat alleen via observatie of verbalisatie gevonden kan worden. In de categorie Inhoud zijn er echter meer problemen gevonden die alleen door middel van verbalisaties opgespoord zijn dan via observaties. Dit is niet alleen het geval wanneer naar de gehele hoofdcategorie gekeken wordt, maar ook wanneer er naar de problemen in de subcategorieën Begrijpelijkheid en Volledigheid gekeken wordt. Bij de subcategorie relevantie is er geen eenduidig beeld.

De hoofdcategorie Navigatie laat geen verschil zien tussen het aantal problemen dat opgespoord is via observaties en verbalisaties. Dit komt mogelijk door de tegengestelde resultaten bij de subcategorieën. In de categorieën Gebruikersgemak en Snelheid is een groter deel van de problemen via observaties gevonden, terwijl dit bij de categorie Label via verbalisaties is.

De categorie Design laat geen eenduidig beeld zien voor het aantal problemen dat gevonden is door middel van verbalisatie of observatie. Ditzelfde geldt voor de bijbehorende subcategorie Ontwerp. De problemen in de subcategorie Lay-out zijn echter wel vaker via verbalisatie gevonden dan observatie.

Hieruit kan dus geconcludeerd worden dat er inderdaad verschillen zijn tussen de problemen die geobserveerd en/of geverbaliseerd worden.

## 5.3 Verschillen tussen CTA, RTA en RTE

In deze paragraaf bespreek ik de analyses waarin de concurrent (CTA), retrospectief zonder oogbewegingen (RTA) en retrospectief met oogbewegingen (RTE) conditie vergeleken worden om de vraag te kunnen beantwoorden of er verschillen bestaan tussen condities.

### *Tijd*

Een eerste punt waarnaar gekeken is, is de tijd die men nodig had om de taken uit te voeren. Hiervoor is een significant verschil gevonden ( $F(2,57)=5,55$ ,  $p=.006$ ), waarbij de proefpersonen in de CTA conditie (437,6 seconden, std 113,3) er langer over deden dan de personen in de RTE conditie (324,6 seconden, std 112,8). Met de RTA conditie (362,4 seconden, std 101,0) werden geen verschillen gevonden.

### *Aantal problemen*

Tabel 5.4 laat het gemiddeld aantal problemen per proefpersoon in totaal en per type probleem zien. In totaal zijn door alle proefpersonen samen in de CTA conditie 137 problemen gevonden, 149 in de RTA conditie en 143 in de RTE conditie. Dit is respectievelijk 31,9%, 34,7% en 33,3% van de 429 problemen die in totaal gevonden zijn. De verschillen tussen de condities zijn echter niet significant ( $F(2,57)=0,19$ ,  $p=.83$ ).

In tabel 5.5 staat het aantal unieke problemen dat in elk van de condities gevonden is. Hierbij zijn de problemen van alle proefpersonen met elkaar vergeleken en de problemen die bij meerdere personen gevonden zijn slechts één keer meegeteld. In de CTA, RTA en RTE condities zijn respectievelijk 88 (44,9%), 98 (50%) en 88 (44,9%) van de 196 unieke problemen gevonden.

De derde probleemset bestaat uit de unieke problemen die meerdere personen tegengekomen zijn. Tabel 5.6 laat zien dat in de CTA conditie 44 (64,7%), in de RTA conditie 53 (77,9%) en in de RTE conditie 49 (72,1%) van de 68 problemen gevonden zijn.

### *Overlap tussen problemen*

Van de 88 unieke problemen in de CTA conditie komen er 32 overeen met de 98 uit de RTA conditie, dit is 16,3% van 196 unieke problemen. De overlap tussen de CTA en de RTE conditie is 15,8% van



het totaal aantal unieke problemen. Bij de RTA en de RTE conditie is de overlap iets hoger, namelijk 18,4%. De problemen die in alle drie de condities voorkomen beslaan slechts 10,7% van de 196 unieke problemen.

De overlappende problemen in de probleemset met alle unieke problemen komen allemaal ook terug in de set met unieke problemen die bij meerdere personen gevonden zijn.

Net als in de voorgaande paragraaf zijn de gemiddeldes en percentages uitgesplitst voor de problemen die alleen via observatie (D in figuur 5.1), alleen via verbalisatie (E) en die via beide methoden (F) gevonden zijn. Hieronder bespreek ik of er verschillen zijn tussen het aantal geverbaliseerde en geobserveerde problemen en op basis hiervan geef ik in §5.3.2 antwoord op de deelvraag.

### 5.3.1 Observatie versus verbalisatie

Wanneer de problemen uitgesplitst zijn per conditie is steeds het grootste aandeel van de problemen via beide methodes gevonden. Bij het totaal aantal probleemdetecties (tabel 5.4) is dit voor CTA 46,7%, voor RTA 47,7% en voor RTE 41,3%. Wanneer gekeken wordt naar de unieke problemen (tabel 5.5) is dit respectievelijk 47,7%, 46,9% en 40,9%. Bij de unieke problemen die bij meerdere personen gevonden zijn (tabel 5.6) zijn deze percentages 56,8% bij CTA, 56,6% bij RTA en 53,1% bij RTE.

Er zijn echter ook problemen die enkel uit de observaties of de verbalisaties naar voren zijn gekomen. Binnen de condities is bij het vergelijken van het totaal aantal probleemdetecties (tabel 5.4) en de unieke problemen (tabel 5.5) zijn er in de CTA en RTE conditie meer problemen gevonden via verbalisaties dan via observaties. Dit verschil is echter alleen significant bij de probleemset met unieke problemen (CTA:  $t(87)=-2,56$ ,  $p=.01$ ), RTE:  $t(87)=-2,19$ ,  $p=.03$ ) en wordt tegengesproken in de probleemset met unieke problemen die bij meerdere personen gevonden zijn (tabel 5.6). Voor de RTA conditie is er geen enkel significant verschil.

Bij een vergelijking van het aantal geobserveerde problemen in de drie condities blijkt dat wanneer naar alle probleemdetecties gekeken wordt (tabel 5.4) in geen van de condities significant meer of minder geobserveerd wordt dan in een andere conditie. Bij het vergelijken van de unieke problemen (al dan niet bij meerdere personen gevonden) kunnen alleen de problemen meegenomen worden die voorkomen in beide condities die getoetst worden. Dus bij een vergelijking tussen het aantal problemen die voortgekomen zijn uit de observaties in de CTA en RTA conditie kan dit slecht berekend worden voor de 32 problemen die deze condities gemeen hebben. In geen van de sets met unieke problemen leidt dit echter tot significante verschillen tussen de condities.

#### ***Inhoud***

Bij de problemen met de inhoud van de website wordt steeds een deel van de problemen via beide methoden gevonden. In de CTA conditie bevat deze 'methode' bij alle drie de probleemsets steeds het grootste deel van de problemen (alle 59,5%, uniek 41,7% en uniek meer personen 56,8%). In de RTE conditie is het grootste gedeelte van de problemen via de verbalisaties gevonden. De RTA conditie laat geen duidelijk beeld zien, bij het totaal aantal probleemdetecties zijn gemiddeld de meeste problemen via alleen verbalisaties gevonden, terwijl de meeste unieke problemen via beide methoden gevonden zijn.

Een vergelijking tussen het aantal geverbaliseerde en geobserveerde problemen laat zien dat in alle drie de condities meer problemen gevonden zijn via de verbalisaties dan de observaties. In de meeste gevallen is dit verschil ook significant (zie tabel 5.4-5.6).

Tabel 5.4: Gemiddeld aantal problemen per proefpersoon in totaal en per type probleem per conditie. Met het resultaat van de gepaarde t-toets binnen elke conditie voor observatie versus verbalisatie en resultaat van de vergelijking van het totaal aantal problemen, de geobserveerde en de geverbaliseerde problemen per conditie.

