

TEKNILLINEN KORKEAKOULU  
Informaatio- ja luonnontieteiden tiedekunta  
Tietotekniikan koulutusohjelma

Marko Saastamoinen

## **ARVIOIJAN KOKEMUKSEN VAIKUTUS KÄYTETTÄVYYDEN ASiantuntija-arviointiin**

Diplomityö

Helsinki, 31.7.2008

Valvoja: Prof. Marko Nieminen, TkT

Ohjaaja: Hannu Kuoppala, FM

TEKNILLINEN KORKEAKOULU Informaatio- ja luonnontieteiden tiedekunta Tietotekniikan koulutusohjelma		DIPLOMITYÖN TIIVISTELMÄ	
Tekijä Marko Saastamoinen		Päiväys 31.7.2008	
		Sivumäärä 98+8	
Työn nimi Arvioijan kokemuksen vaikutus käytettävyyden asiantuntija-arviointiin			
Professuuri Käytettävyys ja käyttöliittymät		Koodi T-121	
Työn valvoja Prof. Marko Nieminen, TkT			
Työn ohjaaja Hannu Kuoppala, FM			
<p>Käytettävyyden asiantuntija-arvioinnissa tutkittavasta tuotteesta etsitään mahdollisia käytettävyysongelmia tuotteen käytettävyyden parantamiseksi. Arvioijan ei tarvitse olla varsinainen asiantuntija, vaan arvioinnin voi käytännössä suorittaa kuka tahansa arviointimenetelmiin tutustuttuaan.</p> <p>Kokemuksella voidaan kuitenkin kiistatta ajatella olevan vaikutusta arviointiin ja sen tuloksiin. Psykologian alalla asiantuntijuuden tutkimus on paljastanut kokeneiden yltävän lukuisilla aloilla kokemattomia parempiin tuloksiin alakohtaisissa erityistehtävissä. Käytettävyyden arvioinnissa kokeneiden on jo aiemmin todettu löytävän kokemattomia enemmän käytettävyysongelmia. Kokemuksen vaikutusta arviointiin on kuitenkin tutkittu varsin vähän muista näkökulmista. Tässä tutkimuksessa kiinnostus kohdistuikin absoluuttisen määrän ohella myös erilaisten ongelma-kokonaisuuksien tutkimiseen sekä arvioinnin tutkimiseen laadullisesti.</p> <p>Kokeneiden ja kokemattomien arvioijien eroja selvitettiin tutkimuksessa empiiriseen vertailevan kokeen avulla. Kuusi kokenutta ja kuusi kokemattonta arvioijaa suorittivat kokeessa asiantuntija-arvioinnin samoille tuotteille. Arvioijat kirjasiivat tekemänsä havainnot ylös itse. Arviointiprosessien ja pohdinnan selvittämiseen käytettiin ääneen ajattelua. Lisäksi arvioijien silmänliikkeet tallennettiin katseenseurantalaitteen avulla. Arviointien tuloksia sekä arviointiprosesseja ja -tapoja analysoitiin määrällisesti ja laadullisesti.</p> <p>Tutkimuksen perusteella kokeneet arvioijat löysivät enemmän käytettävyysongelmia, mutta he myös pääsivät arvioinnissaan kokemattomia syvemmälle. He löysivät enemmän vakavia ongelmia, abstraktiotasoltaan laajempia ongelmakokonaisuuksia sekä esimerkiksi tuotteen konseptia koskevia ongelmia. Arvioijille kertynyt kokemus toimi luotettavana pohjana ongelmien tunnistamiselle. Lisäksi kokeneet arvioijat ottivat paremmin huomioon käytännön tarpeet ongelmien ja korjausehdotusten esittämisessä. Toisaalta yksilöllisten erojen osuus oli ryhmästä riippumatta suuri, mikä puolustaa usean arvioijan käyttämistä arvioinnissa. Lisäksi tutkimuksessa nousi esiin tarve käytettävyysongelman käsitteen määrittelylle arvioijien suoritusten sekä ongelmien vertailun helpottamiseksi.</p>			
Avainsanat käytettävyys, käytettävyyden arviointi, asiantuntija-arviointi, asiantuntijuus			

HELSINKI UNIVERSITY OF TECHNOLOGY Faculty of Information and Natural Sciences Degree Programme in Computer Science and Engineering		ABSTRACT OF MASTER'S THESIS	
Author Marko Saastamoinen	Date July 31, 2008	Pages 98+8	
Title of thesis The Effect of Evaluator Expertise in Usability Expert Evaluation			
Professorship Usability and User Interfaces		Professorship Code T-121	
Supervisor Professor Marko Nieminen, D.Sc.(Tech.)			
Instructor Hannu Kuoppala, M.Sc.			
<p>Expert evaluation is a method to find possible usability problems for developing the usability of a product. The evaluator is not required to be an actual expert of usability. Instead, the evaluation can be done by anyone who is familiar with the evaluation methods.</p> <p>However, the role of expertise can be considered significant in evaluation results. In psychology, the research of expertise has showed experts to be superior in domain-specific tasks in various domains. Usability experts have been found to be better in finding more usability problems than novices. Still, there have been only few approaches to examine other aspects of the effect of expertise. In addition to the number of usability problems, in this thesis, the emphasis was on analyzing different types and levels of usability problems, as well as the qualitative analysis of evaluating a product.</p> <p>An empirical study was arranged to find out the differences in evaluations made by experienced and non-experienced evaluators. Six experienced and six non-experienced evaluators conducted an expert evaluation to given four products. They wrote down their findings themselves during the evaluations. Thinking-aloud method was used to gather information about the evaluation and thinking processes. In addition, to examine the visual perception, an eye-movement study was conducted. Both qualitative and quantitative analyses were used to analyze the evaluation results and the other data.</p> <p>The study showed that the experienced evaluators not only found more usability problems, but they also achieved a more thorough and deeper approach. They found more severe problems, their findings concerned more often larger entities and they found many problems concerning the product concept when the non-experienced evaluators did not. The experienced evaluators used their experience as a basis of their evaluations. In addition, the experienced evaluators formed their findings based on the practical needs. On the other hand, there were significant individual differences in the evaluations. Thus, using multiple evaluators in expert evaluation is recommended. Furthermore, for achieving a valid comparison of usability problems the concept of usability problem needs to be defined.</p>			
Keywords usability, usability evaluation, expert evaluation, expertise			

# SISÄLTÖ

<b>1</b>	<b>JOHDANTO .....</b>	<b>1</b>
1.1	Työn taustaa .....	1
1.2	Tavoitteet .....	2
1.3	Tutkimuksen käytännöllinen ja teoreettinen merkitys.....	3
1.4	Tutkimuskysymykset.....	3
1.4.1	Miten kokemus vaikuttaa arvioinnin tuloksiin? .....	3
1.4.2	Miten kokemus vaikuttaa arviointiprosessiin? .....	4
1.5	Tutkimusmenetelmä .....	4
<b>2</b>	<b>KÄYTETTÄVYYDEN ASiantuntija-ARVIOINTI .....</b>	<b>5</b>
2.1	Käytettävyys .....	5
2.1.1	Käytettävyyden määritelmä .....	5
2.1.2	Käytettävyyden arviointi .....	5
2.2	Käytettävyyden arviointi ilman käyttäjiä .....	6
2.2.1	Heuristinen arviointi .....	6
2.2.2	Kognitiivinen läpikäynti.....	8
2.2.3	Muita menetelmiä ja muunnelmia .....	9
2.3	Asiantuntija-arviointi .....	10
2.3.1	Arvioinnin soveltaminen käytännön arviointiprojektiin .....	11
2.3.2	Käytännön työn asettamat rajoitukset .....	12
2.4	Arvioijan vaikutus arviointiin .....	13
2.4.1	Kokemus ja ongelmien määrä .....	14
2.4.2	Asiantuntija-arvioinnin ennustavuus .....	15
2.4.3	Yksilölliset erot.....	15
<b>3</b>	<b>KOKEMUKSEN KEHITTYMINEN .....</b>	<b>18</b>
3.1	Asiantuntijuus ( <i>expertise</i> ).....	18
3.2	Oppiminen ja taidon kehittyminen.....	19
3.3	Tiedonkäsittelyn rajoitukset.....	20
3.4	Havaitseminen ja tunnistaminen.....	23
<b>4</b>	<b>KOKENEIDEN JA KOKEMATTOMIEN ARVIOIJIEN VERTAILU .....</b>	<b>25</b>
4.1	Tutkimuksen esittely .....	25
4.2	Tutkimukseen osallistuneet arvioijat.....	25
4.3	Koeasetelma .....	27
4.3.1	Arvioinnin kohde.....	27
4.3.2	Arviointipaikka .....	32
4.3.3	Tutkimuksessa kerätty tieto .....	33
4.4	Kerätyn tiedon analysointitavat .....	38
4.4.1	Ääneen ajattelu ja arvioijien antamat kuvaukset.....	38
4.4.2	Arvioinnissa kirjattujen havaintojen tarkastelu .....	40

4.4.3	Silmänliiketutkimuksen data .....	42
<b>5</b>	<b>VERTAILUN TULOKSET .....</b>	<b>45</b>
5.1	Arvioinnissa tehdyt havainnot .....	45
5.1.1	Kirjattujen havaintojen absoluuttinen määrä .....	45
5.1.2	Havaintojen vakavuusluokitus .....	47
5.1.3	Arvioinnin osa-alueet ja näkökulmat .....	50
5.1.4	Havaintojen abstraktiotaso .....	54
5.1.5	Arvioidun osien paremmuuden vertailu .....	57
5.1.6	Korjausehdotukset .....	58
5.2	Arviointitavat ja arvioinnin eteneminen .....	60
5.2.1	Eteneminen tehtävittäin .....	60
5.2.2	Arvioinnin etenemistavat .....	62
5.2.3	Havaintojen tekeminen .....	65
5.2.4	Arviointia ohjanneita asioita ja piirteitä .....	67
5.2.5	Arvioijien oma kuvaus arvioinnista .....	71
5.3	Silmänliiketutkimuksen tulokset .....	73
5.3.1	Ensimmäinen tehtävä .....	74
5.3.2	Toinen tehtävä .....	78
<b>6</b>	<b>JOHTOPÄÄTÖKSET JA POHDINTA .....</b>	<b>83</b>
6.1	Arvioinnin tulokset .....	83
6.1.1	Arvioinnin syvyys .....	83
6.1.2	Yksilölliset erot .....	85
6.1.3	Käytettävyysongelman määrittely .....	86
6.1.4	Käytännön tarpeet tulosten esittämisessä .....	88
6.2	Arviointiprosessi .....	89
6.2.1	Silmäily vai järjestelmällinen läpikäynti? .....	89
6.2.2	Kokemuksen tuoma pohja .....	90
6.2.3	Arviointiin käytettävä aika .....	93
<b>LÄHTEET .....</b>	<b>95</b>	

## LIITTEET

Liite A: Silmänliiketutkimuksen tulokset lämpökarttoina

# 1 JOHDANTO

---

## 1.1 Työn taustaa

Käytettävyyden asiantuntija-arviointi tarkoittaa menetelmää, jolla tutkittavasta tuotteesta etsitään mahdollisia käytettävyysoongelmia ilman todellisten käyttäjien osallistumista. Asiantuntija-arvioinnissa arvioija käy läpi tutkittavan tuotteen käyttäen apunaan erilaisia suunnittelusääntöjä, tarkistuslistoja, tietoa ihmisen toiminnasta sekä omaa kokemustaan toimivista tai puutteellisista käyttöliittymäratkaisuksista.

Asiantuntija-arviointi perustuu usein tarkempiin menetelmiin, joista tunnetuimpia ovat heuristinen arviointi (Nielsen & Molich 1990) ja kognitiivinen läpikäynti (Lewis, Polson, Wharton & Rieman 1990). Nämä kaksi ovat nykyäänkin käytetyimpiä ja tutkituimpia arviointimenetelmiä (Hollinsed & Novick 2007). Etenkin heuristiseen arviointiin keskittyy myös alan opetus eri oppilaitoksissa, koska se on kevyt ja helppo oppia (Riihiahho 2000). Heuristinen arviointi luotiinkin alun perin menetelmäksi, jota voisi soveltaa tiettyjen perustietojen oppimisen jälkeen arviointikokemuksesta riippumatta (Nielsen & Molich 1990).

Kokemuksen vaikutusta asiantuntija-arviointiin jossain muodossa voidaan kuitenkin pitää itsestään selvänä. Kokemuksen vaikutusta ja asiantuntijuuden kehittymistä on psykologian tutkimuksessa tarkasteltu hyvin laajasti eri aloilla, esimerkiksi shakissa, lääketieteessä, fysiikassa ja musiikissa. Tutkimuksissa on tunnistettu selkeitä tekijöitä, jotka nostavat asiantuntijoiden suoritukset selvästi kokemattomia paremmiksi millä tahansa alalla (Ericsson & Smith 1991). Käytettävyyden arvioinnin suorittamisesta on todettu, että kokemuksen karttuessa erilaiset sääntölistat, muut tukiaineistot tai tarkemmat menetelmät sulautuvat asiantuntijan työhön niin, ettei niitä enää näkyvästi käytetä, vaan arvioinnit perustuvat arvioijien omaan kokemukseen (Jeffries, Miller, Wharton & Uyeda 1991; Riihiahho 2000). Arvioijien kokemus on muodostunut heidän aiemmin suorittamistaan arvioinneista sekä heidän ohjaamistaan käytettävyytsteesteistä. Lisäksi heuristinen arviointi on osoittautunut hyvin tehokkaaksi varsinaisten käytettävyydasiantuntijoiden käytössä, vaikka sitä ei aluksi kehitettykään menetelmäksi heidän käyttöönsä (mm. Desurvire, Lawrence & Atwood 1991; Jeffries ym. 1991).

Jo varsin pian heuristisen arvioinnin kehittämisen (Nielsen & Molich 1990) jälkeen useissa tutkimuksissa kokeneiden arvioijien todettiin löytävän määrällisesti kokemattomia enemmän käytettävyysoongelmia (mm. Nielsen 1992; Desurvire, Konziela & Atwood 1992). Heidän myös havaittiin pystyvän ennustamaan käyttäjien kohtaamia ongelmatilanteita kokemattomia paremmin (Desurvire ym. 1991). Lisäksi yksittäisten arvioijien löytämien käytettävyysongelmien määrissä on kokemuksesta riippumatta todettu olevan suuriakin eroja (mm. Virzi 1997; Jacobsen, Hertzum & John 1998; Slavkovic & Cross 1999; Hertzum & Jacobsen 2003).

Edellä mainituissa tutkimuksissa eroja on kuitenkin tutkittu erityisesti löydettyjen käytettävyysongelmien absoluuttisen määrän avulla. Ongelmien absoluuttinen määrä on kuitenkin vain yksi näkökulma arvioinnin laatuun ja kattavuuteen. Se jättää huomioimatta havaintojen muotoilemisen näkökulmia; esimerkiksi millaisia ongelmakokonaisuuksia arvioijat muodostavat, miten ongelmien abstraktiotasot poikkeavat toisistaan ja miten havaitut käytettävyysongelmat muodostetaan tai esitetään niiden korjaamisen kannalta.

Ongelmien luokitteluun on pyritty löytämään erilaisia tapoja (mm. Keenan, Hartson, Kafura & Schulman 1999; Hvannberg & Law 2003; Chattrachart & Lindgaard 2003). Niissä lähtökohta

on kuitenkin arvioinnin tulosten parantaminen käytännön kannalta, ei niinkään eri arvioijien suoritusten vertaaminen keskenään. Lisäksi ne eivät tarjoa keinoja ongelmien vaihtelevien abstraktiotasojen pohdintaan.

Tälle tutkimukselle kiinnostuksen loi toisaalta edellä mainittujen tutkimusten rajoittuminen tiettyihin näkökulmiin, toisaalta halu selvittää kokemuksen vaikutusta arviointiin myös laadullisesta näkökulmasta, erityisesti arviointitapojen kautta. Tässä tutkimuksessa lähtökohdaksi otettiin kokeneiden ja kokemattomien arvioijien empiirinen vertailu. Aiemman tieteellisen tutkimuksen tarkastelu jakaantuu tutkimusaiheen kannalta kahteen keskeiseen teemaan: käytettävyyden arviointiin ja siihen liittyviin tutkimuksiin (luku 2) sekä kokemuksen ja asiantuntijuuden kehittymiseen, jota käsitellään psykologian ja kognitiotieteen kautta (luku 3).