	CTA					RTA					RTE					F	F	F
	Tot	Obser	Beide	Verb	T(19)	Tot	Obser	Beide	Verb	T(19)	Tot	Obser	Beide	Verb	T(19)	(2,57) Tot	(2,57) Obser	(2,57) Verb
<b>Totaal</b>	<b>6,9</b> (3,5)	<b>1,7</b> (1,4)	<b>3,2</b> (1,6)	<b>2,0</b> (1,9)	-0,58, p=.57	<b>7,5</b> (2,6)	<b>2,2</b> (1,7)	<b>3,6</b> (1,3)	<b>1,8</b> (1,0)	1,04, p=.31	<b>7,2</b> (3,1)	<b>2,0</b> (1,3)	<b>3,0</b> (1,2)	<b>2,3</b> (2,2)	-0,59, p=.56	0,19, p=.83	0,48, p=.62	0,41, p=.66
<b>Inhoud</b>	<b>1,9</b> (1,1)	<b>0,1</b> (0,3)	<b>1,1</b> (0,6)	<b>0,7</b> (0,7)	-2,98, p=.01*	<b>1,9</b> (1,1)	<b>0,4</b> (0,5)	<b>0,7</b> (0,7)	<b>0,9</b> (0,8)	-2,13, p=.05*	<b>2,3</b> (1,4)	<b>0,6</b> (0,7)	<b>0,6</b> (0,6)	<b>1,1</b> (1,0)	-1,76, p=.10	0,66, p=.52	<b>4,70,</b> p=.01*	1,08, p=.35
Begrijpelijkheid	0,5 (0,7)	0,1 (0,2)	0,2 (0,4)	0,3 (0,5)	-2,03, p=.06	0,6 (0,8)	0,1 (0,2)	0,2 (0,4)	0,4 (0,7)	-2,10, p=.05*	0,7 (1,0)	0,2 (0,5)	0,1 (0,2)	0,5 (0,7)	-1,42, p=.17	0,30, p=.74	1,20, p=.31	0,30, p=.74
Relevantie	0,9 (0,7)	0,1 (0,2)	0,8 (0,7)	0	1,00, p=.33	0,7 (0,7)	0,2 (0,4)	0,5 (0,6)	0,1 (0,2)	1,38, p=.19	1,0 (0,7)	0,3 (0,5)	0,6 (0,6)	0,1 (0,3)	1,71, p=.10	0,70, p=.50	2,16, p=.12	1,04, p=.36
Volledigheid	0,5 (0,6)	0	0,2 (0,4)	0,4 (0,5)	-3,20, p=.01	0,6 (0,6)	0,1 (0,3)	0,1 (0,3)	0,4 (0,5)	-2,04, p=.06	0,6 (0,8)	0,1 (0,3)	0	0,5 (0,8)	-2,03, p=.06	0,15, p=.86	1,05, p=.36	0,33, p=.72
<b>Navigatie</b>	<b>2,8</b> (2,6)	<b>0,8</b> (1,0)	<b>1,5</b> (1,4)	<b>0,5</b> (1,1)	1,19, p=.25	<b>3,6</b> (2,2)	<b>1,1</b> (1,3)	<b>1,9</b> (1,4)	<b>0,6</b> (0,8)	1,93, p=.07	<b>2,9</b> (1,7)	<b>0,8</b> (0,9)	<b>1,5</b> (1,3)	<b>0,7</b> (0,9)	0,32, p=.76	0,70, p=.50	0,63, p=.54	0,14, p=.87
Gebruikersgemak	0,4 (0,6)	0,2 (0,5)	0,2 (0,4)	0,1 (0,2)	0,81, p=.43	0,4 (0,6)	0,3 (0,5)	0,1 (0,2)	0,1 (0,2)	2,03, p=.06	0,5 (0,8)	0,2 (0,5)	0,2 (0,5)	0,1 (0,3)	0,70, p=.49	0,27, p=.76	0,48, p=.62	0,26, p=.77
Structuur	0,6 (0,7)	0,1 (0,2)	0,4 (0,5)	0,1 (0,3)	-0,57, p=.58	0,7 (0,8)	0,1 (0,4)	0,6 (0,6)	0,1 (0,4)	-0,45, p=.67	0,5 (0,8)	0,1 (0,2)	0,2 (0,4)	0,3 (0,4)	-1,71, p=.10	0,38, p=.68	0, p=1	0,91, p=.41
Label	1,4 (1,4)	0,3 (0,6)	0,9 (0,9)	0,3 (0,6)	0, p=1	2,0 (1,3)	0,3 (0,6)	1,3 (1,1)	0,4 (0,6)	-0,70, p=.50	1,4 (1,1)	0,1 (0,3)	1,1 (1,1)	0,3 (0,6)	-1,00, p=.33	1,34, p=.27	0,90, p=.41	0,47, p=.63
Snelheid	0,6 (0,6)	0,4 (0,5)	0,1 (0,3)	0,1 (0,3)	2,03, p=.06	0,5 (0,8)	0,5 (0,8)	0	0	2,44, p=.03*	0,4 (0,5)	0,4 (0,5)	0,1 (0,2)	0	3,20, p=.01*	0,27, p=.77	0,17, p=.84	2,11, p=.13
Overig	0	0	0	0	nvt	0,1 (0,2)	0	0,1 (0,2)	0	nvt	0,1 (0,3)	0,1 (0,2)	0	0,1 (0,2)	0, p=1	1,03, p=.36	1,00, p=.37	1,00, p=.37
<b>Design</b>	<b>2,1</b> (1,2)	<b>0,8</b> (0,7)	<b>0,5</b> (0,7)	<b>0,8</b> (1,1)	-0,17, p=.87	<b>1,7</b> (1,5)	<b>0,7</b> (0,7)	<b>0,7</b> (0,8)	<b>0,4</b> (0,9)	1,51, p=.15	<b>2,0</b> (1,5)	<b>0,6</b> (0,8)	<b>0,9</b> (0,8)	<b>0,5</b> (0,8)	0,49, p=.63	0,33, p=.72	0,22, p=.81	1,23, p=.30
Ontwerp	0,4 (0,7)	0,2 (0,5)	0,1 (0,2)	0,2 (0,5)	0, p=1	0,5 (0,7)	0,3 (0,6)	0,1 (0,2)	0,2 (0,5)	0,83, p=.42	0,4 (0,5)	0,3 (0,5)	0	0,1 (0,2)	2,03, p=.06	0,38, p=.68	0,57, p=.57	0,38, p=.69
Lay-out	1,7 (1,3)	0,6 (0,6)	0,5 (0,6)	0,7 (1,0)	-0,21, p=.83	1,2 (1,4)	0,4 (0,6)	0,6 (0,8)	0,2 (0,6)	1,29, p=.21	1,6 (1,3)	0,3 (0,6)	0,9 (0,8)	0,5 (0,8)	-0,77, p=.45	0,81, p=.45	1,34, p=.27	1,50, p=.23
<b>Overig</b>	<b>0,2</b> (0,7)	<b>0,1</b> (0,2)	<b>0,1</b> (0,4)	0	1,00, p=.33	<b>0,3</b> (0,7)	0	<b>0,3</b> (0,7)	0	nvt	<b>0,1</b> (0,2)	0	0	<b>0,1</b> (0,2)	-1,00, p=.33	1,02, p=.37	1,00, p=.37	1,00, p=.37

\* betekent  $p < .05$ . De grijs gekleurde hokken geven aan waartussen het significante verschil zich voordoet binnen de conditie. De blauw gekleurde hokken geven aan waartussen het significante verschil zich voordoet tussen de condities.

Tabel 5.5: Unieke problemen per categorie met hoeveelheid geobserveerd en/of geverbaliseerd. Gesplitst voor de drie methoden van hardopdenken; concurrent (CTA), retrospectief zonder oogbewegingen (RTA) en retrospectief met oogbewegingen (RTE).

	CTA								RTA								RTE							
	Tot		Observ		Beide		Verb		Tot		Observ		Beide		Verb		Tot		Observ		Beide		Verb	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
<b>Totaal</b>	<b>88</b>		<b>15</b>	<b>17,0</b>	<b>42</b>	<b>47,7</b>	<b>31</b>	<b>35,2</b>	<b>98</b>		<b>26</b>	<b>26,5</b>	<b>46</b>	<b>46,9</b>	<b>26</b>	<b>26,5</b>	<b>88</b>		<b>20</b>	<b>22,7</b>	<b>36</b>	<b>40,9</b>	<b>32</b>	<b>36,4</b>
<b>Inhoud</b>	<b>24</b>		<b>2</b>	<b>8,3</b>	<b>10</b>	<b>41,7</b>	<b>12</b>	<b>50,0</b>	<b>22</b>		<b>2</b>	<b>9,1</b>	<b>11</b>	<b>50,0</b>	<b>9</b>	<b>40,9</b>	<b>20</b>		<b>6</b>	<b>30,0</b>	<b>2</b>	<b>10,0</b>	<b>12</b>	<b>60,0</b>
Begrijpelijkheid	10		1	10,0	3	30,0	6	60,0	10		0	0,0	5	50,0	5	50,0	11		3	27,3	1	9,1	7	63,6
Relevantie	6		1	16,7	5	83,3	0	0,0	6		1	16,7	4	66,7	1	16,7	5		3	60,0	0	0,0	2	40,0
Volledigheid	8		0	0,0	2	25,0	6	75,0	6		1	16,7	2	33,3	3	50,0	4		0	0,0	1	25,0	3	75,0
<b>Navigatie</b>	<b>41</b>		<b>8</b>	<b>19,5</b>	<b>25</b>	<b>61,0</b>	<b>8</b>	<b>19,5</b>	<b>51</b>		<b>16</b>	<b>31,4</b>	<b>25</b>	<b>49,0</b>	<b>10</b>	<b>19,6</b>	<b>48</b>		<b>8</b>	<b>16,7</b>	<b>28</b>	<b>58,3</b>	<b>12</b>	<b>25,0</b>
Gebruikersgemak	5		2	40,0	2	40,0	1	20,0	7		5	71,4	1	14,3	1	14,3	7		1	14,3	4	57,1	2	28,6
Structuur	6		0	0,0	5	83,3	1	16,7	10		1	10,0	8	80,0	1	10,0	8		0	0,0	4	50,0	4	50,0
Label	24		3	12,5	16	66,7	5	20,8	26		3	11,5	15	57,7	8	30,8	25		1	4,0	19	76,0	5	20,0
Snelheid	6		3	50,0	2	33,3	1	16,7	7		7	100,0	0	0,0	0	0,0	6		5	83,3	1	16,7	0	0,0
Overig	0		0	0,0	0	0,0	0	0,0	1		0	0,0	1	100,0	0	0,0	2		1	50,0	0	0,0	1	50,0
<b>Design</b>	<b>20</b>		<b>4</b>	<b>20,0</b>	<b>5</b>	<b>25,0</b>	<b>11</b>	<b>55,0</b>	<b>20</b>		<b>8</b>	<b>40,0</b>	<b>5</b>	<b>25,0</b>	<b>7</b>	<b>35,0</b>	<b>19</b>		<b>6</b>	<b>31,6</b>	<b>6</b>	<b>31,6</b>	<b>7</b>	<b>36,8</b>
Ontwerp	5		2	40,0	1	20,0	2	40,0	9		5	55,6	1	11,1	3	33,3	5		4	80,0	0	0,0	1	20,0
Lay-out	15		2	13,3	4	26,7	9	60,0	11		3	27,3	4	36,4	4	36,4	14		2	14,3	6	42,9	6	42,9
<b>Overig</b>	<b>3</b>		<b>1</b>	<b>33,3</b>	<b>2</b>	<b>66,7</b>	<b>0</b>	<b>0,0</b>	<b>5</b>		<b>0</b>	<b>0,0</b>	<b>5</b>	<b>100,0</b>	<b>0</b>	<b>0,0</b>	<b>1</b>		<b>0</b>	<b>0,0</b>	<b>0</b>	<b>0,0</b>	<b>1</b>	<b>100,0</b>

Tabel 5.6: Unieke problemen die bij meerdere personen gevonden zijn, per categorie met hoeveelheid geobserveerd en/of geverbaliseerd. Gesplitst voor de drie methoden van hardopdenken; concurrent (CTA), retrospectief zonder oogbewegingen (RTA) en retrospectief met oogbewegingen (RTE).

	CTA								RTA								RTE							
	Tot		Observ		Beide		Verb		Tot		Observ		Beide		Verb		Tot		Observ		Beide		Verb	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
<b>Totaal</b>	<b>44</b>		<b>10</b>	<b>22,7</b>	<b>25</b>	<b>56,8</b>	<b>9</b>	<b>20,5</b>	<b>53</b>		<b>15</b>	<b>28,3</b>	<b>30</b>	<b>56,6</b>	<b>8</b>	<b>15,1</b>	<b>49</b>		<b>12</b>	<b>24,5</b>	<b>26</b>	<b>53,1</b>	<b>11</b>	<b>22,4</b>
<b>Inhoud</b>	<b>12</b>		<b>1</b>	<b>8,3</b>	<b>5</b>	<b>6,0</b>	<b>6</b>	<b>50,0</b>	<b>15</b>		<b>0</b>	<b>0,0</b>	<b>10</b>	<b>66,7</b>	<b>5</b>	<b>33,3</b>	<b>14</b>		<b>3</b>	<b>21,4</b>	<b>4</b>	<b>28,6</b>	<b>7</b>	<b>50,0</b>
Begrijpelijkheid	4		0	0,0	0	0,0	4	100,0	6		0	0,0	4	66,7	2	33,3	16		3	18,8	8	50,0	5	31,3
Relevantie	4		1	25,0	3	75,0	0	0,0	4		0	0,0	4	100,0	0	0,0	3		0	0,0	3	100,0	0	0,0
Volledigheid	4		0	0,0	2	50,0	2	50,0	5		0	0,0	2	40,0	3	60,0	4		0	0,0	1	25,0	3	75,0
<b>Navigatie</b>	<b>19</b>		<b>5</b>	<b>26,3</b>	<b>14</b>	<b>73,7</b>	<b>0</b>	<b>0,0</b>	<b>23</b>		<b>8</b>	<b>34,8</b>	<b>13</b>	<b>56,5</b>	<b>2</b>	<b>8,7</b>	<b>22</b>		<b>4</b>	<b>18,2</b>	<b>16</b>	<b>72,7</b>	<b>2</b>	<b>9,1</b>
Gebruikersgemak	3		1	33,3	2	66,7	0	0,0	2		2	100,0	0	0,0	0	0,0	4		0	0,0	4	100,0	0	0,0
Structuur	4		0	0,0	4	100,0	0	0,0	7		0	0,0	6	85,7	1	14,3	4		0	0,0	3	75,0	1	25,0
Label	8		1	12,5	7	87,5	0	0,0	9		1	11,1	7	77,8	1	11,1	11		1	9,1	9	81,8	1	9,1
Snelheid	5		3	60,0	1	20,0	1	20,0	5		5	100,0	0	0,0	0	0,0	3		3	100,0	0	0,0	0	0,0
Overig	0		0	0,0	0	0,0	0	0,0	0		0	0,0	0	0,0	0	0,0	0		0	0,0	0	0,0	0	0,0
<b>Design</b>	<b>12</b>		<b>4</b>	<b>0,0</b>	<b>5</b>	<b>41,7</b>	<b>3</b>	<b>25,0</b>	<b>13</b>		<b>7</b>	<b>53,8</b>	<b>5</b>	<b>38,5</b>	<b>1</b>	<b>7,7</b>	<b>13</b>		<b>5</b>	<b>38,5</b>	<b>6</b>	<b>46,2</b>	<b>2</b>	<b>15,4</b>
Ontwerp	4		2	50,0	1	25,0	1	25,0	6		5	83,3	1	16,7	0	0,0	4		4	100,0	0	0,0	0	0,0
Lay-out	8		2	25,0	4	50,0	2	25,0	7		2	28,6	4	57,1	1	14,3	9		1	11,1	6	66,7	2	22,2
<b>Overig</b>	<b>1</b>		<b>0</b>	<b>0,0</b>	<b>1</b>	<b>100,0</b>	<b>0</b>	<b>0,0</b>	<b>2</b>		<b>0</b>	<b>0,0</b>	<b>2</b>	<b>100,0</b>	<b>0</b>	<b>0,0</b>	<b>0</b>		<b>0</b>	<b>0,0</b>	<b>0</b>	<b>0,0</b>	<b>0</b>	<b>0,0</b>

Bij de problemen in de subcategorieën Begrijpelijkheid en Volledigheid is dezelfde trend te zien (meer verbalisaties dan observaties), deze is echter alleen bij de unieke problemen (tabel 5.5) in de CTA conditie significant (Begrijpelijkheid:  $t(9)=-2,71$ ,  $p=.02$ , Volledigheid:  $t(7)=-4,58$ ,  $p=.003$ ). In de RTA conditie is het verschil tussen verbalisaties en observaties in de categorie Begrijpelijkheid significant bij zowel het totaal aantal probleemdetecties (tabel 5.4:  $t(19)=-2,10$ ,  $p=.05$ ) als bij de unieke problemen (tabel 5.5:  $t(9)=-3,00$ ,  $p=.02$ ).

In de subcategorie Relevantie is het aantal problemen dat door observaties gevonden is groter dan het aantal gevonden met verbalisaties. Deze verschillen zijn echter voor geen van de probleemsets significant.

Wanneer de drie condities met elkaar vergeleken worden, is er bij de categorie Inhoud alleen een significant verschil wanneer er naar het totaal aantal probleemdetecties gekeken wordt (tabel 5.4:  $F(2,57)=4,70$ ,  $p=.01$ ). Hierbij worden in de RTE conditie meer inhoudelijke problemen opgespoord via observatie opgespoord dan in de CTA conditie. Deze trend is ook in de andere probleemsets te zien.

### ***Navigatie***

In de categorie Navigatie is in alle drie de condities het grootste deel van de problemen gevonden via beide methoden. Ook via alleen observatie en alleen verbalisatie zijn echter weer problemen gevonden. Over het algemeen worden er in alle condities in de categorie Navigatie meer problemen geobserveerd dan geverbaliseerd. Dit verschil is echter alleen bij de problemen die bij meerdere proefpersonen zijn gevonden in de CTA conditie significant ( $t(18)=2,54$ ,  $p=.02$ ).

Bij de subcategorie Snelheid zijn in alle condities en probleemsets meer problemen gevonden via observaties dan verbalisaties. Het verschil tussen de methoden is daadwerkelijk significant in de RTE conditie bij het totaal aantal probleemdetecties (tabel 5.4:  $t(19)=3,20$ ,  $p=.01$ ) en de unieke problemen (tabel 5.5:  $t(5)=5,00$ ,  $p=.004$ ) en in de RTA conditie bij de probleemset met alle probleemdetecties (tabel 5.4:  $t(19)=2,44$ ,  $p=.03$ ).

Ook bij de subcategorie Structuur is een significant verschil gevonden. In de RTE conditie zijn bij de set met unieke problemen meer problemen via verbalisatie dan observatie gevonden. Deze zelfde trend is echter ook in alle andere condities in alle probleemsets gevonden.

Tussen de drie condities onderling bestaan geen significante verschillen tussen het aantal geobserveerde of geverbaliseerde problemen.

### ***Design***

Met welke methode de meeste problemen zijn opgespoord is bij de categorie Design lastig te zeggen. Dit verschilt heel erg per probleemset en conditie. In de RTE conditie worden bij het totaal aantal probleemdetecties (tabel 5.4) en de unieke problemen die bij meerdere personen gevonden zijn (tabel 5.6) bijvoorbeeld het grootste deel van de problemen gevonden via beide methoden, terwijl dit bij de unieke problemen (tabel 5.5) via verbalisaties is.

Tussen het aantal problemen dat binnen elke conditie geobserveerd of geverbaliseerd is, is alleen in de probleemset met unieke problemen die bij meerdere personen gevonden zijn (tabel 5.6) een significant verschil gevonden in de RTA conditie ( $t(12)=2,52$ ,  $p=.03$ ). Hierbij zijn meer problemen via observatie dan verbalisatie aan het licht gekomen. In de andere probleemsets komt dezelfde trend naar voren. Bij de CTA en RTE conditie is het beeld niet eenduidig en zijn er geen significante verschillen gevonden.

Bij de subcategorieën is er in de categorie Ontwerp een significant verschil gevonden in de RTE conditie bij de probleemset met unieke problemen die meerdere keren gevonden zijn (tabel 5.6). Hier zijn meer problemen geobserveerd dan geverbaliseerd ( $t(5)=5,00$ ,  $p=.004$ ). Bij de andere twee probleemsets en condities is dit ook het geval, maar hier zijn de verschillen niet significant.

Ook bij de categorie Lay-out is een significant verschil gevonden. Bij de unieke problemen (tabel 5.5) zijn in de CTA conditie meer problemen via verbalisaties opgespoord dan via observaties ( $t(14)=-2,43$ ,

$p=.03$ ). Deze trend is ook bij de andere sets en in de RTE conditie te zien, maar daar is het verschil niet significant. In de RTA conditie verschilt het per probleemset of er meer geverbaliseerd of geobserveerd wordt.