## 1.2 Tavoitteet

Tämän tutkimuksen päämääränä oli selvittää kokemuksen vaikutusta asiantuntija-arvioinnin tekemiseen prosessina sekä arvioinnista saataviin tuloksiin. Toisin sanoen tavoitteena oli tutkia, miten kokeneiden arvioijien suorittama arviointi eroaa kokemattomien arvioinnista ja millä tavoilla kokeneiden arvioijien asiantuntemus ilmenee. Pohjimmiltaan kun kyseessä on kuitenkin sama menetelmä; aiheuttaako kokemus muutoksen sen soveltamiseen, vai ilmeneekö kokemus jotenkin muuten? <sup>1</sup>

Kokeneiden ja kokemattomien arvioijien löytämien käytettävyysongelmien absoluuttisten määrien vertailun (mm. Nielsen 1992; Desurvire ym. 1992) tueksi tässä tutkimuksessa oli tarkoituksena tutkia löydettyjä käytettävyysongelmia myös muista eri näkökulmista, jotka voisivat paljastaa eroja arvioijien välillä. Tavoitteena oli etenkin tarkastella, miten löydetty ongelmat jakaantuvat eri osa-alueille sekä millä abstraktiotasoilla ongelmia esitetään.

Hypoteesina oli, että löydettyjen ongelmien suuremman määrän lisäksi kokeneet arvioijat kykenevät hahmottamaan ja löytämään arvioitavasta palvelusta suurempia ongelmakokonaisuuksia, jotka vaativat korjauksia. Lisäksi kokeneet saattavat kiinnittää herkemmin huomiota myös esimerkiksi palvelun konseptiin, ei pelkästään käyttöliittymään. Toisaalta taas on mahdollista, että kokemattomat arvioijat pääsevät joissakin tilanteissa parempiin tuloksiin, sillä heidän on jatkuvasti arvioinnin aikana perusteltava itselleen löytämiään ongelmia kokeneita arvioijia tarkemmin.

Arviointiprosessin ja arviointitapojen tutkimisen taustalla oli hypoteesi, jonka mukaan kokeneiden arvioijien käyttämä menetelmä poikkeaisi kokemattomien arviointimenetelmästä ja toisaalta myös tarkemmista opetettavista menetelmistä. Kokeneet arvioijat tekevät arviointeja jatkuvasti työkseen (sekä järjestävät käytettävyystestejä), joten arviointi saattaa perustua enemmän heidän omaan kokemukseensa kuin tarkempiin menetelmiin ja tekniikoihin (Jeffries ym. 1991; Riihiahho 2000).

Lisäksi yhtenä tutkimuksen hypoteesina oli, että arvioijille kertynyt kokemus auttaisi heitä kohti objektiivista ja kattavaa arviointia. Käytettävyyden arviointia työkseen tai muuten paljon tekevien ihmisten asiantuntemus arvioinnista kasvaa, mutta aivan selvää ei ole se, miten se todellisuudessa vaikuttaa arviointiin. On todettu, että eri arvioijat kiinnittävät erilaisista taustoistaan ja muista eroista johtuen huomiota eri seikkoihin, mikä johtaa toisistaan poikkeaviin arviointituloksiin (mm. Virzi 1997; Jacobsen ym. 1998; Slavkovic &

---

<sup>1</sup> Toisaalta kiinnostavaa on myös se, miten ja missä vaiheessa muutos kokemattomasta arvioijasta asiantuntijaksi tapahtuu. Tämän näkökulman selvittäminen jää kuitenkin tämän tutkimuksen ulkopuolelle.

Cross 1999; Hertzum & Jacobsen 2003). Toisaalta vastahypoteesina voisi myös ajatella, että aikaisempi kokemus voisi myös tehdä arvioinnista mekaanisempaa ja vähemmän luovaa, mikä heikentäisi arvioinnin laatua.

### 1.3 Tutkimuksen käytännöllinen ja teoreettinen merkitys

Tutkimuksella on merkitystä sekä käytännön arviointityön kehittämisen että menetelmien tutkimisen kannalta. Käytännön työn kannalta tutkimuksen avulla voidaan kehittää arviointimenetelmiä parhaiten käytännön työn tilanteisiin ja tarpeisiin soveltuviksi. Lisäksi tutkimuksen avulla voidaan tarkastella, missä asioissa kokeneilla arvioijilla on mahdollisesti puutteita, joihin voitaisiin reagoida esimerkiksi käytännön menetelmien ja tekniikoiden tarkentamisen tai lisäkoulutuksen avulla.

Menetelmien osalta myös opetuksen kehittämistä voidaan pitää motiivina tutkimukselle. Jos tutkimuksen avulla havaitaan konkreettisia asioita, joissa kokemattomilla on selvästi puutteita kokemattomiin nähden, voidaan näihin asioihin paneutua tarkemmin käytettävyyden arvioinnin opetuksen yhteydessä. Lisäksi tutkimuksen avulla voidaan opetuksessa etsiä keinoja tukea siirtymistä aloittelevasta arvioijasta kohti asiantuntijuutta.

Teoreettiselta kannalta merkittävää on myös silmänliiketutkimuksen hyödyntäminen asiantuntija-arvioinnin yhteydessä, mitä ei juuri ole ennen tehty. Silmänliiketutkimusta on käytetty käyttäjätestaukseen liittyvissä tutkimuksissa, mutta asiantuntijoiden työskentelyä ei silmänliiketutkimuksen avulla ole liiemmin selvitetty. Tässä suhteessa tutkimus voi antaa täysin uutta tietoa arvioinnista tai arvioijien toiminnasta ongelmien havaitsemisessa ja tunnistamisessa.

### 1.4 Tutkimuskysymykset

Tässä luvussa esitellään työtä varten luodut tutkimuskysymykset. Pääkysymyksiä on kaksi, jotka jakautuvat tarkempiin alakysymyksiin.

#### 1.4.1 Miten kokemus vaikuttaa arvioinnin tuloksiin?

Kokemuksen vaikutusta asiantuntija-arvioinnin tuloksiin on tutkittu aiemmin, mutta lähinnä määrällisesti. Kokeneiden ja kokemattomien arvioijien suorituksia on usein verrattu siten, että mittarina on käytetty löydettyjen ongelmien määrää (mm. Nielsen 1992; Desurvire ym. 1992) tai ennustavuutta käyttäjien kohtaamiin ongelmiin verrattuna (Desurvire ym. 1991). Laadullinen näkökulma tuloksiin on jäänyt vähäiseksi. Laadullista on tosin havaittujen ongelmien vakavuusasteiden tutkiskelu: on havaittu, että vakavat käytettävyysongelmat löytyvät arvioinnissa lievempiä todennäköisimmin (Nielsen 1992; Jacobsen ym. 1998).

Tämän tutkimuksen kannalta keskeinen näkökulma arvioinnin tuloksiin on se, minkälaisiin ja -kokoiisiin asioihin eri arvioijat kiinnittävät huomiota sekä miten havaintojen vakavuuden määrittely ja korjausehdotusten löytäminen näkyvät tuloksissa kokemuksesta riippuen.

Asiantuntija-arvioinnin tuloksiin liittyvät tarkemmat tutkimuskysymykset ovat seuraavat:

- Mikä tekee kokeneemmasta arvioijasta arvioinnin laadun kannalta asiantuntijan?
- Millaisia ongelmakokonaisuuksia kokenut arvioija nostaa esiin kokemattomaan verrattuna?
- Mitä eroja ongelmien havaitsemisessa ja tunnistamisessa on eritasoisten arvioijien välillä?



## 1.4.2 Miten kokemus vaikuttaa arviointiprosessiin?

Työhypoteesina tutkimuksessa on ollut ajatus, että kokemattomat arvioijat noudattavat oppimiaan yksityiskohtaisia arviointimenetelmiä vielä melko tarkasti, kun taas kokeneemmat asiantuntijat suorittavat arvioinnin soveltaen menetelmiä melko vapaasti omien kokemuksiensa tukena (Jeffries ym. 1991; Riihiaho 2000).

Eroja voi ajatella löytyvän siitä, edetäänkö arvioinnissa selkeästi järjestelmällisesti osa-alue kerrallaan, vai enemmän tilanteen, tuotteen tai havaitsemisen ohjaamana. Konkreettisesti kiinnostavaa on se, edetäänkö arviointityössä arvioimalla järjestelmää kokonaisuutena vai keskittyen yhteen tilaan tai näyttöön kerrallaan. Tämä voi johtaa eroihin myös itse arviointimenetelmän soveltamisessa.

Seuraavat tutkimuskysymykset liittyvät arvioinnin etenemiseen, tapaan ja prosessiin:

- Miten asiantuntija-arviointi käytännössä tehdään?
- Miten arvioinnin prosessi eroaa kokemukseltaan eritasoisten arvioijien kesken?
- Mitkä ovat todellisuudessa käytettyjä menetelmiä?
- Millaisia eroja ilmenee arvioijien katseen etenemisessä arvioinnin aikana?

## 1.5 Tutkimusmenetelmä

Tutkimuskysymysten vastausten löytämiseksi tutkimuksessa käytettiin menetelmänä empiiristä vertailua kokeneiden ja kokemattomien arvioijien välillä. Tutkimukseen etsittiin kuusi kokenutta sekä kuusi kokemattomaa arvioijaa. Lisäksi tutkimuksen pilotointiin otettiin mukaan yksi arvioija kummastakin ryhmästä. Kaikki arvioijat suorittivat laboratorio-olosuhteissa arvioinnin samalle kokonaisuudelle. Koeasetelmassa arviointi jaettiin neljään erilliseen suppeaan tehtävään analysoinnin ja vertailun helpottamiseksi. Kaikissa tehtävissä arvioinnin kohteena oli olemassa olevista verkkopalveluista irrotettuja www-sivuja.

Kokeneiden ja kokemattomien arvioijien empiirinen vertailu esitellään tarkemmin luvussa 4.

## 2 KÄYTETTÄVYYDEN ASiantuntija-arviointi

---

### 2.1 Käytettävyys

Käytettävyydellä tarkoitetaan yleisesti tuotteen soveltuvuutta käyttäjille sen tarkoituksenmukaiseen käyttöön. Tutkimusalana käytettävyys perustuu ergonomiaan, jonka tutkimus sai vauhtia sotilasteollisuudesta etenkin toisesta maailmansodasta alkaen. Käytettävyys itsessään tieteellisenä tutkimusalana on vielä verrattain nuori; se on noussut esiin lähinnä tietotekniikan kehityksen myötä viime vuosikymmeninä (Nielsen 1993). Käytettävyuden arviointimenetelmien kehittäminen ja tutkimus käynnistyi laajalti 1990-luvun alkaessa (Hollinsed & Novick 2007).

#### 2.1.1 Käytettävyuden määritelmä

Käytettävyydelle on useita määritelmiä, joista tunnetuimmat ovat kansainvälisen standardointiorganisaatio ISO:n (ISO 9241-11) sekä Jacob Nielsenin määritelmät (Nielsen 1993). ISO:n määritelmän mukaan käytettävyys tarkoittaa ”kuinka tietyt käyttäjät voivat käyttää tuotetta määriteltyjen tavoitteiden saavuttamiseen tuottavasti, tehokkaasti ja miellyttävästi tietyssä käyttöympäristössä”. Määritelmän perusteella käytettävyys jakaantuu kolmeen attribuuttiin:

- *Tuottavuus*: kuinka kattavasti käyttäjät saavuttavat tavoitteensa tuotteella.
- *Tehokkuus*: kuinka paljon resursseja käyttäjiltä vaaditaan suhteessa tavoitteiden saavuttamisen kattavuuteen.
- *Miellyttävyys*: miten positiivisesti käyttäjät suhtautuvat tuotteeseen.

Nielsen määrittelee käytettävyuden yhdeksi tuotteen hyväksyttävyyden ja tarkemmin hyödyllisyyden komponenteista. Nielsenin määritelmän perustana puolestaan on viisi käytettävyysominaisuutta:

- *Opittavuus*: kuinka nopeasti käyttäjät oppivat käyttämään tuotetta.
- *Tehokkuus*: kuinka tehokkaasti tuotetta voi käyttää sen oppimisen jälkeen.
- *Muistettavuus*: kuinka helppoa tuotteen käyttö on muistaa, jos sitä käytetään harvoin tai ajoittain.
- *Virheettömyys*: kuinka hyvin käyttöön liittyvien virheiden välttäminen ja niistä toipuminen onnistuu.
- *Miellyttävyys*: kuinka tyytyväisiä käyttäjät ovat tuotteen käyttöön.

Edellä mainituilla määritelmillä on se merkittävä ero, että Nielsen jättää käytettävyuden ulkopuolelle tuotteen konseptuaalisen puolen, toisin sanoen sen, mitä tuotteella voi tehdä ja vastaako se käyttäjien tarpeita. Määritelmän sisälle jää siis se, kuinka hyvin tuotteiden käyttäminen onnistuu. ISO:n määritelmä puolestaan sisältää myös konseptuaalisen näkökulman (ts. tuottavuus). Määritelmien eroavaisuus on tämän tutkimuksen kannalta merkittävä siinä mielessä, että käytettävyuden arvioinnin tulokset saattavat oletettavasti riippua siitä, miten käytettävyys käsitteenä tulkitaan.

#### 2.1.2 Käytettävyuden arviointi

Käytettävyuden arvioinnin tavoitteena on parantaa tuotteen käytettävyyttä eliminoimalla tuotteesta ja erityisesti sen käyttöliittymästä havaitut ongelmakohdat tai puutteet, jotka voivat aiheuttaa eritasoisia hankaluuksia todellisessa käyttötilanteessa tuotteen kohderyhmää edustaville käyttäjille. Tästä johtuen käytettävyuden arviointimenetelmät ovat tyypillisesti ongelmakeskeisiä; ne perustuvat ongelmien etsimiseen ja havaitsemiseen.

Tuotteesta löydettyjä käytettävyyso ongelmia ei kuitenkaan useimmiten tule pitää arvioinnin lopputuloksena, vaan tärkeää on myös löytää ratkaisut löydetyille ongelmille (Sawyer, Flanders & Wixon 1996; Karat 1997). Arviointia tulisi tehdä eri vaiheissa läpi tuotteen kehityksen ja elinkaaren, ja se tulisi aloittaa mahdollisimman aikaisin, jolloin tuotteeseen tehtävät muutokset ovat halvempia ja helpompia tehdä (ISO 13407).

Käytettävyyden arviointi voidaan jakaa käyttäjien kanssa sekä ilman käyttäjiä sovellettaviin arviointimenetelmiin (Riihiaho 2000). Tyypillisessä käyttäjien kanssa tehtävässä arvioinnissa (käyttäjättestaus, mm. Riihiaho 2000) testataan tuotteen käyttöä todenmukaisilla testitehtävillä mahdollisimman todenkaltaisessa käyttötilanteessa. Tehtävät havainnot perustuvat käyttäjän toiminnan sekä heidän kertomuksensa seuraamiseen ja havainnoimiseen. Testien lisäksi käyttäjiä voidaan ottaa tuomaan käyttäjän näkökulmaa erilaisiin tuotteiden läpikäynteihin. Käyttäjien kanssa arviointi antaa usein parasta tietoa todellista käyttöä vastaavista ongelmakohtista ja parannusta vaativista asioista (Karat 1994), mutta toisaalta se vaatii myös resursseja, kuten asiantuntijoiden kokemusta ja aikaa, joita ei aina ole riittävästi käytettävissä (Nielsen & Molich 1990). Lisäksi testaus käyttäjien kanssa ei välttämättä paljasta kaikkia mahdollisia ongelmia, joten myös muita menetelmiä tarvitaan (Karat 1994).

Ilman käyttäjiä tehtävä arviointi perustuu puolestaan siihen, että arvioija tai joukko arvioijia analysoi tuotteen käytettävyyttä eri näkökulmista, jolloin arvioinnissa käytännössä ennustetaan todellisia käyttäjien kohtaamia ongelmia (Riihiaho 2000). Arviointiin kehitettyjä menetelmiä esitellään lyhyesti seuraavassa luvussa. Tarkempien menetelmien lisäksi arvioinnissa hyödynnetään usein tietämystä ihmisen kognitiivisista prosesseista ja toiminnasta, sekä arvioijan omaa kokemusta erilaisista tuotteista. Yleisesti puhutaankin asiantuntija-arvioinnista, joka voi kattaa erilaisia arviointitapoja ja tarkempia menetelmiä.

## 2.2 Käytettävyyden arviointi ilman käyttäjiä

Tavallisimmat ilman käyttäjiä sovellettavat menetelmät, joihin alan tutkimus on painottunut, ovat heuristinen arviointi ja kognitiivinen läpikäynti. Nämä ovat myös nykyisin alalla sovelletuimpia menetelmiä (Hollinsed & Novick 2007). Ilman käyttäjiä sovellettavat arviointimenetelmät eivät tyypillisesti vaadi pitkää valmistelua tai kokemusta, ja ne ovat kevyitä soveltaa ja oppia (Riihiaho 2000). Menetelmiä voi periaatteessa soveltaa kuka tahansa, ja ensimmäiset menetelmät kehitettiinkin tuotteiden suunnittelijoiden käyttöön (Molich & Nielsen 1990). Etenkin heuristinen arviointi on käyttäjien kanssa suoritettavaan testaukseen verrattuna halpa ja nopea menetelmä (Nielsen 1993).

Asiantuntija-arviointi (ks. 2.3) perustuu useimmiten juuri heuristiseen arviointiin, joka esitellään seuraavassa kappaleessa. Tässä luvussa esitellään myös tunnetuin läpikäynnin menetelmä: kognitiivinen läpikäynti. Lisäksi luvussa mainitaan lyhyesti joitakin muita arviointiin käytettäviä menetelmiä sekä edellä mainituista menetelmistä kehitettyjä muunnelmia.

### 2.2.1 Heuristinen arviointi

Heuristinen arviointi on yleisin menetelmä, johon käytettävyyden asiantuntija-arviointi usein yhdistetään (Nielsen 1993). Molich ja Nielsen (1990) kehittivät oman kokemuksensa sekä erilaisten suunnitteluohjeiden pohjalta suunnittelijoille listan suunnittelusääntöjä, joiden pohjalta he myöhemmin muodostivat heuristiset säännöt käyttöliittymien arviointia varten (Nielsen & Molich 1990).

Heuristinen arviointi perustuu käyttöliittymän näkymän, tilan tai esimerkiksi www-sivun analysointiin ennalta määriteltujen heurististen sääntöjen avulla. Tutkittavasta tuotteesta etsitään kohtia, jotka eivät noudata kyseisiä sääntöjä. Havainnot kirjataan ylös käytettävyyssongelmiksi. Tyypillisesti tutkittavaa tuotetta käydään läpi näkymä tai sivu kerrallaan, ja kunkin kohdalla pohditaan jokaisen heuristisen säännön toteutumista. Tunnetuimmat heuristiset säännöt ovat Nielsenin (1993) 10 heuristista sääntöä:

- Yksinkertainen ja luonnollinen dialogi
- Puhu käyttäjän kieltä
- Minimoi käyttäjän muistikuorma
- Yhtenäisyys
- Palaute
- Selkeät poistumistiet
- Oikopolut
- Hyvät virheilmoitukset
- Virheiden välttäminen
- Ohjeet ja dokumentaatio.

Arviointiin voidaan käyttää myös monia muita arviointisääntöjä, erilaisia suunnittelusääntöjä tai tarkastuslistoja, esimerkiksi Shneidermanin (1998) kahdeksan kultaista sääntöä sekä Smithin ja Mosierin (1986) 60 ohjetta arviointia varten. Lisäksi Nielsen on esitellyt erilaisia listoja eri tilanteisiin ja arvioitaviin tuotteisiin parhaiten soveltuviksi (Nielsen 1994a). Muiden muassa Karoulis ja Pombortsis (2000) ovat todenneet, että www-palvelujen tai -sivujen arviointi vaatii erityistä listaa arvioinnin suorittamiseksi. Esimerkiksi Instone (1997) on muokannut www-sivujen arviointiin soveltuvat heuristiset säännöt Nielsenin sääntöjen pohjalta.

Heuristinen arviointi on menetelmänä kevyt: sitä voi soveltaa nopeasti yksittäisille käyttöliittymän näkymille ja luonnoksille. Varsinaisia järjestelyjä ei tarvita, tuotteen lisäksi tarvitaan vain sovellettava säännöstö, jos arvioija ei sitä ole sisäistänyt. Lisäksi arvioinnin tekijän ei tarvitse olla käytettävyyden asiantuntija (Riihiaho 2000). Heuristinen arviointi suunnattiin alun perin arvioijille, joilla on perustiedot käytettävyyden periaatteista, mutta jotka eivät ole varsinaisia asiantuntijoita (Nielsen & Molich 1990). Tarkoituksena olikin tukea suunnittelijoita oman työn arvioinnissa. Jatkossa kuitenkin todettiin, että menetelmä on erityisen tehokas käytettävyyden asiantuntijoiden käyttämänä (Desurvire ym. 1991; Jeffries ym. 1991).

Heuristisen arvioinnin avulla tutkittavasta tuotteesta löydetään usein vakavimmat käytettävyyssongelmat todennäköisemmin kuin pienet, mutta pieniä ongelmia löytyy kuitenkin määrällisesti selvästi vakavia enemmän (Nielsen 1992). Se paljastaa parhaiten ongelmia tuotteen yksittäisten näkymien ja tilojen sisällä. Järjestelmän toimintojen etenemisen ja käyttäjän toiminnan tukemisen analysointiin se ei puolestaan erityisen hyvin sovellu, koska se ei tarjoa helppoa keinoa tarkastella todellisia käyttötilanteen tehtäviä. Esimerkiksi Sears (1997) toteaa tehtävillä olevan merkittävän roolin olennaisten ongelmien löytämisessä. Toisaalta hän toteaa, että tehtäviä noudattelevassa läpikäynnissä kokemattomat arvioijat hukkaavat helposti fokuksen käyttäjästä, jolloin taas heuristinen lista auttaa heitä keskittymään paremmin. Tämäkään ei kuitenkaan yksin auta, koska heuristiset säännöt eivät välttämättä ohjaa keskittymistä juuri käyttäjän toimintaan.

Suurin haaste arvioinnin laadun ja kattavuuden kannalta onkin juuri sääntölistan käyttäminen. Yleiset heuristiset listat saattavat olla hankalasti sovellettavia abstraktin tasonsa ja tulkinnanvaraisuutensa vuoksi. Esimerkiksi Gould ja Lewis (1983) totesivat, että suunnittelijat voivat käyttää samoja sääntöjä kuin arvioijat, mutta koska säännöt ovat

yleisellä tasolla, ne voivat johtaa harhaan. Konkreettisemmilla ja tarkemmilla säännöillä puolestaan on haasteena kattaa kaikki keskeisimmät näkökulmat pitämällä lista kuitenkin tiiviinä ja kerralla hahmotettavana (vrt. 60 ohjetta arviointiin, Smith & Mosier (1986)).

Lisäksi hyvään tulokseen päästäkseen arvioijan olisi sisäistettävä heurististen sääntöjen merkitys ja tarkoitus, jotta niiden avulla voisi tehdä relevantteja havaintoja. Arvioinnin aloittamisen helppoudesta huolimatta arvioijalle jää siis suuri haaste päästä mahdollisimman objektiiviseen ja syvälliseen tulokseen.

Usein käytännössä suoritettava arviointi pohjautuu tarkkojen heurististen sääntöjen sijaan arvioijan omaan kokemukseen, joka toimii arvioinnissa heurististen sääntöjen tapaan (Jeffries ym. 1991, Nielsen 1993). Toisin sanoen käytännön arvioinnissa arvioitavan tuotteen ominaisuuksia verrataan tällöin kokemukseen ja sen muodostamiin sääntöihin. Kokeneiden arvioijien työhön on saattanut yhdistyä myös läpikäynnin ominaisia piirteitä (Riihaho 2000), jolloin todellinen käytännön menetelmä on itse asiassa yhdistelmä monista yksittäisistä menetelmistä.

### 2.2.2 Kognitiivinen läpikäynti

Kognitiivinen läpikäynti (Lewis ym. 1990; Polson, Lewis, Rieman & Wharton 1992) pureutuu heuristista arviointia tarkemmin tuotteen todelliseen käyttöön ja sillä suoritettaviin tehtäviin, esimerkiksi tietyn toiminnallisuuden käyttämiseen tai navigoimiseen. Läpikäyntiä voi soveltaa erityisesti erilaisiin toimintosarjoihin, joiden eri vaiheissa (ts. näkymässä tai tilassa) arvioidaan toiminnallisuuden ymmärrettävyyttä sekä sitä, kuinka hyvin se ohjaa käyttäjää eteenpäin. Erityisen hyödyllinen se on oppimisen helppouden tarkastelussa, eli tuotteilla, joita käyttäjän pitäisi pystyä käyttämään saman tien ilman opettelua (Riihaho 2000).

Tarkasteltava tuotteen osa jaetaan usein osatehtäviin, joita tarkastellaan yksitellen. Jokaisessa tarkasteltavassa vaiheessa analysoidaan seuraavien kysymysten toteutuminen (Wharton, Rieman, Lewis & Polson 1994):

- Ymmärtääkö käyttäjä, että tietty osatehtävä johtaa hänen tavoitteeseensa?
- Löytääkö käyttäjä tarjolla olevan soveltuvan toiminnon?
- Yhdistääkö käyttäjä toiminnon kyseiseen osatehtävään?
- Näkeekö käyttäjä edistymisen kohti tavoitetta osatehtävän jälkeen?

Kognitiivisen läpikäynnin onnistumisen kannalta keskeistä on sen huolellinen valmistelu. Ennen läpikäyntiä on muun muassa määriteltävä, ketkä ovat tuotteen käyttäjiä, mitä tehtäviä tuotteella tyypillisesti tehdään, mitä näistä tehtävistä erityisesti tutkitaan ja mikä on oikea tai oletettu tapa suorittaa tehtävät (Wharton ym. 1994). Sopivien tehtävien valitsemisen tärkeyttä korostetaan erityisesti usein (mm. Polson ym. 1992; Sears 1997).

Kognitiivisen läpikäynnin avulla arvioinnissa saadaan nostettua esiin toiminnallisuuteen ja etenemiseen liittyviä ongelmakohtia, jotka todennäköisesti tulisivat vastaan todellisessa käyttötilanteessa. Lewis ym. (1990) totesivat tutkimuksessaan sen paljastavan puolet käyttäjien kohtaamista ongelmista. Hertzum ja Jacobsen, jotka ovat tutkineet laajasti arvioijien välisiä eroja eri menetelmillä, havaitsivat tutkimuksessaan (1999), että kognitiivisen läpikäynnin strukturoidusta prosessista johtuen arvioijien väliset erot ovat pienempiä ja havaintojen saturoituminen on vähäisempää kuin heuristista arviointia tehtäessä. Tästä huolimatta yksilöllinen hajonta arvioijien tulosten välillä on suuri. He näkivät kognitiivisen läpikäynnin ongelmiksi ankkuroitumisen sekä stereotypisoinnin: arvioija toisaalta toimii itse käyttäjänä arvioinnin aikana ja siten edustaa itse käyttäjää, toisaalta arvioija mieltää

todellisen käyttäjän liian suppeasti ja jättää huomioimatta käyttäjien diversiteetin. Hollinsed & Novick (2007) huomauttavat, että kognitiivinen läpikäynti sopii myös kokemattomille arvioijille, mutta arvioinnin onnistumiselle erityinen haaste on edustavien tehtävien määrittely.

Heuristiseen arviointiin verrattuna kognitiivinen läpikäynti on varsin työläs ja uuvuttava menetelmä, minkä vuoksi sitä todellisuudessa käytetään heuristista arviointia harvemmin (Riihiaho 2000). Lisäksi jo toiminnalliselle tuotteelle käytettävyydestä on usein parempi menetelmä. Vaikka todellisten tehtävien tarkastelun tärkeyttä arvioinnin tulosten kannalta on korostettu (Sears 1997), on myös todettu, että kokeneiden arvioijien soveltamana kognitiivinen läpikäynti ennustaa todellisia ongelmia heuristista arviointia heikommin, sillä heuristinen arviointi kuitenkin ohjaa analysoimaan useita eri näkökulmia ja ulottuvuuksia (Desurvire ym. 1992). Käyttöliittymän yksityiskohtiin tai toiminnallisuuden ulkopuolisiin asioihin liittyviä havaintoja menetelmällä ei johdonmukaisesti saadakaan kerättyä.

### 2.2.3 Muita menetelmiä ja muunnelmia

Heuristisen arvioinnin ja kognitiivisen läpikäynnin lisäksi arviointiin ilman käyttäjiä on sovellettu lukuisia muitakin menetelmiä, kuten esimerkiksi *GOMS* (Card, Moran & Newell 1983), jolla ennustetaan kokeneiden käyttäjien suoritusta, sekä *Guideline review* (mm. Jeffries ym. 1991; Shneiderman 1998), jolla tuotetta arvioidaan käyttöliittymän suunnittelua koskevien suositusten avulla. Shneiderman (1998) mainitsee myös yhtenäisyystarkistukset (*Consistency inspection*) ja formaalit käytettävyystarkistukset (*Formal usability inspection*).

Nämä muut menetelmät ovat kuitenkin harvinaisempia, ja asiantuntija-arviointi käytännössä perustuu heuristiseen arviointiin ja kognitiiviseen läpikäyntiin, joten kyseisiä menetelmiä ei tässä työssä tarkastella.

Heuristisen arvioinnin ja kognitiivisen läpikäynnin eroavaisuudet ja tietyt puutteet ovat saaneet käytettävyyden tutkijat hakemaan arvioinnin tueksi myös sellaisia menetelmiä, joiden avulla näiden molempien vahvuudet saataisiin huomioitua yhdessä arvioinnissa.

Tunnetuin yhdistelmä on heuristinen läpikäynti (*Heuristic walkthrough*) (Sears 1997). Heuristinen läpikäynti yhdistää hyvät puolet heuristisesta arvioinnista, kognitiivisesta läpikäynnistä sekä käytettävyysläpikäynnistä (Sears 1995). Strukturoituneisuudessaan menetelmä asettuu pohjana olleiden menetelmien välimaastoon, ja se on helppo oppia ja käyttää. Menetelmän tarkoituksena on tukea sekä heurististen sääntöjen ohjaamaa ongelmien poimimista että käyttäjän todellisten tehtävien tarkastelua läpikäynnin muodossa. Menetelmän suorittaminen jakautuu kahteen osaan:

- Osa 1: Tehtäviin perustuva arviointi arvioijalle luotujen tehtävien avulla vastaa tuotteen intuitiivisuuteen – mitä käyttäjä voi tehdä, löytyykö haluttu toiminto, osaako käyttäjä toimia, saako käyttäjä riittävän palautteen.
- Osa 2: Vapaamuotoinen arviointi vastaa heuristista arviointia, mutta sitä tuetaan todellisilla käyttötilanteilla ja tehtävillä, jotka on käyty läpi ensimmäisessä osassa.

Sears (1997) toteaa tutkimuksensa perusteella heuristisen läpikäynnin olevan kattavampi kuin kognitiivinen läpikäynti, mutta toisaalta tavoitteeseen keskittyneempi kuin heuristinen arviointi. Menetelmä on tämän tutkimuksen kannalta erityisen mielenkiintoinen siksi, että arvioijien on käytännössä todettu yhdistävän heuristiseen arviointiinsa läpikäynnin piirteitä (Riihiaho 2000). Erityisesti kokeneiden arvioijien suorittama strukturoimaton asiantuntija-arviointi voi siis mahdollisesti muistuttaa jollain tavalla heuristista läpikäyntiä. Sears ei

verrannut menetelmää kokeneiden ja kokemattomien arvioijien avulla, mutta oletti myös kokeneiden olevan tehokkaampia kyseisellä menetelmällä.

## 2.3 Asiantuntija-arviointi

Edellisissä kappaleissa kuvattiin ilman käyttäjiä tehtävää arviointia erilaisten kehitettyjen menetelmien valossa. Käytännön arviointityössä puhutaan kuitenkin usein yleisemmin asiantuntija-arvioinnista, minkä vuoksi tähän käsitteeseen pureudutaan tässä luvussa erikseen. Tuloksesta riippuen asiantuntija-arviointi voidaan toisaalta nähdä yläkäsitteenä erilaisille ilman käyttäjiä suoritettaville arviointimenetelmille, toisaalta taas se voidaan nimensä mukaisesti mieltää käytettävyyden asiantuntijan tekemäksi arvioinniksi.