Tussen de drie condities bestaat geen significant verschil in het aantal problemen dat is opgespoord via de observaties of de verbalisaties.

### 5.3.2 Conclusie

In deze paragraaf zijn analyses uitgevoerd om te achterhalen of er tussen de condities verschillen zijn in de soort problemen die opgespoord worden en de verhouding tussen geverbaliseerde en geobserveerde problemen.

In zowel de concurrent (CTA) als de retrospectieve methoden met en zonder oogbewegingen (RTA, RTE) condities zijn problemen opgespoord door middel van observaties en verbalisaties. Een groot deel hiervan is via beide methoden gevonden, maar er zijn ook problemen die alleen via observaties gevonden zijn of alleen via verbalisaties. De verschillen tussen de condities voor de hoeveelheid problemen die via de drie 'methoden' gevonden zijn, zijn niet significant. De overlap tussen de problemen in de verschillende condities is laag. Slechts tussen de 16% en 19% van de problemen wordt in twee condities gevonden en slechts 10% in alle drie.

In alle drie de condities zijn problemen gevonden in alle probleemcategorieën. In de categorie Inhoud zijn veel problemen via beide methoden gevonden. Van de overige problemen, die slechts via één methode gevonden zijn, zijn er in alle condities meer opgespoord via de verbalisaties. Dit geldt ook voor de subcategorieën Begrijpelijkheid en Volledigheid.

In de categorie Navigatie zijn juist meer problemen aan het licht gekomen door observaties. Dit is het geval in alle drie de condities. De subcategorie Snelheid laat ook in zowel de CTA als RTA en RTE conditie dezelfde verhouding zien. Ook in de categorie Structuur bestaat er een verschil tussen het aantal problemen dat geverbaliseerd of geobserveerd wordt. Hier is het aandeel van verbalisaties echter groter. Dit is echter ook weer voor alle drie de condities hetzelfde.

Bij de Design problemen zijn er in alle drie de condities meer problemen gevonden via observatie dan door verbalisatie. Dit is ook het geval in de subcategorie Ontwerp. Bij de categorie Lay-out zijn er in de CTA en RTA conditie meer problemen via verbalisatie opgespoord, bij de RTA conditie is dit niet eenduidig te zeggen.

Tussen de drie condities zijn er geen verschillen tussen het aantal alleen geobserveerde of geverbaliseerde problemen, niet in totaal en niet in de verschillende probleemcategorieën.

Al met al zijn er tussen de drie condities maar weinig verschillen naar voren gekomen.

## 5.4 Verschillen tussen proefpersonen

Deze paragraaf bespreekt de verschillen tussen groepen van proefpersonen. In §5.4.1 kijk ik naar de verschillen tussen verschillende leeftijdscategorieën, in §5.4.2 opleiding en tot slot in §5.4.3 naar geslacht.

### 5.4.1 Leeftijd

#### *Tijd*

Voor de tijd die de proefpersonen nodig hadden om de taken uit te voeren is er geen sprake van een significant verschil tussen de verschillende leeftijdscategorieën ( $F(3,56) = 3,47, p=.07$ ).

#### *Problemen*

In tabel 5.7 is te zien hoeveel problemen de proefpersonen gemiddeld ondervonden in elk van de leeftijdsgroepen. Uit de Anova's blijkt dat er inderdaad verschillen zijn. Voor jongeren in de categorie

tot 29 jaar zijn minder problemen gevonden dan voor ouderen vanaf 40 jaar ( $F(3,56)=4,32, p=.01$ ). In de subcategorieën is dit verschil terug te vinden bij Navigatie ( $F(3,56)=7,03, p<.001$ ). Binnen deze categorie bereikt het verschil het niveau van significantie bij de subcategorieën Label en Snelheid.

**Tabel 5.7: Gemiddeld aantal problemen per proefpersoon per leeftijdscategorie voor de verschillende categorieën.**

	18-29	30-39	40-54	55+	F (3,56)=
<b>Totaal</b>	<b>5,5 (2,5)</b>	<b>6,2 (2,7)</b>	<b>8,5 (3,2)</b>	<b>8,2 (2,8)</b>	4,32, p=.01*
<b>Inhoud</b>	<b>2,0 (0,8)</b>	<b>1,8 (1,1)</b>	<b>1,7 (1,2)</b>	<b>2,4 (1,5)</b>	1,11, p=.35
Begrijpelijkheid	0,6 (0,7)	0,3 (0,7)	0,4 (0,6)	0,9 (1,0)	1,84, p=.15
Relevantie	0,8 (0,6)	0,9 (0,5)	0,9 (0,8)	0,8 (0,7)	1,18, p=.33
Volledigheid	0,6 (0,6)	0,5 (0,7)	0,5 (0,6)	0,7 (0,7)	0,28, p=.84
<b>Navigatie</b>	<b>1,4 (1,3)</b>	<b>2,7 (1,5)</b>	<b>4,1 (2,7)</b>	<b>4,0 (1,8)</b>	7,03, p<.001*
Gebruikersgemak	0,2 (0,5)	0,3 (0,5)	0,5 (0,7)	0,7 (0,7)	1,91, p=.14
Structuur	0,3 (0,6)	0,6 (0,8)	0,5 (0,6)	0,9 (0,8)	2,61, p=.06
Label	0,8 (0,8)	1,6 (1,2)	2,3 (1,7)	1,7 (1,1)	3,76, p=.02*
Snelheid	0,2 (0,4)	0,3 (0,5)	0,7 (0,7)	0,7 (0,8)	2,93, p=.04*
Overig	0	0	0,1 (0,4)	0,1 (0,2)	1,23, p=.31
<b>Design</b>	<b>2,0 (1,3)</b>	<b>1,7 (1,3)</b>	<b>2,3 (1,5)</b>	<b>1,7 (1,5)</b>	0,66, p=.58
Ontwerp	0,3 (0,6)	0,3 (0,4)	0,5 (0,6)	0,6 (0,7)	1,15, p=.34
Lay-out	1,8 (1,2)	1,4 (1,2)	1,8 (1,4)	1,1 (1,3)	1,13, p=.35
<b>Overig</b>	<b>0,1 (0,3)</b>	<b>0,1 (0,3)</b>	<b>0,4 (0,9)</b>	<b>0,1 (0,5)</b>	1,21, p=.32

Voor elk van de categorieën problemen is ook getoetst of er verschillen zijn tussen de leeftijdscategorieën in de manier waarop de problemen gevonden zijn, via de observaties of de verbalisaties. Hieruit blijkt dat er bij jongeren over het totaal en in de hoofdcategorie Navigatie meer problemen alleen opgespoord zijn via de observaties dan bij ouderen. Terugkijkend naar figuur 5.1 (pagina 28) betekent dit dat groep D groter is dan groep E. In tabel 5.8 staan de significante verschillen.

**Tabel 5.8: Gemiddeld aantal problemen per proefpersoon per leeftijdscategorie. Significante verschillen tussen het aantal alleen geobserveerde problemen in de leeftijdscategorieën.**

	18-29		30-39		40-54		55+		O	F (3,56)=
	O	V	O	V	O	V	O	V		
<b>Totaal</b>	<b>1,3 (0,8)</b>	<b>2,0 (1,5)</b>	<b>1,5 (1,1)</b>	<b>1,5 (2,0)</b>	<b>2,2 (1,6)</b>	<b>2,5 (1,7)</b>	<b>2,7 (1,7)</b>	<b>1,8 (1,7)</b>	3,55, p=.02*	
<b>Navigatie</b>	<b>0,3 (0,4)</b>	<b>0,4 (0,6)</b>	<b>0,4 (0,5)</b>	<b>0,5 (0,9)</b>	<b>1,3 (1,3)</b>	<b>0,7 (0,9)</b>	<b>1,4 (1,1)</b>	<b>0,6 (1,1)</b>	6,36, p=.001*	
Structuur	0	0,2 (0,5)	0	0,2 (0,4)	0	0,1 (0,3)	0,2 (0,4)	0,2 (0,4)	2,87, p=.05*	
Snelheid	0,1 (0,3)	0	0,2 (0,4)	0	0,5 (0,7)	0,1 (0,3)	0,7 (0,7)	0,1 (0,2)	3,09, p=.03*	
<b>Design</b>										
Ontwerp	0,1 (0,3)	0,2 (0,5)	0,1 (0,3)	0,1 (0,3)	0,3 (0,5)	0,2 (0,6)	0,5 (0,7)	0	3,19, p=.03*	

*O zijn de problemen gevonden door alleen observaties, V zijn de problemen gevonden door alleen verbalisaties.*

## 5.4.2 Opleiding

### **Tijd**

Voor de verschillende opleidingsniveaus is er geen significant verschil in de tijd die de proefpersonen nodig hebben om de taak uit te voeren ( $F(2,57)=0,51, p=.61$ ).

### **Problemen**

In tabel 5.9 is te zien dat er over het algemeen geen verschil is tussen de verschillende opleidingsniveaus. Alleen bij problemen met de structuur van de site is er sprake van een significant verschil. Uit de post-hoc Scheffé test blijkt echter dat er geen verschil tussen twee van de niveaus bestaat, alleen een algemeen verschil.

**Tabel 5.9: Gemiddeld aantal problemen per proefpersoon per opleidingsniveau voor de verschillende categorieën.**

	Laag	Midden	Hoog	F (2,57)=
<b>Totaal</b>	<b>7,2 (2,9)</b>	<b>6,8 (3,0)</b>	<b>7,5 (3,3)</b>	0,25, p=.78
<b>Inhoud</b>	<b>1,9 (1,2)</b>	<b>1,8 (1,1)</b>	<b>2,3 (1,2)</b>	0,95, p=.39
Begrijpelijkheid	0,8 (0,9)	0,5 (0,8)	0,5 (0,7)	0,62, p=.54
Relevantie	0,7 (0,5)	0,8 (0,7)	1,0 (0,8)	1,02, p=.37
Volledigheid	0,4 (0,5)	0,5 (0,6)	0,8 (0,8)	1,72, p=.19
<b>Navigatie</b>	<b>3,8 (2,1)</b>	<b>2,6 (2,0)</b>	<b>2,9 (2,3)</b>	1,66, p=.20
Gebruikersgemak	0,3 (0,6)	0,4 (0,6)	0,5 (0,7)	0,46, p=.63
Structuur	0,9 (0,9)	0,4 (0,6)	0,4 (0,7)	3,27, p=.05*
Label	1,9 (1,5)	1,4 (1,2)	1,5 (1,2)	0,82, p=.45
Snelheid	0,6 (0,6)	0,4 (0,7)	0,4 (0,7)	0,23, p=.79
Overig	0,1 (0,3)	0	0,1 (0,2)	1,25, p=.29
<b>Design</b>	<b>1,4 (1,1)</b>	<b>2,1 (1,3)</b>	<b>2,1 (1,7)</b>	1,77, p=.18
Ontwerp	0,3 (0,5)	0,4 (0,7)	0,4 (0,7)	0,15, p=.86
Lay-out	1,1 (0,9)	1,7 (1,2)	1,7 (1,6)	1,52, p=.23
<b>Overig</b>	<b>0,1 (0,2)</b>	<b>0,2 (0,8)</b>	<b>0,2 (0,5)</b>	0,54, p=.59

Bij het vergelijken van het aantal geobserveerde of geverbaliseerde problemen bij de proefpersonen uit de verschillende opleidingsniveaus is geen significant verschil gevonden voor het totaal aantal problemen en de meeste subcategorieën. Wel zijn er in de categorie Design en de subcategorie Lay-out door de hoger opgeleiden meer problemen geverbaliseerd dan door de lager opgeleiden. Tabel 5.10 laat deze significante verschillen zien.

**Tabel 5.10: Gemiddeld aantal problemen per proefpersoon per opleidingsniveau. Significante verschillen tussen het aantal alleen geverbaliseerde problemen in de opleidingsniveaus.**

	Laag		Midden		Hoog		V	F (2,57)=
	O	V	O	V	O	V		
<b>Totaal</b>	<b>1,4 (0,3)</b>	<b>1,6 (0,4)</b>	<b>1,7 (0,4)</b>	<b>1,5 (0,3)</b>	<b>1,2 (0,3)</b>	<b>1,9 (0,4)</b>	2,62, p=.08	
<b>Design</b>	<b>0,9 (0,2)</b>	<b>0,2 (0,1)</b>	<b>0,6 (0,1)</b>	<b>1,1 (0,2)</b>	<b>0,7 (0,1)</b>	<b>1,0 (0,2)</b>	4,30, p=.02*	
Lay-out	0,7 (0,2)	0,2 (0,1)	0,6 (0,1)	1,0 (0,2)	0,4 (0,1)	0,9 (0,2)	3,33, p=.04*	

### 5.4.3 Geslacht

#### *Tijd*

Tussen mannen en vrouwen is er geen verschil in de tijd die ze gemiddeld nodig hebben om de taak uit te voeren ( $t(58) = 1,38, p = .17$ ).

#### *Problemen*

Tabel 5.11 laat zien dat er over het algemeen geen verschil is tussen het gemiddeld aantal problemen dat bij mannen en vrouwen gevonden is. Alleen in de subcategorie Label (bij Navigatie) zijn er meer problemen geconstateerd bij vrouwen dan bij mannen.

**Tabel 5.11: Gemiddeld aantal problemen per proefpersoon per geslacht voor de verschillende categorieën.**

	Man	Vrouw	t
<b>Totaal</b>	<b>6,4 (2,6)</b>	<b>7,8 (3,2)</b>	t(58) = -1,94, p = .06
<b>Inhoud</b>	<b>3,0 (1,3)</b>	<b>2,1 (1,2)</b>	t(58) = -0,87, p = .39
Begrijpelijkheid	0,4 (0,7)	0,8 (0,9)	t(58) = -1,56, p = .13
Relevantie	0,9 (0,7)	0,8 (0,6)	t(58) = -0,26, p = .80
Volledigheid	0,6 (0,7)	0,6 (0,6)	t(58) = 0,05, p = .96
<b>Navigatie</b>	<b>2,5 (1,8)</b>	<b>3,6 (2,4)</b>	t(58) = -1,99, p = .05
Gebruikersgemak	0,5 (0,6)	0,3 (0,70)	t(58) = 0,93, p = .35
Structuur	0,4 (0,6)	0,7 (0,9)	t(56,0) = -1,57, p = .12
Label	1,1 (1,0)	1,9 (1,4)	t(56,3) = -2,48, p = .02*
Snelheid	0,4 (0,6)	0,6 (0,7)	t(58) = -1,23, p = .23
Overig	0,1 (0,3)	0,0 (0,2)	t(58) = 0,70, p = .49
<b>Design</b>	<b>1,9 (1,3)</b>	<b>1,9 (1,5)</b>	t(58) = -0,22, p = .83
Ontwerp	0,3 (0,5)	0,5 (0,7)	t(58) = -0,92, p = .36
Lay-out	1,5 (1,2)	1,5 (1,4)	t(58) = 0,20, p = .85
<b>Overig</b>	<b>0,1 (0,7)</b>	<b>0,2 (0,6)</b>	t(58) = -0,31, p = .76

Uit de analyses van het aantal alleen geverbaliseerde of geobserveerde problemen blijkt dat er in totaal bij mannen minder problemen geobserveerd zijn dan bij vrouwen. Dit zelfde doet zich voor in de subcategorie Ontwerp bij Design. Deze verschillen staan in tabel 5.12.

**Tabel 5.12: Gemiddeld aantal problemen per proefpersoon per geslacht. Significante verschillen tussen het aantal alleen geobserveerde problemen bij mannen en vrouwen.**

	Man		Vrouw		t
	O	V	O	V	
<b>Totaal</b>	<b>1,5 (1,1)</b>	<b>1,8 (1,6)</b>	<b>2,3 (1,6)</b>	<b>2,1 (1,9)</b>	t(58) = -2,06, p = .04*
<b>Design</b>					
Ontwerp	0,1 (0,3)	0,2 (0,5)	0,4 (0,6)	0,1 (0,4)	t(47,7) = -2,18, p = .03*

#### 5.4.4 Conclusie

Bij het vergelijken van de leeftijdscategorieën is een verschil naar voren gekomen tussen de jongeren en de ouderen. Jongeren zijn minder problemen tegen gekomen en dan vooral in de categorie Navigatie. Dit verschil zit vooral in het aantal observaties, bij jongeren zijn in het algemeen en de probleemcategorie Navigatie minder problemen geobserveerd dan bij ouderen.

Tussen hoger en lager opgeleiden is geen algemeen verschil gevonden in het gemiddeld aantal problemen dat per persoon gevonden is. Alleen in de subcategorie Structuur (bij Navigatie) zijn er voor hoger opgeleiden minder problemen geconstateerd. Over het algemeen kan dus gezegd worden dat er nauwelijks een verschil is tussen het aantal problemen dat voor de verschillende opleidingsniveaus gevonden is. Uit het vergelijken van de geverbaliseerde en geobserveerde problemen blijkt dat de hoger opgeleiden meer problemen in de categorie Design en subcategorie Lay-out hebben geverbaliseerd dan de lager opgeleiden.

Ook bij mannen en vrouwen is er nauwelijks verschil tussen het aantal problemen dat gevonden wordt. Alleen in de categorie label zijn er bij vrouwen meer problemen opgespoord dan voor mannen. Bij het vergelijken van de aantallen observaties en verbalisaties is er wel een verschil. Bij mannen zijn minder problemen geobserveerd. Dit is ook te zien in de subcategorie Ontwerp (bij Design).



## 6 Conclusies

### 6.1 Beantwoording van de deelvragen

In deze paragraaf geef ik op basis van de hiervoor gepresenteerde resultaten een antwoord op de deelvragen van dit onderzoek en plaats ik deze in een breder kader.

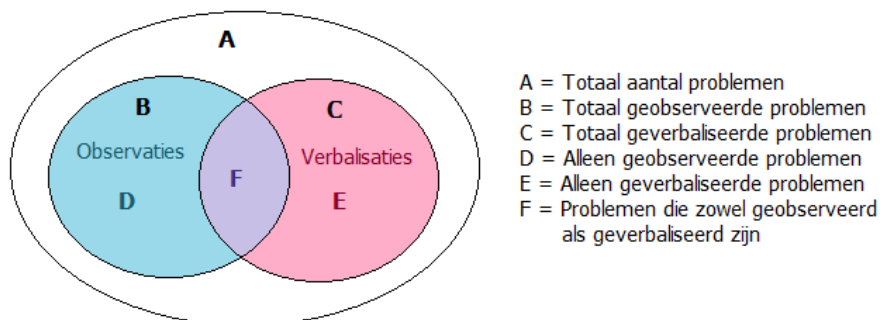
#### 6.1.1 Problemen via observaties en verbalisaties

1. *Welke problemen kan men aanwijzen door middel van schermobservaties (eyetracking data en het klikpad/muisbewegingen)?*
2. *Welke problemen kan men opsporen door hardopdenken?*
3. *In hoeverre zijn de problemen die opgespoord worden door observaties en de hardopdenkmethode hetzelfde of verschillend?*

#### **Aantal**

In dit onderzoek is het totaal aantal problemen dat met observaties en verbalisaties gevonden wordt ongeveer gelijk. Uit de analyses blijkt dat een groot deel van de problemen via beide methodes opgespoord kan worden. Maar bij beide methoden is er een aanzienlijk deel dat uniek is (deel D en E in figuur 6.1). Bij de verbalisaties is dit aandeel unieke problemen over het geheel genomen groter dan bij de observaties. Wanneer er echter gekeken wordt naar de unieke problemen die door meer dan één persoon gevonden worden (net als Vereijken, Van der Burg en De Bode gedaan hebben) is er geen verschil is tussen de twee methoden.