Tässä työssä asiantuntija-arviointi määritellään edellä mainituista tulkinnoista lähinnä jälkimmäisen mukaan, kuitenkin rajautumatta varsinaisiin asiantuntijoihin. Toisin sanoen asiantuntija-arvioinnin voivat suorittaa kaikki käytettävyyttä tutkimusalana tuntevat, sitä opiskelleet tai parhaillaan opiskelevat henkilöt, joiden kokemus voi kuitenkin vaihdella selvästi. Asiantuntija-arvioinnin taustalla voi olla monia erilaisia arviointitapoja, tarkempia yksityiskohtaisia menetelmiä tai arvioijien omaa kokemusta. Puutteellinen kokemus arvioinnista ei estä sen suorittamista, sillä esimerkiksi heuristisen arvioinnin on todettu olevan nopeasti opittavissa ja sovellettavissa <sup>2</sup> (Riihiaho 2000). Yleisesti asiantuntija-arvioinnin tekijä voi olla käytettävyyden asiantuntija, mutta se ei ole välttämätöntä (Nielsen & Mack 1994). Toisaalta arviointi nähdään välillä myös niin riippuvaksi arvioijan kokemuksesta ja taidoista, että kokemattomien arvioijien arvioinnin tukemista erilaisilla apukeinoilla pidetään välttämättömänä (Karoulis & Pombortsis 2000). Tässä tutkimuksessa lähtökohtana arvioinnille pidetään sitä, että arvioijalla on tiedossaan vähintään perusteet käytettävyydestä ja sen arvioinnista.

Lähes poikkeuksetta käytettävyyttä opiskelevat ja sitä kautta käytettävyydelle päätyvät henkilöt lähtevät liikkeelle tietyistä tarkemmista menetelmistä (ks. 2.2). Käytännön työssä ja kokemuksen kertyessä ne kuitenkin usein sulautuvat osaksi arvioijalle muodostuvaa omaa arviointitapaa ja kokemusta, eikä arvioinnissa rajoituta yksin niihin (Jeffries ym. 1991; Riihiaho 2000). Asiantuntija-arviointi voikin sisältää erilaisissa suhteissa näitä menetelmiä, tietämystä ihmisen kognitiivisista prosesseista ja toiminnasta sekä yksinkertaisesti arvioijan omaa kokemusta. Jo heuristisen arvioinnin menetelmän esittelyn yhteydessä Nielsen ja Molich (1990) totesivat, että vaikka ideaalilanteessa tuotetta verrataan suunnittelusääntöihin, sääntöjen määrystä ja luonteesta johtuen useimmat arvioijat tekevät arvioinnin oman intuitionsa sekä terveen järjen avulla, toisin sanoen yksinkertaisesti katsomalla tuotetta. Lisäksi samalla kun Nielsen (1993) kirjassaan esittelee tunnetut heuristiset sääntönsä, hän toteaa, että arvioinnissa nojataan usein omaan kokemukseen.

Sääntöjen merkitystä arvioinnissa ei kuitenkaan voida pitää yhdentekevänä. Bastien ja Scapin (1995) vertailivat asiantuntijoiden tekemiä arviointeja ensin ilman sääntöjä ja sitten sääntöjen kanssa: puolet käytti toisella kerralla sääntöjä apunaan, mutta puolet teki sen edelleen vain kokemuksensa avulla. Tutkimuksessa asiantuntijoiden havaittiin hyötyvän säännöstöstä: sääntöjä käyttäneet arvioijat löysivät enemmän ongelmia kuin kokemuksensa luottaneet arvioijat. Lisäksi Sears (1997) totesi kokemattomien arvioijien niin ikään hyötyvän heuristisista säännöistä, mikä onkin varsin luonnollista. Asiantuntijoiden säännöistä saamaa hyötyä selittävät ainakin osittain arvioijien tuloksissa näkyvät yksilölliset erot (mm. Virzi

---

<sup>2</sup> Kokemuksella ja arvioijalla yleisemmin on toki iso vaikutus arvioinnin tuloksiin (ks. 2.4). Tuotteen parantamiseksi käytännössä arviointi voidaan kuitenkin aina nähdä hyödylliseksi, vaikka tulokset eivät olisikaan kattavia.

1997; Jacobsen ym. 1998; Hertzum & Jacobsen 2003), joita heuristinen lista jossain määrin tasoiittaa.

Asiantuntija-arviointi on joskus nähty opittavuutensa ja edullisuutensa vuoksi poissulkevana vaihtoehtona käytettävyydestille. Jeffries ja Desurvire (1992) painottavat, että vaikka asiantuntija-arviointi tukee käytettävyydestä menetelmänä hyvin, se ei voi korvata sitä tulosten näkökulmasta. Osa ongelmista löytyy vain testauksen avulla, ja heidän mukaansa päästäkseen edes lähelle testauksen tuloksia arvioijien olisi oltava asiantuntijoita. Lisäksi arvioijia pitäisi olla useita, sillä yksittäisen kokeneenkin arvioijan tekemä asiantuntija-arviointi on kattavuuden kannalta huonoin vaihtoehto. Asiantuntija-arvioinnin lähtökohtaisesti edullinen hinta myös nousee siitä, että menetelmä tuottaa paljon pieniä ongelmia sekä havaintoja, jotka eivät todellisuudessa edes ole ongelmia lainkaan.

Käytännön työ ja sen asettamat tarpeet muokkaavat asiantuntija-arviointia tarkoitukseen sopivaksi (Sawyer ym. 1996). Käytettävyyden arviointipalveluita tarjoavat yritykset myös usein tarjoavat palvelunaan juuri asiantuntija-arviointia erittelemättä sitä tarkemmin. Tämän käsitteen sisällä asiantuntija voi soveltaa aina tilanteeseen parhaiten soveltuvia keinoja ja tuottaa arvioinnin tilanneelle asiakkaalle tehokkaasti tärkeimmät tulokset: useimmiten asiakkaan kannalta merkittävimmät käytettävyysongelmat ja ehdotukset niiden korjaamiseksi.

### 2.3.1 Arvioinnin soveltaminen käytännön arviointiprojektiin

Asiantuntija-arviointia pidetään käytännön arviointiprojekteihin hyvin soveltuvana. Sitä voidaan soveltaa aikaisessa tuotekehitysvaiheessa, sen avulla löydetään paljon ongelmia – ja vakavia ongelmia. Lisäksi se on menetelmänä halpa. Näiden etujen haasteena on kuitenkin se, että laadukkaisiin tuloksiin päästäkseen se vaatii käytettävyyden asiantuntemusta ja useita arvioijia (Jeffries ym. 1991), joita ei aina ole saatavilla. Lisäksi laadukas arviointi vaatisi todellisuudessa aina kunnollisen perehtymisen tuotteeseen, vaikka arviointi sinänsä olisikin nopea tehdä (Shneiderman 1998). Muiden muassa Law ja Hvannberg (2004) kyseenalaistavat heuristisen arvioinnin tehokkuuden (käytettyjen resurssien suhde tuloksiin), ja ovat pyrkinet parantamaan tehokkuutta erilaisten strategioiden avulla.

Tieteelliseen tutkimukseen verrattuna käytännön arviointityötä koskevat erilaiset tarpeet. Vaikka käytettävyydetutkimus ja käytettävyyden arviointi perustuukin tieteelliseen tutkimukseen, käsitykseen ihmisen toiminnasta ja erilaisiin teorioihin, niin käytännön arviointiprojektissa usein keskeisin tavoite on yksinkertaisesti vain parantaa tuotetta löytämällä todelliset ongelmakohdat ja niitä korjaavat ratkaisuehdotukset. Tällöin käytännön arvioinnissa keskeiseen rooliin nousevat tehtyjen havaintojen jäsentely ja priorisointi, niiden muotoilu, esittäminen ja kommunikointi asiakkaalle ymmärrettävästi sekä toteuttamiskelpoisten parannusehdotusten laatiminen. Lisäksi käytännön arviointityössä saatetaan keskittyä parantamaan jotakin tutkittavan tuotteen osa-aluetta, jolloin osa mahdollisista havainnoista ei ole välttämättä asiakkaan kannalta merkityksellisiä.

Sawyer ym. (1996) ovat tutkineet useiden käytännön arviointiprojektien avulla asiantuntija-arvioinnin vaikuttavuutta, toisin sanoen sitä, kuinka paljon asiakas hyötyy arvioinnista todellisuudessa sekä kuinka paljon tuote todellisuudessa paranee arvioinnin myötä. He ovat löytäneet kuusi keskeistä tekijää asiantuntija-arvioinnin vaikuttavuuteen: vakavuusluokituksen laatiminen, tarkat parannusehdotukset, palautteen pyytäminen asiakkaalta, arviointi aikaisessa vaiheessa, tutustuminen tuotteeseen ja tuoteperheen kontekstin huomioiminen. Erityisesti he korostavat arvioinnin laadun kannalta tuotteeseen ja tuoteperheeseen tutustumisen merkitystä. Shneiderman (1998) korostaa sitä, että arvioijien



pitäisi päästä todellisen käyttötilanteen ja -ympäristön mukaiseen kontekstiin esimerkiksi kurssien, manuaalien ja tutoriaalien avulla. Perehtyminen on mielenkiintoista siksi, että monissa tieteellisissä tutkimuksissa tämä lähestymistapa on jäänyt hyvin vähäiseksi. Lisäksi Sawyer ym. (1996) nostavat esiin kahden arvioijan yhteistyön tehokkuuden käytännössä. He ovat todenneet yhteisarvioinnin olevan tehokas yhteisen kuvan muodostamisessa, kommunikoinnissa sekä hyvien parannusehdotusten laatimisessa.

Yhdeksi merkittäväksi asiantuntija-arvioinnin pulmaksi on nähty se, että se ei tarjoa systemaattista keinoa havaittujen ongelmien korjaamiseksi (mm. Nielsen & Molich 1990). Käytännön työn kannalta tämä on suuri haaste, sillä arvioinnin tilannut asiakas harvoin haluaa lopputuloksena pelkkää listaa havaituista ongelmakohtista, vaan tavoitteena on kehittää parempi tuote. Shneiderman (1998) on sitä mieltä, että vaikka arvioijat nostavat ongelmat hyvin esiin, niin ratkaisut ja uudelleen suunnittelu tulisi jättää suunnittelijoille.

Käytännön arvioinnissa arvioinnin tuloksista laaditaan tyyppillisesti raportti, joka toimitetaan arvioinnin tilaajalle. Tällöin iso rooli on tulosten esittämisessä ymmärrettävästi ja havainnollisesti. Capra (2006) on esittänyt huolensa erityisesti ongelmakuvausten laadusta. Hän kehitti ohjeet hyvän ongelmakuvausten laatimiseen sekä menetelmän ongelmakuvausten laadun arviointia varten. Vertaillen kokeneita ja kokemattomia arvioijia mittareillaan hän havaitsi kokeneiden tuottavan laadukkaampia ongelmakuvaus-  
kuvia.

Arvioinnin tasoon vaikuttaa myös tutkittavan tuotteen kehitysaste. Vaikka esimerkiksi heuristinen arviointi voidaan tehdä jo aikaisessa tuotekehitysvaiheessa, niin Nielsen ja Molich (1990) totesivat, että se soveltuu parhaiten käyttöliittymän siistimiseen, ei isoihin muutoksiin. Käytännössä kuitenkin varhaisen kehitysvaiheen prototyyppistä on järkevää etsiä juuri konseptuaalisia ja isompia rakenteellisempia ongelmakohtia ja jättää yksityiskohdat vielä kokonaan sivuun. Julkaisua odottavan tai jo käytössä olevan järjestelmän osalta puolestaan voi olla viisasta kerätä pienempiä ongelmakohtia, jotka korjaamalla käytettävyyttä voitaisiin kokonaisuudessaan parantaa – sen sijaan, että siinä vaiheessa tehtäisiin suuria muutoksia järjestelmän rakenteeseen tai konseptiin. Tämä raja riippuu myös hieman siitä, miten käytettävyys arvioinnin aikana ylipäänsä ymmärretään (ks. 2.1.1). Joka tapauksessa käytännössä olennaista onkin määritellä selkeät tavoitteet arvioinnin tuloksille.

Käytettävyyden menetelmien tutkimisen alkuaikoina tutkimuksissa arvioidut tuotteet olivat usein selkeästi rajattuja, paikoin yksinkertaisiaakin kokonaisuuksia. Käytännössä arvioitavat tuotteet saattavat kuitenkin usein olla monimutkaisia. Wharton ym. (1992) esimerkiksi totesivat, että kognitiivinen läpikäynti ei sovellu suoraan monimutkaisiin järjestelmiin, ainakin edustavien tehtävien merkitys silloin korostuu entisestään.

Arvioijien välisiä eroja tutkineet Hertzum ja Jacobsen (2003) nostavat pohdinnassaan avoimeksi kysymykseksi sen, voidaanko arvioinnilla saada valideja tuloksia eroista huolimatta. Käytännön kannalta varmasti kaikki havainnot, jotka eivät perustu virhetulkintoihin, ovat kuitenkin arvokkaita tuotteen kehittämisen kannalta. Erojen merkitys korostuu arvioijien keskinäisiä suorituksia vertailtaessa, mikä ei kuitenkaan ole kovin oleellista käytännön arvioinnissa.

### 2.3.2 Käytännön työn asettamat rajoitukset

Käytännön arviointiprojekteja koskevat samat säännöt kuin muitakin liike-elämän projektiluontoisia töitä: resurssit ovat rajatut, ja ne on kohdennettava tehokkaasti ja viisaasti tutkimusten tavoitteiden kannalta. Käytännön resursseilla aika on aina rajattu (Sears 1997),

joten se pitää käyttää mahdollisimman tehokkaasti hyödyksi. Rajattu aika asettaa vaatimukset arviointiprosessin tehokkuudelle sekä priorisoinnille, koska kaikkia laajojen järjestelmien ominaisuuksia tai puolia ei usein käytössä olevilla rajallisilla resursseilla voida käydä läpi. Vaikka esimerkiksi kognitiivinen läpikäynti tarkoitettiin alun perin suunnittelijoiden ja kehittäjien käyttöön, se ei kuitenkaan sovi suoraan heidän työnsä nopeaan ja kiireiseen luonteeseen (Wharton ym. 1992). Lisäksi projektihenkilöstö on käytännössä aina rajallinen.

Asiantuntija-arvioinnin tieteellinen tutkimus on osoittanut, että kattavan tuloksen saavuttaminen vaatii aina useiden arvioijien käyttämistä. Eri arvioijat löytävät eri ongelmia, ja suurempi määrä arvioijia löytää enemmän ongelmia kuin pienempi määrä samassa yhteisajassa (Virzi 1997). Nielsen ja Molich (1990) osoittivat, että heuristisessa arvioinnissa 3-5 arvioijaa on hyvä määrä kattavan tuloksen saavuttamiseksi ottaen huomioon käytetyt resurssit. Nielsen ja Landauer (1993) totesivat hyötysuhteen olevan suurimmillaan neljällä arvioijalla, mutta olivat myös sitä mieltä, että jopa 16 arvioijan käyttäminen on tulosten kannalta kustannusten arvoista. Lisäksi Nielsen (1992) havaitsi, että arvioijien kokemus vaikuttaa ratkaisevasti tarvittavien arvioijien määrään: kokemattomia arvioijia pitäisi olla 14, jotta löydetäisiin 75 % ongelmista, kun kokeneita arvioijia vastaavan tulokseen tarvittaisiin kolme. Toisaalta ylipäänsä kahdesta kolmeen arvioijaa riittäisi, jos arvioijat olisivat asiantuntijoita sekä käytettävyydessä että tuotetta koskevalla alalla. Sawyer ym. (1996) ovat havainneet käytännön arviointiprojekteissa, että kahden arvioijan ryhmät työskentelevät tehokkaasti ja tuottavat tehokkaasti yhtenäisiä tuloksia.