De verwachting voor deze deelvragen was dat er met observaties meer problemen gevonden zouden worden dan via verbalisaties (Vereijken 2009, Van der Burg 2009 en De Bode 2009). Dat er in dit onderzoek minder unieke problemen via observaties dan verbalisaties naar voren kan deels liggen aan de probleemset die geanalyseerd is. Een andere verklaring kan liggen in de manier waarop de probleemsets tot stand zijn gekomen. Ik heb problemen die afgeleid kunnen worden uit een onjuist antwoord op de vraag uit de taak zowel bij de observaties als de verbalisaties gerekend. Wanneer ik dit niet had gedaan, was de groep met problemen die via de verbalisaties opgespoord zijn (C in figuur 6.1) een stuk kleiner geweest.



*Figuur 6.1: Visualisatie van data*

#### **Soorten**

Zowel met de observaties als met de hardopdenkmethode zijn problemen gevonden in alle probleemcategorieën. Zowel de hoofdcategorieën Inhoud, Navigatie en Design als de bijbehorende subcategorieën. In de onderzoeken van Vereijken (2009), Van der Burg (2009) en De Bode (2009) was het verschil tussen de twee methodes dat aan de hand van de oogbewegingen vooral veel problemen in de categorie Navigatie en de subcategorie Begrijpelijkheid (bij Inhoud) gevonden werden. Terwijl dit voor de hardopdenkmethode meer verspreid was over alle hoofd- en subcategorieën. De observaties in dit onderzoek geven dus een beeld dat meer divers is dan alleen de analyse van oogbewegingen.

Het vergelijken van het aantal problemen dat uniek is voor de observaties of voor de verbalisaties per soort probleem laat echter wel een aantal verschillen zien.

In de categorie Inhoud en de bijbehorende subcategorieën Begrijpelijkheid en Volledigheid zijn meer problemen gevonden die alleen via verbalisaties aan het licht zijn gekomen dan alleen via observaties. Bij de subcategorie Relevantie is er geen eenduidig beeld naar voren gekomen.

Dat in de categorie Begrijpelijkheid meer problemen gevonden zijn door de verbalisaties is het tegenovergestelde van wat De Bode (2009) en Van der Burg (2009) vonden. Mogelijk komt dit door de definitie die ik voor de categorie Relevantie gebruik heb. In de categorie Relevantie heb ik bijvoorbeeld niet alleen problemen opgenomen waarbij de informatie op de site niet voldoet "aan de behoeften van de lezer" zoals Van der Burg (2009) en De Bode (2009), maar ook wanneer er een onvolledig antwoord wordt gegeven op de vragen. Vooral bij de vraag naar de voorwaarden waaraan men moet voldoen om de Starterslening te krijgen waren er veel aspecten die de proefpersonen zouden moeten noemen. Vaak noemden ze echter alleen de laatste voorwaarde die ze zagen. Dit hoeft niet te betekenen dat ze de tekst niet goed begrepen hebben. De kans is groot dat ze de eerder genoemde voorwaarden simpelweg niet als de doorslaggevende voorwaarden zagen of dat ze geen zin hadden om deze allemaal op te noemen. Een onjuist antwoord op de vraag hebben De Bode en Van der Burg echter wel als begripsprobleem gecategoriseerd.

Voor de categorie Navigatie zijn over het algemeen de meeste problemen via beiden methoden gevonden. Er is ook geen (significant) verschil tussen het totaal aantal problemen dat alleen geobserveerd of alleen geverbaliseerd is. Vereijken (2008) De Bode (2009) en Van der Burg (2009) vonden juist dat er vooral veel problemen in deze categorie opgespoord werden via observaties. In de subcategorieën Gebruikersgemak en Snelheid laat dit onderzoek hetzelfde zien, maar in de categorie Labels is het resultaat tegenovergesteld, er zijn meer problemen geverbaliseerd.

Bij de categorie Design is er geen eenduidig beeld voor de manier waarop de problemen gevonden worden. In de drie geanalyseerde probleemsets wordt de ene keer meer geverbaliseerd terwijl in de andere meer geobserveerd wordt. Ditzelfde beeld laat de subcategorie Ontwerp zien. Bij de categorie Lay-out is er wel een verschil, daar zijn meer problemen geverbaliseerd dan geobserveerd.

Er bestaat dus vooral in de categorieën Inhoud en Navigatie een verschil tussen de aantallen problemen die via observatie of verbalisatie gevonden kunnen worden.

### **6.1.2 Verschillen tussen CTA, RTA en RTE**

#### *4. Verschilt het per conditie van de hardopdenkmethode welke problemen er gevonden worden en hoe dit in verhouding staat met de geobserveerde problemen?*

In dit onderzoek zijn bij het kijken naar de aantallen problemen bij de drie varianten van de hardopdenkmethode geen verschillen gevonden. Niet voor het totaal aantal problemen, noch voor het aantal in elke probleemcategorie. Dit is een ander resultaat dan in het onderzoek van Ball e.a. (2007). Zij vonden meer problemen in de retrospectieve conditie met oogbewegingen (RTE) dan in de concurrent conditie (CTA). Tussen de retrospectieve conditie zonder oogbewegingen (RTA) en de andere condities vonden zij ook geen verschillen. Een mogelijke verklaring hiervoor zou kunnen zijn dat Ball e.a. (2007) gebruik maakten van de relaxte manier van hardopdenken, dus aansporingen door middel van vragen. Hertzum e.a. (2009) hebben laten zien dat deze manier van hardopdenken meer reactiviteit veroorzaakt, terwijl dit bij de klassieke manier, die in dit onderzoek is gebruikt, minder het geval is. In het onderzoek van Van den Haak e.a. (2007, 2009), dat ook gebruik heeft gemaakt van de klassieke manier van hardopdenken heeft echter ook geen verschillen gevonden.

De overlap tussen de problemen in de drie condities is laag. Tussen twee condities is de overlap tussen de 16% en 19% en tussen alle drie is dit slechts 10%. Er worden in alle drie de condities dus veel 'nieuwe' problemen opgespoord.

Bij het vergelijken van de verbalisaties en observaties binnen de drie condities is hetzelfde beeld te zien: er wordt niet duidelijk meer geobserveerd dan geverbaliseerd. Ook bij de categorie Inhoud is het beeld in alle drie de condities hetzelfde, er worden meer problemen geverbaliseerd dan geobserveerd. Bij de categorie Navigatie is dit precies andersom (meer observatie dan verbalisatie), maar dit geldt ook weer voor alle drie de condities. Voor de problemen in de categorie Design zijn helemaal geen significante verschillen gevonden. Tot slot zijn de condities onderling vergeleken. Hier is uitgekomen dat er geen verschil is in het aantal observaties of verbalisaties voor de drie condities. Dit komt overeen met wat Van den Haak e.a. (2007, 2009) in hun onderzoeken naar gemeentewebsites vonden.

Uit dit alles kan geconcludeerd worden dat er nauwelijks verschillen bestaan tussen de drie condities. Blijkbaar maakt het voor de hoeveelheid problemen die gevonden worden niet uit welke variant van de hardopdenkmethode gebruikt wordt. Omdat de overlap tussen de problemen die via de verschillende varianten aan het licht zijn gekomen laag is, is het echter wel nuttig om alle drie de varianten toe te passen.

### 6.1.3 Verschillen tussen proefpersonen

5. *Zijn proefpersoonkenmerken (opleiding, leeftijd, geslacht) van invloed op de soort en hoeveelheid problemen die gevonden worden?*

#### **Leeftijd**

De leeftijd van de proefpersonen lijkt daadwerkelijk van invloed te zijn op het aantal problemen dat gevonden wordt. Bij jongeren zijn minder problemen gevonden dan bij ouderen. Dit was vooral het geval in de categorie Navigatie. Dat bij jongeren minder problemen zijn opgespoord voldoet aan de verwachting die Elling e.a. (niet gepubliceerd) hadden bij hun onderzoek naar de online tool Infocus. Volgens hen zouden jongeren minder problemen hebben met de online tool, omdat ze er meer ervaring mee hebben en er daarom meer mee op hun gemak zouden zijn. Dit zou ook kunnen gelden voor het zoeken van informatie, veel jongeren gebruiken het internet dagelijks zoveel dat ze er niet van schrikken om informatie te moeten gaan zoeken. Een aantal ouderen leek hier echter wel moeite mee te hebben. Er zijn echter ook genoeg ouderen die ook veel gebruik maken van de computer, zowel tijdens hun werk als privé, deze verklaring is dus niet afdoende. Een mogelijke andere verklaring zou kunnen zijn dat jongeren op jongere leeftijd zijn begonnen met het werken op de computer en internet en dat ze daar hun voordeel uit kunnen halen bij het zoeken van de informatie die zij nodig hebben.