Käytännön arviointeja koskee siis haaste; miten optimoidaan käytettävien arvioijien määrä siten, että tulokset ovat kattavia, mutta toisaalta pysytään käytettävissä olevissa resursseissa. Lisäksi kokeneita asiantuntijoita ei aina ole käytettävissä kovin paljon, joten muita ratkaisuja on etsittävä. Esimerkiksi Karoulis ja Pombortsis (2000) pohtivat, miten opiskelijoiden tekemien arviointien tuloksia voitaisiin parantaa kokeneisiin arvioijiin nähden, sillä opiskelijoita olisi helpompi saada tekemään arviointeja, mutta toisaalta ero kokeneisiin on kuitenkin huomattava.

Resurssien lisäksi käytännön työhön vaikuttavat konkreettiset tavoitteet. Usein arvioinneissa määritetään yhdessä arvioinnin tilanteen asiakkaan kanssa arvioinnin rajaukset; mitä erityisesti tutkittavasta tuotteesta halutaan selvittää, mikä juuri kyseisellä hetkellä on merkittävää kehitystyön kannalta ja mihin osa-alueisiin erityisesti kiinnitetään huomiota. Usein asiakkaiden kanssa voidaan muotoilla jopa tarkkoja tutkimuskysymyksiä, joihin haetaan vastauksia. Erityisen tärkeää on, että arviointi tuottaa asiakkaalle tärkeät ja kiinnostavat tulokset (Sawyer ym. 1996). Rajattaessa arviointialuetta on selvää, että arvioinnin aikana joitakin osakokonaisuuksia tai ominaisuuksia jää arvioinnin ulkopuolelle, ja arvioijan on pidettävä tämä mielessä. Ilman rajausta esiin nousisi paljon havaintoja, jotka eivät kuitenkaan välttämättä olisi merkityksellisiä juuri kyseisessä tilanteessa. Arvioijan vastuulle jää siis arviointitilanteessa keskittyä olennaisiin ja ennalta sovittuihin asioihin.

## 2.4 Arvioijan vaikutus arviointiin

Arvioijan vaikutusta arviointiin on tutkittu aiemmin useista näkökulmista. Arvioijan vaikutus ei näy arvioinnissa pelkästään kokemuksen kautta, vaan myös esimerkiksi henkilökohtaisina eroina. Tässä luvussa esitellään aiempia havaintoja ongelmien määrän, arvioinnin ennustavuuden, yksilöllisten erojen sekä ongelmien kategorisoinnin kautta.

### 2.4.1 Kokemus ja ongelmien määrä

Kokeneiden arvioijien on todettu löytävän asiantuntija-arvioinnissa selvästi aloittelijoita enemmän käytettävyyso ongelmia. Esimerkiksi Nielsen (1992) vertaili tutkimuksessaan heuristisessa arvioinnissa kokemattomia arvioijia, käytettävyyden asiantuntijoita sekä sellaisia arvioijia, jotka olivat asiantuntijoita sekä käytettävyydessä että kyseisen tuotteen ongelma-alueella (kaksoisasiantuntijuus). Nielsen havaitsi, että kokemattomat arvioijat löysivät käytettävyyso ongelmia vähiten, asiantuntijat selvästi enemmän, ja kaksoisasiantuntijat kaikkein eniten. Desurvire (1994) havaitsi omissa tutkimuksissaan, että asiantuntijat löysivät selvästi kokemattomia enemmän ongelmia sekä heuristisessa arvioinnissa että kognitiivisessa läpikäynnissä. Hänen mukaansa asiantuntijat löysivät heuristisella arvioinnilla suhteessa parhaiten vakavia ongelmia, mutta löysivät kattavasti myös pieniä ongelmia. Lisäksi Jeffries ym. (1991) havaitsivat, että asiantuntijat löysivät yli 50 % todellisista ongelmista, mikä oli selvästi kokemattomia enemmän. Lisäksi tutkimuksissa on todettu, että arvioijan kokemuksesta riippumatta yksittäiset vakavat käytettävyyso ongelmat löytyvät arvioinneissa muita ongelmia varmemmin (Nielsen 1992; Jakobsen ym. 1998).

Edellä mainituissa tutkimuksissa kokeneiden arvioijien suoritusta kokemattomiin nähden on verrattu löydettyjen ongelmien absoluuttisen määrän avulla. Tästä herää kuitenkin kysymys, miten voidaan selvittää kaikkien ongelmien todellinen määrä, johon suorituksia vertaillaan. Nielsen käytti vertailukohtanaan itse luomaansa listaa, jota hän sitten täydensi tutkimuksen myötä. Desurvire puolestaan vertasi ongelmia käytettävyydestin avulla löydettyihin ongelmiin, toisin sanoen todellisiin käyttäjien kohtaamiin ongelmiin. Jeffries ym. taas muodostivat ongelma-avaruuden yksittäisten arviointien pohjalta yhdistäen. On selvää, että mikään näistä tavoista ei ole aukoton: itse muodostettu lista perustuu vain tiettyjen, vaikkakin kokeneiden asiantuntijoiden näkemykseen, testin avulla löydetään erilaisia ongelmia kuin asiantuntija-arvioinnilla, eikä millään yksittäisillä arviointimenetelmillä löydetä kaikenlaisia ongelmia. Tästä syystä ongelmien määrien vertailu varsinkin eri tutkimusten välillä voi johtaa harhaan.

Lukuisissa tutkimuksissa sovellettu käytettävyyso ongelmien absoluuttisen määrän vertailu tuntuukin arveluttavalta juuri siksi, että ongelmien laskeminen ei puutu ainakaan suoraan ongelmien abstraktiotasoihin: mikä itse asiassa on yksi ongelma, mikä taas useista yksityiskohdista muodostuva laajempi ongelmakokonaisuus. Miten määriä voidaan luotettavasti vertailla, jos ongelmat määritellään ja muodostetaan eri tavoin? Sears (1997) sanoo, että kaikkien ongelmien määrän hahmottaminen on hankalaa. Hänen mukaansa aiemmat tutkimukset eivät olleet ottaneet kantaa, mikä on oikeastaan vähäinen ongelma, josta ne puhuvat: ovatko jotkin ongelmat niin vähäisiä, etteivät ne ole oikeastaan käytettävyyso ongelmia? Arviointimenetelmien tutkimuksen alkuaikoina 1990-luvun alussa ongelman määrittely ei ehkä noussut esille siksi, että tutkimuksissa käytetyt tuotteet olivat melko rajattuja, jolloin myös löydettyistä ongelmista oli helppoa muodostaa konsensus. Todellisuudessa arvioitavat tuotteet saattavat kuitenkin olla hyvinkin monimutkaisia, jolloin havainnot eivät välttämättä liiku samalla abstraktiotasolla. Doubleday, Ryan, Springett & Sutcliffe (1997) huomauttavat, että arviointien subjektiivisuus ongelmien määrittelyssä tekee ongelmien vertailun vaikeaksi, toisin sanoen ongelmat eivät ole yhteismitallisia. Yhden arvioijan kirjaama ongelma voi käsittää useita toisen arvioijan kirjaamia ongelmia, ja jos arvioija nostaa esiin tietyn yksityiskohdan, tarkoittaako se erillistä ongelmaa vai samaa kokonaisuutta kuin laajempi kuvaus?

Arvioinnin riippuvuutta arvioijan kokemuksesta ja taidoista pidetään usein yhtenä suurimpana haasteena arvioinnin tulosten laadulle ja vertailukelpoisuudelle. Kokemattomat ja kokeneet arvioijat toimivat eri tavoin, mikä vaikuttaa väistämättä arvioinnin tuloksiin ja

korostaa asiantuntijuuden merkitystä arvioinnissa (Karoulis & Pombortsis 2000). Karoulis ja Pombortsis nostavat esiin kysymyksen, miten kokemattomien arvioijien työtä voitaisiin tukea niin, että erot kokeneisiin olisivat pienempiä. Skov ja Stage (2005) toteavat noviisien oppivan nopeasti, mutta sanovat eron olevan ongelmien tunnistamisessa. He kehittivätkin työkalun (Conceptual tool for problem identification), joka auttaa noviiseja pääsemään parempaan tulokseen niin ongelmien määrässä kuin vakavuusluokituksessa.

Vaikka tulokset kokeneiden ja kokemattomien arvioijien vertailuista ovatkin kiistattomia, jättävät ne vähemmälle huomiolle joitakin näkökulmia, kuten erot arviointitavoissa ja prosesseissa sekä ongelmien määrittelyssä ja luokittelussa abstraktiotasojen ja eri osa-alueiden kautta. Lisäksi Chattratchart ja Lindgaard (2008) painottavat, että asiantuntijan (*expert*) ja aloittelijan (*novice*) ero olisi ylipäänsä määriteltävä selkeästi, pelkkä oletus käytännön työn kokemuksesta ei riitä erotteluun.

## 2.4.2 Asiantuntija-arvioinnin ennustavuus

Ongelmien määrän vertailun ohella yhtenä tapana tarkastella arvioinnin kattavuutta ja todellista vaikutusta on käytetty sen ennustavuutta, toisin sanoen sitä, kuinka hyvin arvioinnin tulokset vastaavat tuotteen todellisen käytön ongelmia. Desurvire ym. (1991; 1992) havaitsivat, että kokeneiden arvioijien tulokset ennustavat todellista käyttöä kokemattomia paremmin. Tutkimuksessa verrattiin arvioijien havaitsemia ongelmia käytettävyydestin avulla löydettyihin ongelmiin.

Onkin selvää, että arvioijan kokemuksen merkitys tulee pelkästään asiantuntija-arviointikokemuksesta, vaan hyvin vahvasti myös käytettävyydesteistä, joita arvioija on ollut järjestämässä tai seuraamassa. Testien avulla arvioija on saanut tuntuman todellisiin käyttäjiin, ja tyyppillisiin heidän kohtaamiinsa ongelmiin. Sears (1997) sanoo kokemattomien arvioijien hukkaavan heuristisessa arvioinnissa fokuksen käyttäjästä. Heuristiset säännöt toisaalta auttavat heitä fokusoitumaan, mutta ei kuitenkaan välttämättä käyttäjään. He nojautuvat ja rajautuvat sääntöihin, mikä voi johtaa virheisiin. Esimerkkinä Sears mainitsee tilanteen, jossa tuotteesta puuttuva eksplisiittinen Ohje-toiminto johtaa heuristisen säännön rikkomiseen, vaikka näkymä olisikin riittävän yksinkertainen käytettäväksi ilman ohjetta.

Kokeneet arvioijat saattavat käyttää arvioinnin tukena skenaarioita, joilla he saavat arviointiin mukaan todellisen käytön mukaisia tehtäviä ja tilanteita (Nielsen 1995; Virzi 1997). Lisäksi kokeneiden arvioijien suorittamaan asiantuntija-arvioon sisältyy heuristisen arvioinnin lisäksi usein myös todellisen käytön mukaisten tehtävien läpikäyntiä (Riihiaho 2000), jolloin arvioinnin tuloksilla on luonnollisesti myös paremmat mahdollisuudet vastata todellisessa käytössä kohdattuja ongelmia. Kokeneet arvioijat toimivat arvioinneissa ikään kuin todellisten käyttäjien korvaajina (Virzi 1997).

## 2.4.3 Yksilölliset erot

Yksittäisten arvioijien välillä on todettu olevan suuria eroja myös arviointikokemuksesta riippumatta. Arvioinnin voi suorittaa periaatteessa kuka tahansa, joten henkilöiden taustat voivat olla hyvinkin erilaisia. Myös käytävyyden asiantuntijoiden taustat ovat erilaisia; osalla voi olla tekninen, osalla kognitiivisen psykologian, osalla taas graafisen suunnittelun koulutus. Tästä johtuen on selvää, että eri arvioijat ovat herkempiä havaitsemaan erilaisia asioita arviointia suorittaessaan. Lisäksi arvioijat havaitsevat ongelmia yksilöllisesti. Useissa tutkimuksissa on havaittu, että eri arvioijat löytävät eri ongelmia (mm. Virzi 1997). Osa arvioijista löytää paljon selviä ongelmia, osa taas määrällisesti vähemmän – mutta myös vaikeammin löydettäviä ongelmia (Nielsen & Molich 1990). Desurvire (1994) toteaa, että

hänen tutkimuksessaan parhaan arvioijan ja parhaan menetelmän yhdistelmä löysi 44 % käytettävyysongelmista. Yksilöllisistä eroista johtuen arviointiin suositellaan useita arvioijia kattavan arvioinnin saavuttamiseksi (mm. Nielsen & Molich 1990; Nielsen & Landauer 1993; Virzi 1997). Lisäksi parina työskentelevien arvioijien on todettu löytävän ongelmia enemmän kuin itsenäisten arvioijien (Karat, Campbell & Fiegel 1992).

Hertzum ja Jacobsen ovat tutkineet kattavasti yksilöllisiä eroja eri arviointimenetelmien soveltamisessa, kuten heuristisessa arvioinnissa (Hertzum & Jacobsen 2003), kognitiivisessa läpikäynnissä (Hertzum & Jacobsen 1999) ja käytettävyydestin analysoinnissa (Jacobsen ym. 1998). Tutkimuksissa on tyypillisesti verrattu kunkin arvioijan löytämiä käytettävyyso ongelmia ongelmien yhteismäärään, eli joukkoon uniikkeja käytettävyyso ongelmia. Yksilölliset erot näkyvät sekä kokeneilla että kokemattomilla arvioijilla, niin vakavissa kuin kosmeettisissakin ongelmassa, ongelmien etsimisessä ja vakavuusluokituksissa sekä monimutkaisissa ja yksinkertaisissa tuotteissa (Hertzum & Jacobsen 2003). Heidän mukaansa erot syntyvät siitä, että arviointi vaatii tulkintaa; vaikka osa ongelmista onkin itsestään selviä, niin useimmat vaativat tarkempaa vuorovaikutuksen analyysia. Erilaisiin tulkintoihin puolestaan vaikuttavat esimerkiksi arvioijien kognitiiviset kyvyt, kokemus, motivaatio, persoonallisuus sekä taitojen hallitseminen.

Slavkovic ja Cross (1999) havaitsivat tutkiessaan kokemattomien arvioijien haastaville tuotteille suorittamia arviointeja, että yksilölliset erot korostuvat entisestään monimutkaisten järjestelmien arvioinnissa. Kaikki arvioijat ohittivat useita kategorioita, mutta olivat tehokkaita joissakin. Ylipäänsä arvioijat löysivät selvästi vähemmän ongelmia kuin Nielsenin tutkimuksissa oli havaittu. He esittävät, että haastavien järjestelmien arviointiin ei riitä 5-10 kokematon arvioijaa, koska he keskittyvät eri osiin ja erilaisiin asioihin arvioinneissaan. Yksilölliset erot näkyvät myös kokeneiden arvioijien kesken. Bastien ja Scapin (1995) vertasivat heuristisia sääntöjä käyttäneiden ja pelkästään omaan kokemukseensa nojanneiden asiantuntijoiden tuloksia keskenään, ja totesivat asiantuntijoiden hyötyvän heurististen sääntöjen käyttämisestä. Tämä on mielenkiintoista siksi, että Karat ym. (1992) kyseenalaistavat sääntöjen hyödyllisyyden kokeneilla arvioijilla.

Yhden henkilön suorittama arviointi on joka tapauksessa väistämättä jollakin tavalla subjektiivinen, vaikuttivat siihen sitten mitkä tahansa seikat henkilön kokemuksessa tai taustassa. Hertzum, Jacobsen ja Molich (2002) toteavat, että ensimmäinen askel selviytyä yksilöllisistä eroista on tiedostaa ja hyväksyä erot. Heidän tutkimuksessaan tuli esimerkiksi ilmi, että yhdentoista arvioijan joukossa vain 9 % ongelmista oli sellaisia, jotka vähintään kaksi arvioijaa löysi. Tästäkin huolimatta arvioijilla oli yhteisymmärrys arvioinnin tuloksista.

Henkilökohtaisiin eroihin saattavat johtaa myös erot ongelmien määrittelyssä. Eri arvioijat mieltävät käytettävyyso ngelman käsitteen eri tavoin. Ongelmien vaihtelevien abstraktiotasojen lisäksi ongelman määrittelyyn vaikuttavat tarkastelunäkökulma ja ongelmien osa-alueet sekä se, mitä ylipäänsä pidetään käytettävyyso ngelmina. Hertzum ja Jacobsen (2003) toteavat, että ilman käytettävyyso ngelman määrittelyä (ongelmakriteeristöä) kaikki havainnot helposti määritellään käytettävyyso ngelmiksi. Heidän mukaansa tutkimuksissa olisi määriteltävä käytettävyyso ngelma, muuten eri arvioijat saattavat kirjata eri havainnot samasta ilmiöstä. He kuitenkin toteavat määrittelyn olevan vaikeaa, koska ongelmat liittyvät tuotteeseen, tehtäviin ja arvioinnin tavoitteisiin. Myös Doubleday ym. (1997) nostavat esiin, että eri arvioijat saattavat muodostaa samasta havainnoista erilaisia ja -kokoisia ongelmia, joiden vertailu keskenään on vaikeaa.

Ongelmia on luokiteltu perinteisesti sääntölistojen aiheiden, vakavuuden sekä muiden suoraviivaisten tapojen avulla (mm. Cuomo & Bowen 1994; Karat 1994). Myöhemmin

Keenan ym. (1999) kehittivät käytettävyysongelmille taksonomian, toisin sanoen eräänlaisen kehyksen ongelmien globaalien tason luokitteluun, vertailuun ja analysointiin. Hvannberg ja Law (2003; ks. myös Vilbergsdóttir, Hvannberg & Law 2006) esittivät luokittelu- ja määrittelymenetelmän käytettävyysongelmille (*Classification of Usability Problems, CUP*). Chattratichart ja Lindgaard (2003) muodostivat ongelmien tarkastelua varten ongelmaprofiilit ja jakoivat tarkasteltavat ongelmat aiheen mukaisiin osa-alueisiin (sisältö, navigointi, grafiikka, toimintojen tehokkuus, muotoilu ja layout, kieli, apu ja virheilmoitukset). Vaikka eri lähestymistavat ongelmien luokitteluun ja analysointiin jossain määrin poikkeavat toisistaan, yleisesti on kuitenkin selvästi havaittu se, että ongelmia on tarkasteltava tarkemmin kuin pelkän määrän perusteella. Edellä mainitut luokittelukeinot perustuvat lähinnä käytännön tarpeisiin ja tuotteiden parantamiseen, eivät kuitenkaan arviointien tulosten keskinäiseen vertailuun.

### 3 KOKEMUKSEN KEHITTYMINEN

---

Kokemuksen kertymistä ja asiantuntijuuden muodostumista on tutkittu pitkään psykologian alalla. Tässä luvussa esitetään kootusti ja tiivistetysti kokemukseen liittyviä teemoja erityisesti kognitiivisesta psykologiasta ja kognitiotieteestä.

#### 3.1 Asiantuntijuus (*expertise*)

Asiantuntijuudella<sup>3</sup> tarkoitetaan yksinkertaistettuna sellaisia ominaisuuksia, kykyjä ja mekanismeja, jotka tekevät henkilön suorituksesta ylivoimaisen valtaosaan muista verrattuna (Ericsson & Smith 1991). Tietyllä alalla asiantuntijaksi kutsuttavalla henkilöllä on alaltaan erityistä tietämystä ja taitoa, jotka ovat kehittyneet ammattimaisen harjoittelun ja käytännön kokemuksen avulla. Käytännössä laajamittainen harjoitus kehittää asiantuntijuuden, joka auttaa henkilöä selviytymään omaan alaansa kuuluvista tehtävistä useimpia muita paremmin (Anderson 2000). Yleistä ja kattavaa määritelmää siitä, miten asiantuntijan suoritus eroaa edukseen, on mahdotonta antaa, mutta joka tapauksessa alakohtaisesti keskeisissä tehtävissä hänen suorituksensa ilmenee ylivoimaisena alalla kokemattomiin nähden.

Asiantuntijuutta ja sen kehittymistä on tutkittu pitkään vertaamalla kokemukseltaan eritasoisten henkilöiden – useimmiten aloittelijoiden (*novice*) ja asiantuntijoiden (*expert*) tai muuten kokeneiden henkilöiden – suorituksia tutkimusalaan liittyvissä tehtävissä. Nimensä mukaisesti asiantuntijuus koskee aina jotakin rajattua alaa, jolta on havaittavissa selkeitä alakohtaisia tehtäviä. Asiantuntijuutta onkin tutkittu lukuisilla yksittäisillä aloilla, joista tunnetuimpia lienevät shakki (mm. Simon & Chase 1973), fysiikka (mm. Chi, Feltovich & Glaser 1981), lääketiede (mm. Patel & Groen 1991) ja musiikki (mm. Sloboda 1991).

Alakohtaisuudesta huolimatta asiantuntijuudelle on löytynyt selvästi yhtenäisiä piirteitä ja ominaisuuksia. Asiantuntijat pystyvät käsittelemään omaan alaansa ja tehtäviinsä liittyviä asioita aloittelijoita tehokkaammin ja varmemmin (Saariluoma 1990). Tämä johtuu muun muassa siitä, että he muodostavat tutuista käsiteltävistä asioista laajempia ja tehokkaampia kokonaisuuksia, eivätkä joudu käsittelemään samoja asioita aina yksityiskohdista lähtien. Asiantuntijat kykenevät ylipäänsä kiertämään luontaisia tiedonkäsittelyn rajoja (ks. 3.3) aloittelijoita paremmin. Asiantuntijoille on kehittynyt kyseisellä ongelma-alueella tehokkaita skeemoja ja malleja, jotka nopeuttavat olennaisten asioiden havaitsemista ja tunnistamista suoraan monimutkaisen päättelyn sijaan (Ericsson & Smith 1991). Esimerkiksi lääketieteessä (mm. Patel & Groen 1991) ja fysiikassa (Anzai 1991) on havaittu, että asiantuntijat tyypillisesti tunnistavat relevantin kokonaisuuden nopeasti yksittäisten pienten vihjeiden avulla (esim. oireet lääketieteessä tai käytettävissä olevat tiedot fysiikan ongelmissa). Lisäksi asiantuntijoilla on yksinkertaisesti huomattavasti enemmän alakohtaista osaamista, joka on alakohtaisissa tehtävissä palautettavissa pitkäkestoisesta muistista prosessoinnin tueksi.

Monilla aloilla asiantuntijan käsite (*expert*) määritellään siten, että asiantuntija on harjoittanut omaa alaansa jopa kymmenen vuotta (Ericsson & Smith 1991). Tämän työn käsittelemän tutkimusalan, eli käytettävyytutkimuksen asema on kuitenkin hyvin erilainen verrattuna luvun alussa mainittuihin, paljon tutkittuihin aloihin. Käytettävyytutkimus on alana verrattain nuori; vaikka käytettävyyden juuret ovatkin jo sinänsä vanhemmassa

---

<sup>3</sup> Tässä työssä termiä asiantuntijuus käytetään suomennoksena englanninkielen *expertise*-käsitteelle. Vastaavasti psykologian määrittelemä *expert*-käsite suomennetaan asiantuntijaksi. Vaikka kyseisiä suomenkielisiä termejä käytetäänkin tavallisessa kielessä laajemmassa merkityksessä, tässä työssä niiden määritelmä on rajattu mainittuihin käsitteisiin.

ergonomiassa, niin laajemmin käytettävyyttä on tutkittu käytännössä vasta viime vuosikymmeninä tietotekniikan kehittymisen myötä (Nielsen 1993; Hollinsed & Novick 2007). Lisäksi ala on edelleen hyvin kapea, vaikka tietoisuus käytettävyydestä onkin kasvanut. Nämä seikat johtavat siihen, että asiantuntijan määritelmän perusteella puhtaasti käytettävyytutkimuksen asiantuntijoita on tässä vaiheessa olemassa suhteellisen vähän moniin muihin aloihin verrattuna. Näin ollen myös asiantuntijuuden tutkiminen käytettävyytutkimuksen alalla on haastavampaa. Nielsen (1992) totesi tutkimuksessaan, että käytettävyydasiantuntijalle ei ole virallista määritelmää. Omassa tutkimuksessaan hän määritteli käytettävyydasiantuntijan käytettävyyden alalta olevan tutkinnon tai usean vuoden työkokemuksen perusteella. Toisaalta Chattratchart ym. (2008) ovat sitä mieltä, että käytettävyyden asiantuntijan (*expert*) ja kokemattoman arvioijan (*novice*) erot on määriteltävä, pelkkä oletus käytännön työstä ei riitä.

Edellä mainituista syistä johtuen tässä tutkimuksessa, ja etenkin sen empiirisessä osuudessa keskitytäänkin puhumaan kokeneista arvioijista, ei varsinaisista asiantuntijoista. Kokenut arvioija määritteli tässä tutkimuksessa siten, että hän on suorittanut käytettävyyden asiantuntija-arviointeja ja käytettävyyttestejä täysipäiväisesti työkseen vähintään kahden vuoden ajan. Kokematon arvioija puolestaan on suorittanut yliopistotasolla käytettävyyden peruskursseja, opetellut tyypillisimmät käytettävyytutkimuksen menetelmät ja kokeillut niitä käytännössä, mutta ei ole vielä soveltanut niitä työssään.

Asiantuntijuuden määritelmä johtaa myös pieneen ristiriitaan tämän tutkimuksen käsittelemän menetelmän, käytettävyyden asiantuntija-arvioinnin suhteen. Miten voidaan puhua yleisesti asiantuntija-arvioinnista, jos asiantuntijuuden määritelmän mukaisesti asiantuntijoiden joukko on rajattu? Tämä ongelma kuitenkin sivuutetaan tässä työssä ajattelemalla asiantuntija-arviointia menetelmänä, jolla voidaan tehdä relevantteja havaintoja käytettävyysongelmista aina kunkin arvioijan osaamiseen perustuen. Tämä näkökulma tukee sitä, että arvioijan ei välttämättä tarvitse olla asiantuntija (ks. 2.3).

## 3.2 Oppiminen ja taidon kehittyminen

Kokemuksen kehittyminen – siinä missä ihmisen kykyjen kehittyminen ylipäänsä – perustuu oppimiseen (Anderson 2000). Oppiessaan ihminen muodostaa opittavasta asiasta mentaalaisia malleja tai mielikuvia, joita hän kykenee soveltamaan uusissa tilanteissa. Oppimista tarkastellaan psykologiassa usein taitojen kehittymisen kautta (Anderson 2000). Oman kiinnostuksen tai jonkin ulkoisen syyn pohjalta ihminen alkaa harjoittaa tiettyä taitoa. Harjoittaminen saattaa olla puhtaasti tekemistä tai sitten systemaattista harjoittelua. Vaikka systemaattinen harjoittelu useimmiten alalla kuin alalla johtaa parempiin tuloksiin, kehittyä harjoitettu taito myös pitkäjäksoisen tekemisen avulla (Ericsson & Smith 1991).

Joka tapauksessa nykykäsityksen mukaan oppiminen ja taitojen kehittyminen perustuvat yksinkertaisesti pääosin harjoittamiseen. Ihmisillä on toki synnynnäisiä perittyjä ominaisuuksia, jotka joko edesauttavat tai vaikeuttavat oppimista ja taidon kehittämistä, mutta yksinään ne eivät johda asiantuntijuuteen (Ericsson & Smith 1991). Tavallisessa kielessä puhutaan usein lahjakkuudesta, mutta todellisuudessa taitavaksi kehittymisessä pääpaino on yksinkertaisesti harjoittamisella, lahjakkuuden osuus taidosta on lopulta hyvin pieni<sup>4</sup>. Tässä työssä synnynnäiset ominaisuudet jätetään huomioimatta, ja eroja tarkastellaan puhtaasti kokemuksen ja oppimisen synnyttäminä.

---

<sup>4</sup> Lahjakkuus voi toki tukea taidon oppimista tarjoamalla suotuisat edellytykset taidon harjoittamiselle. Suuri merkitys voi olla myös siinä, että lahjakkuuden havaitsemisen ansiosta taidon harjoittaminen alkaa riittävän aikaisin.



Lisäksi opittu taito on usein hyvin kapea-alaista, eikä sen siirtäminen toiselle alalle ole kovin helppoa (Anderson 2000). Tämä johtuu siitä, että taitavuus tai asiantuntijuus tietyllä alalla vaatii tiettyjen yksityiskohtaisten ja alakohtaisten erikoistehtävien osaamista, jotka eivät siinä tarkkuudessaan toistu muilla aloilla. Abstraktimmat ja yleisemmät proseduurit ja mentaaliset mallit (skeemat) voivat toki olla joillakin aloilla yhteisiä, mikä helpottaa siirtymistä ja oppimista toiselle alalle, etenkin jos alat ovat lähellä toisiaan<sup>5</sup>. Kuitenkin asiantuntijuus vaatii nimenomaan alan tietämystä kuvaavan pyramidin kärjen hallintaa, mikä voi syntyä vain sillä kyseisellä alalla (Ericsson & Smith 1991). Toisaalta luovuus voi joskus olla sitä, että pystyy tuomaan toisaalla käytetyn ajattelumallin omalle alalleen (Saariluoma 1990). Tämä voi tosin vaatia osaamista kahdella alalla.

Yksinkertaisimmillaan oppiminen voidaan nähdä opittavaan taitoon liittyvän tiedon tallettamisena säilömuistiin, eli pitkäkestoiseen muistiin. Varsinainen tiedon käsittely taidon harjoittamisen aikana tapahtuu työmuistissa. Työmuistissa tietoa käsitellään siten, että se voidaan varastoida säilömuistiin myöhempää käyttöä varten. Taidon oppimisen ero pelkkään tietoon on siinä, että opittu taito myös säilyy muistissa pitkään, kun pelkkä tieto voi kadota nopeasti (Anderson 2000).

Ihmiselle kertyvä taitoon liittyvä tietämys voi olla luonteeltaan deklarativista tai proseduraalista (Anderson 2000). Deklaratiivinen tietämys perustuu opittuihin faktoihin, peruskäsitteisiin ja niiden jäsentämiseen. Proseduraalinen tietämys puolestaan pohjautuu tekemiseen ja toimintaan, ja on siten syvempää kuin deklarativinen (Anderson 2000). Aluksi taitoon liittyvä tietämys on puhtaasti deklarativista, minkä jälkeen proseduraalinen tietämys vahvistuu. Kokeneen henkilön taito kuitenkin jossain määrin muodostuu edellä mainittujen tietomuotojen yhdistelmänä. Tiedosta voidaan puhua myös eksplisiittisenä tai implisiittisenä tietona: aluksi tieto on eksplisiittistä, mutta kokemuksen kehittyessä henkilön implisiittinen tieto lisääntyy.

Anderson (2000) esittelee taidon kehittymisen tarkastelemisen kolmen kehitysvaiheen avulla (Fitts & Posner 1967): kognitiivisen, assosiatiivisen ja autonomisen. Kognitiivinen vaihe perustuu deklarativiseen tietämykseen, eli ihminen pitää mielessään taitoon liittyviä faktoja ja ominaisuuksia, harjoittelee niitä ja yrittää muistella niitä tarvitessaan. Assosiatiivisessa vaiheessa taitoon ja siihen liittyvään suoritukseen kuuluvien osien väliset yhteydet vahvenevat ja suorituksia varten muodostuu selkeitä proseduureja. Toisin sanoen proseduraalinen tietämys vahvistuu. Samalla helpot virheet perussuorituksessa vähenevät asteittain. Autonomisella tasolla taitoon liittyvä suoritus automatisoituu ja nopeutuu, ja prosessointi vaatii entistä vähemmän resursseja.

### 3.3 Tiedonkäsittelyn rajoitukset

Käytettävyyssarvioinnin suorittaminen voidaan nähdä luonteeltaan yhdenlaisena ajatteluprosessina. Ajattelun psykologiaan kuuluvat olennaisesti käsitteet ongelmanratkaisu, päättely ja päätöksenteko, joilla tarkoitetaan ajattelutehtävien eri muotoja. Nämä kaikki on nähtävissä myös osina käytettävyyssarvioinnin tekemistä. Kaikkiin edellä mainittuihin käsitteisiin liittyy myös hyvin kiinteästi ihmisen työmuistin käsite. Työmuistilla tarkoitetaan muistia, joka vastaa ihmisen aktiivisen toiminnan aikana tapahtuvasta tiedon käsittelystä (Saariluoma 1988a). Toiminnan ja ajattelun aikana tarvittavat tiedot tallentuvat työmuistiin, jossa niitä käytetään tehtävän aikana. Jotkut tiedot saattavat päätyä pitkäkestoiseen muistiin, toiset taas katoavat tehtävän suorittamisen jälkeen.