#### **Opleiding**

Op basis van de resultaten kan geconcludeerd worden dat er bij lager opgeleiden niet meer problemen gevonden zijn dan bij hoger opgeleiden. Alleen met het vinden van de informatie lijken ze meer moeite te hebben; in de subcategorie Structuur (bij Navigatie) zijn er meer problemen gevonden voor lager opgeleiden dan hoger opgeleiden. De vergelijking tussen de problemen die alleen geverbaliseerd zijn, heeft laten zien dat er voor het totaal aantal problemen geen verschil is tussen hoger en lager opgeleiden. Dit gaat in tegen de verwachting die ik had op basis van de onderzoeken van Elling e.a. (niet gepubliceerd) en Haas (2009). Zij hebben beide gevonden dat de verbalisaties van hoger opgeleiden een betere kwaliteit hebben dan die van lager opgeleiden. Dit zou dus voor dit onderzoek hebben kunnen betekenen dat er bij hoger opgeleiden meer problemen via observaties zouden worden gevonden. Blijkbaar is deze redenering niet juist en leiden betere verbalisaties niet tot meer geverbaliseerde problemen.

Kijkend naar de verschillende soorten problemen bij de hoger opgeleiden, zijn in de categorie Design, en daarbij vooral in de subcategorie Lay-out, meer problemen gevonden die geverbaliseerd zijn dan bij de lager opgeleiden. Deze lay-out problemen hebben voornamelijk betrekking op de pagina's waar de antwoorden op de vragen uit de taak staan. Blijkbaar konden de hoger opgeleiden als ze de juiste informatie eenmaal gevonden hadden meer van een afstandje kijken naar deze teksten en hoe ze eruitzien dan de lager opgeleiden. Het lijkt er dus op dat de kwalitatief betere verbalisaties wel degelijk kunnen leiden tot meer probleeminformaties, maar dat dit vooral gebeurt nadat men op de

juiste pagina is gekomen. Mogelijk kost het zo snel mogelijk vinden van de informatie zoveel van de aandacht dat er minder gelet wordt op mogelijke problemen met de manier waarop de pagina ingedeeld is (de lay-out).

Er zijn dus ook voor de verschillende opleidingsniveaus daadwerkelijk verschillen gevonden.

### ***Geslacht***

Tussen mannen en vrouwen is geen verschil gevonden voor het totaal aantal geconstateerde problemen. Kijkend naar de verschillende soorten problemen is er alleen een verschil gevonden bij de problemen met de labels van de links op de site. Hiermee hebben meer vrouwen problemen gehad dan mannen.

Bij het vergelijken van de problemen die alleen gevonden zijn via observatie of verbalisatie is er een verschil gevonden tussen het totaal aantal problemen. Bij mannen zijn minder problemen via observatie gevonden dan bij vrouwen. Dit is echter niet te wijten aan het totaal aantal problemen dat de proefpersonen uit de twee groepen tegengekomen zijn. Het verschil daartussen is, zoals vermeld, niet significant. Waar het wel aan ligt is mogelijk een nuttig onderwerp voor vervolgonderzoek.

## **6.2 Antwoord op de hoofdvraag**

*In hoeverre zijn schermobservaties (muis- en oogbewegingen) te gebruiken als zelfstandige probleemopsporende evaluatiemethode?*

Bij de beantwoording van de deelvragen is naar voren gekomen dat er bij zowel de observaties als bij de verbalisaties redelijk veel unieke problemen naar voren komen. De overlap tussen de gevonden problemen via observaties en via de hardopdenkmethode is groot, maar de verbalisaties zijn zeker nodig om tot een breed beeld van problemen te komen. In dit opzicht zijn schermobservaties alleen niet voldoende om de usability van een website te beoordelen.

Het is bovendien nuttig om zowel jongeren als ouderen, hoger en lager opgeleiden en mannen en vrouwen te vragen een website te evalueren. Zij vinden allemaal een aanzienlijk deel van de problemen.

Door middel van beide methoden is een verscheidenheid aan soorten problemen te vinden. In de drie hoofdcategorieën, Inhoud, Navigatie en Design, zijn via beide methoden problemen geconstateerd. Dit in tegenstelling tot onderzoeken waarin alleen de oogbewegingen zijn gebruikt om problemen op te sporen (Vereijken 2009, Van der Burg 2009, De Bode 2009). Hierbij werden via de oogbewegingen alleen problemen in een beperkt aantal categorieën gevonden (Navigatie en Begrijpelijkheid). Het is dus zeker nuttig om niet alleen te kijken naar oogbewegingen, maar ook de handelingen met de muis mee te nemen. Dan is het mogelijk om een grotere verscheidenheid aan problemen op te sporen.

## 7 Discussie

### 7.1 Probleempunten

#### *Kwaliteit oogbewegingen*

In dit onderzoek vormen de oogbewegingen een belangrijk deel van de observaties. Bij bijna de helft van de proefpersonen was hier een probleem mee. In een aantal gevallen (13) vielen de oogbewegingen deels weg, maar nadat de proefpersoon weer ging verzitten kwamen ze weer terug. Bij vier personen kwamen de oogbewegingen daarna echter niet meer terug en moest de rest van de analyse alleen gebaseerd worden op de muisbewegingen. Op de website van Apeldoorn was het mogelijk om redelijk wat te observeren, ook als er geen oogbewegingen waren. Dit omdat er op de startpagina via het menu of de zoekmogelijkheid van het digitaal loket gewerkt moet worden. Het menu klapt uit en om verder te gaan moet de muis over de subonderwerpen verplaatst worden. Dit zijn dus allemaal zaken die gezien kunnen worden. Op andere sites staat het hele menu al in beeld. Als er dan geen oogbewegingen zijn, is het niet alleen onmogelijk om te bepalen wat de proefpersoon leest op de doelpagina, maar kan men ook niet vaststellen wat de hij op de startpagina doet.

Naast het wegvallen van de oogbewegingen kwamen de oogbewegingen van minstens 10 proefpersonen ook niet helemaal overeen met wat ze precies aan het lezen waren. Vaak bevonden de oogbewegingen zich dan een stuk boven de regel die gelezen of de link die bekeken wordt. Wanneer ze bovenaan begonnen te lezen was dit te zien, dan begonnen ze een stuk boven de eerste regel, in het witte deel van de pagina. In deze gevallen was het wel te achterhalen wat men precies las: meestal waren de oogbewegingen dan één of twee regels boven wat de proefpersoon precies aan het lezen was. Begon men echter ergens halverwege de tekst, dan was dit niet mogelijk. Bovendien gingen sommige proefpersonen wel eens verzitten waardoor de ogen dus vanuit een andere hoek gemeten worden en niet meer overeenkomen met de instellingen aan het begin van de sessie. In deze gevallen is het dus mogelijk dat ik niet heb kunnen achterhalen of de oogbewegingen zich op de juiste plaats bevonden.

Hier moet in onderzoek zeker rekening mee gehouden worden en het is naar mijn mening een belangrijke reden om ook andere observaties op te nemen in de analyses. Die zijn altijd te zien en niet onderhevig aan veranderingen in de houding van de proefpersoon.

#### *Praten met onderzoeksleider*

Bij het bekijken van de filmpjes waarin proefpersonen aan het werk waren op de website gebeurde het een aantal keer dat de proefpersoon niet helemaal begreep wat de bedoeling was. Sommige proefpersonen gaven bijvoorbeeld geen antwoord op het moment dat ze het gevonden hadden, maar pas nadat ze meer informatie bekeken hadden. Dit was echter niet zozeer een probleem.

Er waren daarentegen ook proefpersonen die niet begrepen wat ze moesten doen. Hierover gingen ze dan vragen stellen aan de onderzoeksleider en die kon dan niets anders dan antwoord geven. Hierdoor werd de het werken van de proefpersoon echter wel beïnvloed. Een deel van de oogbewegingen en handelingen die op zo'n moment plaatsvonden waren dus niet echt gericht op het uitvoeren van de taak, maar meer verbonden aan het luisteren. Het was echter niet eenvoudig om vast te stellen wanneer het luisteren precies ophield en wanneer de proefpersoon weer inhoudelijk aan het werk ging. Dit zou de resultaten beïnvloed kunnen hebben. Daarom is het in vervolgonderzoek beter om zoveel mogelijk te voorkomen dat proefpersonen met de onderzoeksleider gaan praten.

#### *Probleembeschrijvingen*

Ik heb geprobeerd om voor elk probleem dat ik tegenkwam aan te geven wat er gebeurde, wat de oorzaak ervan was en wat het tot gevolg had. Door dit zo gestructureerd te doen hoopte ik beter te kunnen achterhalen welke problemen hetzelfde waren. Terwijl ik bezig was met het bekijken van de overlap tussen de verschillende problemen kwam ik er echter achter dat de beschrijvingen niet altijd goed genoeg waren. Dan was ik een deel vergeten of vroeg ik me bij het nalezen ervan af wat ik ook al weer precies bedoelde. Het lijkt me daarom nuttig om wanneer een soortgelijk onderzoek nogmaals uitgevoerd wordt, gebruik te maken van probleembeschrijvingen die nog strikter zijn.

Hierbij zou dan bovendien onderscheid gemaakt kunnen worden voor de manier waarop een probleem gevonden is en waarop de oorzaak achterhaald wordt. In dit onderzoek ben ik bij de observaties begonnen met onderscheid maken tussen de zaken die uit de handelingen met de muis te halen zijn (S), zaken die uit de oogbewegingen te halen zijn (O) en problemen die blijken uit de antwoorden (A). Meestal was er sprake van een combinatie van de drie omdat de zowel de oogbewegingen als de muishandelingen dan nodig waren om de onjuiste antwoorden te verklaren. Een voorbeeld hiervan is de situatie waarbij de proefpersoon een deel van de relevante informatie over de Starterslening niet bereikte door niet voldoende naar beneden te scrollen en wat hij wel zag alleen te scannen. Het was dan onmogelijk om het goede antwoord te geven.