---

<sup>5</sup> Aiemmalla tietämyksellä voi tosin olla myös negatiivinen ja hidastava vaikutus siirryttäessä uudelle alalle. Opitut skeemat saattavat ohjata ihmistä harhaan uudella alalla, jos ne eivät sellaisenaan sovellu alalle.

Ajattelutehtävät vaativat joidenkin asioiden pitämistä mielessä prosessoinnin aikana. Näitä asioita kutsutaan mieltämisyksiköiksi (Saariluoma 1990). Selvin ajattelua koskeva työmuistin piirre on sen rajoitettu kapasiteetti. Vaikka työmuistin koko voi aavistuksen vaihdella henkilöstä riippuen<sup>6</sup>, on työmuistissa yleensä mahdollista pitää kerrallaan vain muutamia mieltämisyksiköitä (Saariluoma 1990). Jos täyteen työmuistiin tallentuu uusi yksikkö, poistuu samalla jokin edellisistä. Tämä asettaa rajat tietojen käsittelemiselle ajattelutehtävien aikana. Rajojen kiertämiseen on kuitenkin olemassa erilaisia keinoja ja tekniikoita. Taitava ajattelu, taito ja kokemus voidaankin tavallaan nähdä juuri erilaisina tapoina kiertää työmuistin rajoituksia (Saariluoma 1990).

Ongelmanratkaisu on keskeinen osa ajattelun psykologiaa. Itse asiassa ajattelu monesti nähdään ongelmanratkaisuna (Saariluoma 1988b). Ongelmalla tarkoitetaan sitä, että ihmisellä on tavoite, jonka saavuttaminen vaatii tiettyjä ratkaisuja, toimenpiteitä ja päätöksiä.

Aiemmissa tutkimuksissa on havaittu, että ongelmanratkaisussa tietyn ongelma-alueen kannalta kokemattomien henkilöiden suoritus on heikompi kokeneisiin nähden. Esimerkiksi lääketieteen asiantuntijuuden tutkimus on osoittanut, että kokeneet lääkärit kykenevät tehokkaampaan ongelmanratkaisuun kuin aloittelevat lääketieteen opiskelijat (Patel & Groen 1991). Tietyllä alalla kokemattomien henkilöiden heikompi ongelmanratkaisun taso johtuu pääosin työmuistin rajallisesta kapasiteetista (he esimerkiksi tekevät virheitä enemmän) (Saariluoma 1988a). Kokemattomien työmuisti täyttyy helposti tehtävän kannalta epäoleellisesta informaatiosta, tai yksinkertaisesti työmuistissa ei ole tilaa kaikille tehtävän vaatimille yksittäisille tiedoille. Kokemattomien skeemat ovat niin pieniä, toisin sanoen mieltämisyksiköt ovat niin pieniä yksityiskohtia, että olennaisia osia tiedosta jää työmuistin ulkopuolelle (Ericsson & Smith 1991). Kokeneiden skeemat puolestaan ovat muodostuneet ongelma-alueella niin laajoiksi, että työmuistiin talletettavat mieltämisyksiköt ovat selvästi kokemattomia suurempia. Niiden avulla työmuistiin saadaan mahtumaan laajempia kokonaisuuksia, eli tehtävien kannalta olennainen tieto. Eli vaikka työmuistin koossa ei ole kokemuksesta johtuvia eroja, ero muodostuu siitä, minkälaisia mieltämisyksiköitä työmuistiin voidaan tallettaa (Saariluoma 1990). Suurempien mieltämisyksiköiden käyttämisestä puhutaan myös pitkäkestoisena työmuistina.

Toinen merkittävä tapa kiertää työmuistin rajoja on ongelma-avaruuden rajoittaminen. Ongelma-avaruudella tarkoitetaan käytännössä erilaisten vaihtoehtojen, valintojen ja ratkaisujen joukkoa. Kokematon joutuu käymään läpi koko ongelma-avaruutta – tai koska se ei käytännössä ole mahdollista, tekemään enemmän sattumanvaraisia ratkaisuja ja valintoja, jotka eivät perustu tehtävän kannalta olennaiseen tietämykseen. Kokeneet puolestaan pystyvät sulkemaan ulkopuolelle ennalta huonoiksi tai soveltumattomiksi tuntemiaan vaihtoehtoja ja keskittymään ongelmanratkaisussa niihin, jotka ovat tehtävän kannalta tehokkaita (Saariluoma 1988b).

Kokeneet pystyvät erottamaan ongelma-alueen tehtävien kannalta olennaisen epäolennaisesta (Saariluoma 1988b). Esimerkiksi kokeneet shakinpelaajat eivät välttämättä laske mielessään merkittävästi kokemattomia pidempiä siirtosarjoja, vaan he kykenevät tunnistamaan pelin kannalta olennaisia nappula-asetelmia, joilla ongelma-avaruutta voidaan rajoittaa (de Groot 1965). Siinä missä myös epäolennaiset asiat kuormittavat kokemattomien työmuistia, kokeneiden kyky rajautua olennaiseen johtaa parempiin suorituksiin ja tuloksiin. Kokeneiden skeemat sallivat heidän muodostaa ongelma-alueella laajempia ja tuttuja kokonaisuuksia, joita he voivat käyttää mieltämisyksiköinä.

---

<sup>6</sup> Työmuistin koon vaihtelu on tosin äärimmäisen pientä.

Työmuistin koko vaikuttaa myös päätöksentekoon. Päätöksenteko tarkoittaa tilannetta, jossa henkilöllä on useita toimintavaihtoehtoja, joiden joukosta hänen on valittava yksi. Päätöksen tekeminen vaatii eri vaihtoehtojen punnitsemista ja asettamista paremmuusjärjestykseen. Kokemus vaikuttaa järjestyksen laatuun; koska järjestämiseen ja päätösvaihtoehtojen punnitsemiseen vaikuttavat useat eri päätösattribuutit, auttaa kokemus löytämään attribuuteista ne seikat, jotka johtavat rationaaliseen päätökseen. Kokemattomien päätökset eivät välttämättä ole yhtä rationaalisia, koska kyky suhteuttaa eri attribuutteja ei ole yhtä kehittynyt. Työmuistin kapasiteetti ei riitä pitämään muistissa kaikkia vaihtoehtoja ja attribuutteja, vaan valikointi niiden välillä on välttämätöntä. Valikointi voi johtaa helposti virhepäätöksiin, jolloin kokemuksen merkitys sopivan valikoinnin kannalta korostuu. (Saariluoma 1990)

Päätöksenteon rationaalisuuteen vaikuttavat niin sanotut päätöksenteon heuristiikat, joilla tiedostamatta kierretään työmuistin rajoja. Heuristiikat ohjaavat päätöksentekoa; monesti ne helpottavat päätöksentekoa, mutta ne saattavat myös johtaa siinä harhaan. Tyypillisiä esimerkkejä tutkituista heuristiikoista ovat (Tversky & Kahneman 1974):

- Edustavuus: Tyypillisten esimerkkien esiintymistä pidetään todennäköisenä.
- Saavutettavuus: Ilmiöitä, joista on helppo muistaa esimerkkejä, pidetään todennäköisinä.
- Ankkuroituminen: Alustava ja suuntaa-antava arvo johtaa arvioiden suuntautumiseen sitä kohti.

Kolmas luvun alussa mainituista ajattelutehtävien muodoista on päättely, joka pohjautuu formaalille logiikalle. Päättely tarkoittaa sitä, että joistakin lähtöoletuksista seuraa tiettyjä johtopäätöksiä. Päättelyssä ihminen käyttää tietoaan ja soveltaa niitä tiettyihin tilanteisiin. Erilaiset lähtöoletukset, usean päättelyn yhdistelmät ja päättelyyn vaikuttavien muuttujien suuri määrä vaikeuttavat päättelyä ja aiheuttavat siksi usein myös virheitä (Saariluoma 1988b), mikä johtuu muiden ajattelumuotojen tapaan pääosin työmuistin rajoittuneesta kapasiteetista. Päätöksenteon tapaan ihminen soveltaa myös päättelyyn erilaisia heuristiikkoja, jotka yksinkertaistavat päättelyä.

Mielenkiintoisin esimerkki päättelyn heuristiikoista tämän tutkimuksen kannalta on päättely osana käytettävyyden arviointia: arvioitaessa tuotetta on kyettävä päättelemään, ovatko arvioinnin kuluessa havaittavat asiat käytettävyysongelmia, ja kuinka vakavia sellaisia. Käytettävyyсарvioinnin aikana arvioijat soveltavat heuristisia sääntöjä, jotka ovat käytännössä hyviksi ja toimiviksi todettuja ohjeita tai sääntöjä. Heuristiset säännöt toimivat nyrkkisääntöjen tapaan: ne ohjaavat useimmiten oikeaan päätelmään, mutta saattavat jossain tilanteessa myös johtaa harhaan. Heurististen sääntöjen avulla arvioijilla on päättelyyn tueksi valmis pohja, eikä arviointiin liittyvää päättelyä tarvitse tehdä jokaisen havaittavan asian kohdalla alusta alkaen. Toisaalta joitakin heuristisia sääntöjä on arvosteltu juuri siitä, että niiden avulla ei voida huomioida kontekstia riittävän hyvin, eli ne ovat turhan yleisellä tasolla kunnollisten tulosten saavuttamiseksi. Tunnetuimmat heuristiset säännöt käytettävyyden arviointiin ovat Nielsenin kymmenen heuristista sääntöä (Nielsen 1993), jotka on lueteltu luvussa 2.2.1.

Vaikka tässä luvussa on perusteltu tietyllä alalla kokeneiden henkilöiden ajattelun tehokkuutta kokemattomiin verrattuna lähinnä työmuistin käyttämisen tavoilla, asiantuntijoiden etu ei perustu pelkästään työmuistiin, vaan myös pitkäkestoisen muistin käyttöön. Sen lisäksi että asiantuntijat kykenevät muodostamaan alallaan laajempia kokonaisuuksia mieltämisyksiköiksi työmuistin käyttöön, he kykenevät myös muistamaan suurempia määriä näitä erilaisia malleja ja kokonaisuuksia, mikä on pitkäkestoisen muistin ominaisuus (Andersson 2000). Eli kokeneet henkilöt ovat kokemuksen kehittyessä

tallettaneet pitkäkestoiseen muistiinsa lukuisia malleja, jotka ovat myöhemmin hyödynnettävissä uudelleen mieltämysyksiköinä samankaltaisissa tehtävissä.

### 3.4 Havaitseminen ja tunnistaminen

Edellisissä luvuissa on tarkasteltu kokemuksen kehittymistä ja kokemuksen vaikutusta siihen, miten ongelma-alueella havaittuja asioita voidaan käsitellä tehokkaasti ongelma-alueen tehtävien kannalta. Tässä luvussa keskitytään siihen, miten nämä havainnot itse asiassa tehdään ja miten ongelma-alueen kannalta keskeisiä asioita tunnistetaan.

Asioiden havaitseminen perustuu aistimiseen. Aistiensa avulla ihminen rekisteröi ympäröivästä maailmasta erilaisia aistiärsyksiä, jotka voivat olla elektromagneettista säteilyä tai kemiallisia ja mekaanisia ärsyksiä (Andersson 2000). Aistiärsykkeet siirtyvät signaaleina aivoihin, jossa varsinaiset havainnot muodostuvat koko tiedonkäsittelyprosessin tuloksena ärsykkeiden tulkitsemisen pohjalta. Ihmisen havainnot perustuvat tavallisesti usean eri aistin tuottamiin ärsyksiin (Andersson 2000). Tyypillisen käytettävyyssarvioinnin tekeminen pohjautuu tästä huolimatta kuitenkin pääosin näköaistiin, paitsi tietysti silloin, jos arvioinnin kohteena on fyysinen tuote tai äänikäyttöliittymä. Koska aistiminen on mekaanista toimintaa, eikä henkilön kokemus vaikuta aistiärsykkeiden muodostumiseen, ei aistiminen sinänsä ole kuitenkaan tämän tutkimuksen kannalta mielenkiintoista<sup>7</sup>. Kiinnostavaa sen sijaan on, miten aistiärsyksiin perustuva havaitseminen eroaa kokemuksesta riippuen.

Koska havaintojen muodostuminen tapahtuu aivoissa, siihen vaikuttavat voimakkaasti ihmisen aiemmat kokemukset ja tiedot. Aiemmat kokemukset ohjaavat havaitsemista niin, että jotkut asiat havaitaan nopeammin ja varmemmin, jotkut taas saattavat jäädä kokonaan havaitsematta. Kyse ei ole yksin siitä, että tutut asiat havaittaisiin helpommin, vaan enemmänkin siitä, että kiinnostaviksi ja mielekkäiksi nähtävät asiat havaitaan varmemmin. Esimerkiksi jokin tuttu asia voi olla aiemmin mielletty turhaksi tai epäoleelliseksi jossain tehtäväympäristössä, jolloin se saatetaan automaattisesti sivuuttaa havaitsemisessa.

Näköhavainnot ovat siis tämän tutkimuksen kannalta havaitsemisen kiinnostavin osa. Esitietoisten havaintojen muuttuessa tietoisiksi näköhavainnot jäsentyvät piirreohjattusta havaitsemisesta, jossa kohteesta kerätään ja tutkitaan yksityiskohtia, käsiteohjattuun havaitsemiseen, jossa yksityiskohdista muodostetaan nopeasti hypoteeseja sisällön tulkitsemista varten (Andersson 2000). Hypoteeseja verrataan havaitsemistilanteessa havainnon sisältöön; jos ne eivät sovi, ne hylätään ja muodostetaan uusia. On kuitenkin huomattava, että tämä hypoteesien vertailu on pääosin tiedostamatonta, toisin sanoen henkilö ei kykene eksplisiittisesti kertomaan hypoteesiensa vertailusta. Hypoteesit muodostuvat edellä kuvattuun tapaan ihmisen aiempien kokemusten ja tiedon pohjalta, joten niiden laatu ja oikeellisuus riippuvat vahvasti kertyneestä kokemuksesta.

Näköhavainnon jäsentymistä ohjaavat voimakkaasti myös hahmolait, joilla tarkoitetaan pitkälti synnynnäisiä tapoja yhdistellä havaittavia piirteitä. Hahmolakeja ovat esimerkiksi läheisyys, samanlaisuus, jatkuvuus, tuttuus, valiomuotoisuus, yhteinen liike, yhteenliittyminen ja sulkeutuvuus. Koska hahmolait – kuten muutkaan esitietoisen vaiheen automaattiset ilmiöt – eivät kuitenkaan ole suorassa yhteydessä kokemukseen ja sen kehittymiseen, niiden tarkempi käsittely sivuutetaan tässä yhteydessä.

---

<sup>7</sup> Tosin mekaanisen aistimisen ja sen rajoitusten tunteminen on eräs avainasioita käytettävyystudkimuksen ja käyttöliittymäsuunnittelun kannalta. Arvioijien kokemuksen vaikutuksen tutkimisen kannalta se ei kuitenkaan ole keskeinen teema, minkä vuoksi se tässä työssä sivuutetaan.

Havaitseminen perustuu siis ihmisen oppimisiin malleihin ja skeemoihin, jotka aktivoituvat tiettyjen aistiärsykkeiden pohjalta. Kokemus jollakin alalla vaikuttaa siis luonnollisesti siihen, millaisia malleja, skeemoja tai ennakkokäsityksiä henkilöllä on havaitsemisensa tueksi. Esimerkiksi näköhavaintojen osalta skeemat ohjaavat niin katseen etenemistä kuin ärsykkeiden tulkittamista havainnoiksi (Saariluoma 1990). Näiden laajuus ja vahvuus on tutussa ongelmaympäristössä pääsääntöisesti etu, mutta koska aivot tavallaan pakottavat havaitsemista näihin opittuihin malleihin, voi se luonnollisesti johtaa myös virheisiin. Lisäksi on huomattava, että muistiin tallentuvat aina tulkitut havainnot, eivät objektiiviset aistiärsykkeet.