Voor de verbalisaties heb ik onderscheid gemaakt tussen een probleem dat daadwerkelijk door de proefpersoon werd genoemd (VP) en een probleem dat ik achterhaalde door hun woorden te interpreteren (VI). Om de oorzaak van het probleem te achterhalen was het echter af en toe ook nodig om te bekijken wat er op het scherm gebeurde en wat de oogbewegingen waren. Uiteindelijk heb ik deze coderingen uit de analyses weggelaten doordat de code af en toe wel erg lang werd (zeker als ik bij de verbalisaties ook muis- en oogbewegingen nodig had). In vervolgonderzoek kunnen deze coderingen misschien wel gebruikt worden, maar dan gesplitst voor de gebeurtenissen die het probleem aan het licht brengen en de gebeurtenissen die de daadwerkelijke oorzaak aangeven.

## **7.2 Suggesties voor vervolgonderzoek**

### ***Hardopdenken***

Tijdens het analyseren van de filmpjes in de retrospectieve condities bleek af en toe dat de proefpersonen ook tijdens het uitvoeren van de taak hardop vertelden wat ze dachten. Dit was in een aantal gevallen niet hetzelfde als wat ze achteraf vertelden. Wanneer deze verbalisaties meegenomen waren in de analyses waren mogelijk nog meer problemen met de website aan het licht gekomen. Omdat een deel van het onderzoek bestond uit de vergelijking van de condities was het echter niet mogelijk om dit te doen, het was nodig om een strikte scheiding te maken tussen het concurrent en retrospectief hardopdenken. In vervolg onderzoek zou het echter nuttig kunnen zijn om bijvoorbeeld een conditie toe te voegen waarin zowel tijdens als na de taakuitvoering verteld wordt wat men doet en denkt.

### ***Kwalitatieve analyse oogbewegingen***

Uit dit onderzoek is gebleken dat het nuttig is om de oogbewegingen en andere observaties met elkaar te combineren. In dit onderzoek zijn de oogbewegingen echter alleen bekeken, er zijn geen analyses uitgevoerd waarin is vastgesteld hoe vaak ergens op gefixeerd werd en hoelang de verschillende fixaties en saccades precies duurden. Het voordeel hiervan is dat de verschillen in oogbewegingen tussen proefpersonen meegenomen kunnen worden. Sommige personen lezen bijvoorbeeld sneller dan anderen en de fixaties zijn ook niet bij iedereen even lang. Het toevoegen van de exacter gemeten fixatieduur en frequentie zou er echter voor kunnen zorgen dan de observaties iets objectiever worden.

### ***Problemen in relatie met werkproces***

In de conclusie met betrekking tot verschillen tussen de opleidingsniveaus van de proefpersonen (§6.1.3), is geconstateerd dat bij hoger opgeleiden meer problemen via de verbalisaties zijn gevonden dan bij de lager opgeleiden. Dit was het geval in de categorie Design, en dan meer specifiek met de lay-out van de pagina's. Deze problemen deden zich voornamelijk voor op de pagina's met informatie om de vragen uit de taak te beantwoorden. In eerdere onderzoeken (Elling e.a. niet gepubliceerd. Haas 2009) bleek de kwaliteit van verbalisaties van hoger opgeleiden beter te zijn. Uit de resultaten in dit onderzoek zou mogelijk afgeleid kunnen worden dat de betere verbalisaties vooral tot meer geverbaliseerde problemen leiden als de benodigde informatie gevonden is. Eerder lijken ze zich vooral bezig te houden met het zo snel mogelijk vinden van de juiste pagina en nemen ze niet echt afstand van waar ze mee bezig zijn. Voor vervolgonderzoek zou het interessant kunnen zijn om te achterhalen of het inderdaad afhankelijk is van het moment in het zoekproces wanneer er meer problemen geverbaliseerd worden.

## Bronvermelding

- Ball, L.J., N. Eger, N., R. Stevens & J. Dodd (2007) Cueing retrospective verbal reports in usability testing through eye-movement replay. In: L.J. Ball, M.A. Sasse, C. Sas, T.C. Ormerod, A. Dix & T. McEwan (eds) *People and Computers XXI: HCI...but not as we know it 1*, 129-137. Swinton: British Computer Society.
- Bode, M. de (2009) Weet ik wat jouw ogen zien? Een onderzoek naar de mogelijkheden van eyetracking als zelfstandige probleemopsporende methode in usability onderzoek. Masterscriptie. Universiteit Utrecht: Utrecht.
- Boren, M. T. & J. Ramey (2000) Thinking aloud: Reconciling theory and practice. *IEEE Transactions on professional communication* 43 (3), 261-278.
- Burg, J. van der (2009) Zien zonder te luisteren. Kan eyetracking gebruikt worden zonder de hardopdenkmethode bij probleemopsporend usability onderzoek naar websites? Masterscriptie. Universiteit Utrecht: Utrecht.
- Chen, M., J. Anderson & M. Sohn (2001) *What can a mouse cursor tell us more? Correlation of eye/mouse movement on web browsing*. Proceedings of the Conference on Human Factors in Computing Systems, 31 maart – 5 april 2001 (281-282), Seattle, Washington.
- Cooke, L. (2005) Eye tracking: how it works and how it relates to usability. *Technical Communication* 52 (4), 456-463.
- Cooke, L. & E. Cuddihy (2005) *Using eye tracking to address limitations in think-aloud protocol*. Proceedings of the IEEE International Professional Communication Conference, juli 2005 (653-658), Limerick, Ierland.
- Ehmke, C. & S. Wilson (2007) Identifying web usability problems from eye-tracking data. In: L.J. Ball, M. A. Sasse, C. Sas, T.C. Ormerod, A. Dix & T. McEwan (eds) *People and Computers XXI HCI...but not as we no it 1*.
- Elling, S., L. Lentz & M. de Jong. Focus on Infocus: A tool for evaluating informational websites. Ongepubliceerd.
- Elling, S., L. Lentz, M. de Jong (2007). Website Evaluation Questionnaire: Development of a research-based tool for evaluating informational websites. In: Wimmer, M.A., H.J. Scholl & A. Grönlund (Eds.): *EGOV 2007, LNCS 4656* (pp.293-304). Berlin Heidelberg: Springer-Verlag.
- Guan, Z., S. Lee, E. Cuddihy & J. Ramey (2006) The validity of the stimulated retrospective thinking-aloud method as measured by eye Tracking. Proceedings of the Conference on Human Factors in Computing Systems, 22-27 april 2006 (1253-1262), Montréal, Quebec, Canada.
- Haak, M.J. van den, M.D.T de Jong & P.J. Schellens (2003) Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & Information Technology* 22 (5), 339-351.
- Haak, M.J. van den, M.D.T de Jong & P.J. Schellens (2004) Employing think-aloud protocols and constructive interaction to test the usability of online library catalogues: A methodological comparison. *Interacting with computers* 16, 1153-1170.
- Haak, M.J van den, M.D.T. de Jong & P.J. Schellens (2007) Evaluation of an informational web site: Three variants of the think-aloud method compared. *Technical Communication* 54 (1), 58-71.

Haak, M.J van den, M.D.T. de Jong & P.J. Schellens (2009) Evaluating municipal websites: A methodological comparison of three think-aloud variants. *Government Information Quarterly* 26, 193-202.

Haas, E. (2009) 'Ik moet even wennen aan die stippeltjes' Een onderzoek naar de toegevoegde waarde van het terugzien van de oogbewegingen op de verbalisaties bij retrospectief hardopdenken, uitgevoerd op een gemeentelijke website. Masterscriptie. Universiteit Utrecht: Utrecht.

Hertzum, M., K.D. Hansen & H.H.K. Andersen (2009) Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, 28 (2), 165 - 181.

Hornbæk, K. (2010) Dogmas in the assessment of usability evaluation methods. *Behaviour & Information Technology* 29 (1), 97-111.

Lentz, L., P. Mak & H. Pander Maat (2006) Oogbewegingsregistratie en gebruikersonderzoek. *Tijdschrift voor Taalbeheersing* 28 (3), 233-244.

Manhartsberger, M. & N. Zellhofer (2005) Eye tracking in usability research: What users really see. In: Usability Symposium (2005) *Empowering software quality: How can usability engineering reach these goals?* OCG publication vol 198, 141-152.

Mueller, F. & A. Lockerd (2001) Cheese: Tracking mouse movement activity on websites, a tool for user modelling. Proceedings of the Conference on Human Factors in Computing Systems, 31 maart – 5 april 2001 (281-282), Seattle, Washington.

Nielsen, J., T. Clemmensen & C. Yssing (2002) *Getting access to what goes on in people's heads? - Reflections on the think-aloud technique*. Proceedings of the second Nordic conference on Human-computer interaction, 19-23 oktober 2002 (101-110), Aarhus, Denemarken. New York: ACM.

Rodden, K., X. Fu, A. Aula & I. Spiro (2008) *Eye-mouse coordination patterns on web search results pages*. Proceedings of CHI 2008, 5-10 april 2008 (2997-3002) Florence, Italië.

Vereijken, E. (2008) 'Ik zie, ik zie wat jij (niet) ziet'. Een verkenning van de mogelijkheden van eyetracking voor probleemopsporend usability onderzoek naar websites. Masterscriptie. Universiteit Utrecht: Utrecht.

Williams, T.R., C. Mulligan, K. Koprowicz, J. Miller, C. Reimann & d. Wang (2005) Does isolating a visual element call attention to it? Results of an Eye-tracking investigation of the effects of isolation on emphasis. *Technical Communication* 52 (1), 21-26.