Monien alojen tutkimuksissa on osoitettu, että alallaan kokeneet henkilöt pystyvät tunnistamaan suorituksensa aikana tyypillisiä ja toistuvia malleja, kuvioita ja kaavoja (Ericsson & Smith). Tunnistaminen voi tapahtua useiden eri aistihavaintojen perusteella ja niiden yhteistyössä, mutta tyypillisimmin se tapahtuu visuaalisesti (esim. shakki, Simon & Chase 1973) tai auditiivisesti (esim. musiikki, Sloboda 1991). Esimerkiksi kokeneet shakkipelaajat kykenevät tunnistamaan aiemmista pelintilanteista tuttuja nappula-asetelmia ja muodostelmia, joita he voivat käyttää pelitilanteen ymmärtämiseen ja strategian luomiseen pohtimatta juurikaan yksittäisiä nappuloita pelilaudalla. Kokemattomat taas joutuvat arvioimaan tilannetta pitkälti yksittäisten nappuloiden, niiden suhteiden ja yksittäisten siirtosarjojen vertailun kautta. Itse asiassa kokeneiden shakkipelaajien ero kokemattomiin ei synny niinkään pidempien siirtosarjojen laskemisen kautta – kuten voisi olettaa, vaan opittujen nappulamodostelmien ja niitä seuraavien tilanteiden hahmottamisen avulla (de Groot 1965).

Näköhavaintoja ja niiden muodostumista voidaan tutkia monin eri tavoin, mutta objektiivista dataa katseen etenemisestä ja kiinnittymisestä voidaan kerätä esimerkiksi silmänliiketutkimuksen avulla. Silmänliiketutkimuksessa käytetään apuna katseenseurantalaitteistoa, joka rekisteröi ihmisen katseen kiinnittymisen tiettyihin kohteen pisteisiin (fiksaatiopisteet) sekä siirtymät näiden pisteiden välillä (sakkadit). Havaitseminen tapahtuu vain fiksaatiopisteiden avulla, siirtymien aikana havaitsemista ei tapahdu. Silmänliiketutkimusta onkin käytetty käytettävyydetutkimuksessa keräämään tietoa tuotteen käyttäjän katseen etenemisestä ja ohjautumisesta tuotteen suhteen. Silmänliiketutkimus kertoo totuudenmukaisesti katseen liikkeen ja kiinnittymisen, mutta se ei kuitenkaan kerro varmasti, ovatko katsotut kohteet havaittu tai miten havainnot on tulkittu. Suurin haaste silmänliiketutkimuksessa onkin kerätyn datan analysoinnissa: miten voidaan olla varmoja, kuinka ihminen on todella tulkinut näkemänsä, ja onko hän ylipäänsä ymmärtänyt sen. Joka tapauksessa kaiken muun mahdollisen ihmisen toiminnasta kerättävän tiedon tukena silmänliiketutkimus tuo mielenkiintoisen lisänäkökulman sekä joissakin tapauksissa varmistuksen tutkimuksessa havaittaville ilmiöille.

## 4 KOKENEIDEN JA KOKEMATTOMIEN ARVIOIJIEN VERTAILU

---

Tässä luvussa esitellään tutkimuksen empiirinen osuus. Luvussa esitellään tutkimus, koasetelma, sekä kerätyn tiedon analysointitavat. Empiirisen osuuden tulokset esitellään luvussa 5.

### 4.1 Tutkimuksen esittely

Ericsson ja Smith (1991) mainitsevat yhtenä asiantuntijuuden tutkimisen keskeisenä haasteena paremmuuden kriteerien määrittelyn: miten voidaan valita selkeät mitattavat tai tutkittavat piirteet, joiden avulla paremmuus voidaan kiistattomasti osoittaa. Esimerkkinä he mainitsevat shakin, jossa eritasoisten pelaajien ratkaisuja samassa tilanteessa verrataan keskenään. Paremmuus ratkeaa selvittämällä, kuka tekee tilanteessa parhaan siirron – mutta miten voidaan määrittellä, mikä on paras siirto? Sama ongelma koskee myös tätä tutkimusta: mikä tekee arvioijan suorituksesta toista paremman? Yksinkertaisena mittarina on usein käytetty löydettyjen käytettävyysongelmien määrää, mutta tämä lähestymistapa ei kuitenkaan ota juuri huomioon ongelman määrittelyn problematiikkaa eikä sitä, miten ongelmat nostetaan esiin (ks. 2.4). Tästä syystä tässä tutkimuksessa arvioijien välisiä eroja haluttiin selvittää ongelmien määrän lisäksi laajemmin myös muista näkökulmista. Kriteereitä paremmuudelle ei kuitenkaan etukäteen laadittu, vaan tutkimuksessa etsittiin erilaisia eroja eritasoisten arvioijien välillä ja pyrittiin löytämään eroille syitä.

Tutkimuksessa vertailtiin kokeneiden ja kokemattomien arvioijien suorittamien asiantuntija-arviointien eroja sekä arviointiprosessin että tulosten osalta. Tutkimuksessa joukko kokemukseltaan eritasoisia arvioijia suoritti laboratorio-olosuhteissa arvioinnin samoille tuotteille: neljälle erilliselle verkkosivuston osalle. Arviointiin liittyvää tietoa kerättiin ennen arvioinnin aloittamista, sen aikana sekä arvioinnin jälkeen. Arvioinnin aikana kiinnostus kohdistui erilaisten menetelmien ja tapojen seuraamiseen suoritettaessa arviointia samoille tuotteille. Arvioinnin tuloksia puolestaan vertailtiin pääasiassa arvioijien kirjaamien havaintojen ja suullisesti antamien korjausehdotusten pohjalta.

Tutkimukseen otettiin mukaan kuusi kokenutta arvioijaa sekä kuusi aloittelevaa arvioijaa. Näiden lisäksi molemmista ryhmistä oli tutkimuksen pilottivaiheessa mukana yksi arvioija. Myös pilottiarvioijat otettiin harkinnan jälkeen huomioon joidenkin osien analysoinnissa, jos koasetelma ja ohjeistus eivät näiltä osin muuttuneet pilotin jälkeen.

### 4.2 Tutkimukseen osallistuneet arvioijat

Vertailun onnistumisen ja luotettavuuden kannalta tutkimukseen oli tärkeää saada arvioijia, jotka olivat selvästi joko kokeneita tai kokemattomia arvioijia. Toisaalta kaikkien piti tuntea asiantuntija-arviointi menetelmänä ja osata soveltaa sitä. Nielsen (1992) totesi tutkimuksessaan, että käytettävyyssiantuntijalle ei ole virallista määritelmää. Omassa tutkimuksessaan hän määritteli käytettävyyssiantuntijan käytettävyyden alalta olevan tutkinnon tai usean vuoden työkokemuksen perusteella. Aloittelijoiden asemassa puolestaan on käytetty monissa tutkimuksissa opiskelijoita, jotka opiskelevat alaa, mutta joille ei vielä ole kertynyt pitkää kokemusta (mm. Nielsen & Molich 1990).

Tässä tutkimuksessa arvioijat valittiin niin, että kaikki olivat opiskelleet arviointimenetelmiä ja kokeilleet niitä käytännössä. Aloittelijat määriteltiin tutkimuksessa niin, että he eivät olleet vielä tehneet työkseen käytettävyyden arviointia ilman käyttäjiä tai käyttäjien kanssa. Kokeneemmilla asiantuntijoilla puolestaan oli oltava vähintään kahden vuoden käytännön

työn tuoma kokemus arvioinneista ja käytettävyydesteistä. Asiantuntija-arviointien lisäksi myös käytettävyydestien kokemus on tutkimuksen kannalta merkittävä. Käytettävyydestit tuovat käytännön kokemusta käyttäjän toiminnasta, jota voi käyttää hyödyksi asiantuntija-arvioinneissa.

Tutkimukseen osallistuneet kokeneet arvioijat olivat käytettävyydetutkimukseen erikoistuneen Adage Oy:n asiantuntijoita. Tutkimuksen tuloksia arvioitaessa on pidettävä mielessä se, että kokeneet arvioijat olivat yhdestä ja samasta yrityksestä. Yrityksen toimintatavat ja kulttuuri vaikuttavat varmasti arvioijien työhön henkilökohtaisten erojen ja kokemuksen lisäksi. Kokemattomat arvioijat poimittiin mukaan tutkimukseen Teknillisen korkeakoulun Käytettävyyden arviointi -kurssilta, jolla he olivat juuri päässeet opettelemaan käytettävyydestausta ja käytettävyyden asiantuntija-arviointia. Kaikki kokemattomat arvioijat olivat siis opiskelijoita, kokeneet pääosin valmistuneita.

Kaikki tutkimukseen osallistuneet henkilöt tulivat mukaan vapaaehtoisesti ja ilman palkkiota. Motiivina kokemattomille arvioijille oli tutustua paremmin asiantuntija-arviointiin ja oppia siitä lisää. Kokeneille puolestaan kiinnostavaa oli päästä vertaamaan omaa arviointiaan muiden kokeneiden arvioijien työhön.

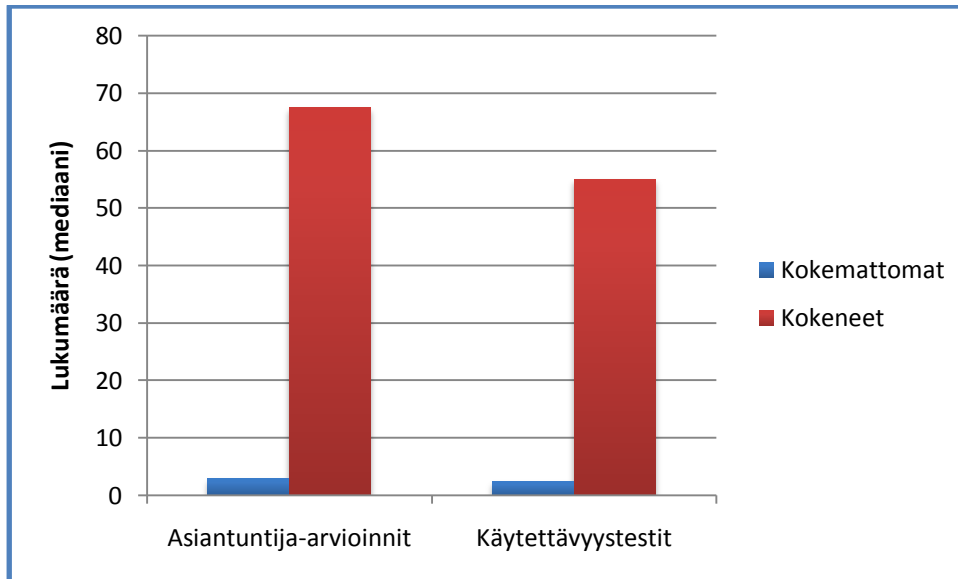
Taulukko 1 esittää tutkimukseen osallistuneiden arvioijien taustan ja kokemuksen käytettävyyden arvioinnista. Kokemattomista arvioijista neljä (kuudesta) oli tekniikan opiskelijoita Teknillisestä korkeakoulusta, yksi Taideteollisen korkeakoulun opiskelija ja yksi kognitiotieteiden opiskelija Helsingin yliopistosta. Kaikkien kokemus käytettävyyden arvioinnista rajoittui viimeisimmän vuoden aikana suoritettuihin käytettävyydetutkintoihin. Suoritettujen asiantuntija-arviointien määrä vaihteli kahdesta viiteen ennen tutkimusta. Käytettävyydesteistä heillä oli kokemusta yhden ja neljän testin väliltä.

Kokeneista arvioijista viiden tausta oli lähinnä Teknillinen korkeakoulu, yksi oli koulutukseltaan filosofian maisteri. Suoritettujen asiantuntija-arviointien määrä vaihteli välillä 40–300, käytettävyydestien välillä 20–200. Kokeneet arvioijat olivat tutkimushetkellä tehneet käytettävyyden arviointia työssään kolmesta kahdeksaan vuotta. (ks. Taulukko 1)

**Taulukko 1.** Taulukossa esitetään tutkimukseen osallistuneet varsinaiset arvioijat (ilman pilotiarvioijia). Arvioijat 1-6 edustivat kokemattomia arvioijia, 7-12 puolestaan kokeneita.

Nro	Ammatti	Kokemus ajallisesti	A-arviointien määrä	Käyt.testien määrä
1	Opiskelija, tekn. yo.	1 v	5	1
2	Opiskelija, tekn. yo.	1 v	2	2
3	Opiskelija, tekn. yo.	1 v	4	4
4	Opiskelija, tekn. yo.	1 v	3	3
5	Opiskelija, tait. yo.	1 v	3	3
6	Opiskelija, fil. yo.	½ v	3	2
7	Laboratoriopäällikkö	8 v	300	200
8	Johtava asiantuntija	5 v	60	40
9	Johtava asiantuntija	5 v	75	30
10	Käytettävyydasiantuntija	6 v	60	70
11	Kehittämiskojohtaja	9 v	180	100
12	Käytettävyydasiantuntija	3 v	40	20

Kuva 1 esittää kaaviossa mediaanien avulla kokeneiden ja kokemattomien arvioijien suorittamien asiantuntija-arviointien ja käytettävyydestien määrät. Kaaviossa selvästi näkyvä ero on myös tilastollisesti merkitsevä.



**Kuva 1.** Tutkimukseen osallistuneiden arvioijien suorittamien asiantuntija-arviointien ja käytettävyydestien määrä (mediaani).

## 4.3 Koeasetelma

Tässä luvussa esitellään tutkimuksen koeasetelma. Koeasetelmaan kuuluvat arvioinnin kohteena käytetyt tuotteet, kokeen järjestämiseen käytetty tila sekä kokeessa kerätyt tiedot.

### 4.3.1 Arvioinnin kohde

Arvioitavina tuotteina tutkimuksessa käytettiin neljää erillistä osaa www-pohjaisista selainkäyttöliittymistä, joista kukin muodosti oman tehtävänsä. www-palvelut ovat nykyisin tyypillisimpiä arvioitavia tuotteita, joten näin voitiin olla varmempia asiantuntijuudesta ja kokemuksesta vertailuasetelman saavuttamiseksi. Yleisyytensä lisäksi www-palvelut ovat kiinnostava arvioinnin kohde myös siksi, että ne ovat luonteeltaan erilaisia verrattuna arviointimenetelmien kehittämisen alkuaikoina käytettyihin tuotteisiin. Esimerkiksi heuristisen arvioinnin on todettu soveltuvan hyvin myös www-palveluihin, mutta erilaisesta ongelmajoukosta johtuen niiden on todettu vaativan usein erilaisia heuristisia listoja (Instone 1997; Karoulis & Pombortsis 2000). Lisäksi www-palveluiden jatkuva käyttö johtaa siihen, että arvioija on aina jollain tasolla myös itse palvelun käyttäjä (Karoulis & Pombortsis 2000).

Arvioinnin laajuus rajattiin niin, että arvioinnin suorittamiseen käytettävä aika oli rajallinen, mutta kuitenkin mahdollisimman luonnollinen koeasetelmasta huolimatta. Koetilanteessa osien arviointiin annettiin yhteensä puolentoista tunnin aika. Ajan rajoittaminen oli välttämätöntä, sillä kokeen kokonaiskesto haluttiin rajata maksimissaan kahteen tuntiin. Rajoitettu aika toki vaikuttaa arvioinnin suorittamiseen, mutta toisaalta se myös luo tutkimukselle mielenkiintoisen näkökulman: miten rajattu aika vaikuttaa työskentelyyn. Ajan rajaamisen avulla tutkimuksessa mahdollisesti esiintyvät arvioijien väliset erot on myös mahdollisesti selvemmin havaittavissa, koska ne eivät ala tasoittua pelkästään pidemmän käytettävissä olevan ajan vaikutuksesta. Ajan rajaaminen oli perusteltua myös siksi, että käytännön resursseilla aika on aina rajattu (Sears 1997).

Koeasetelmaan otettiin mukaan kaksi yksittäistä sivua, jotka oli irrotettu kokonaisista www-palveluista, yksi monisivuinen tilausprosessi sekä yksi useampaan välilehteen jakautuva sivukokonaisuus. Valinnalla pyrittiin selvittämään eroja arvioitaessa sekä yksittäisiä sivuja että useista sivuista koostuvia kokonaisuuksia. Vaiheittaisen tilausprosessin arvioiminen



vaatii yksittäisten sivujen arvioinnin lisäksi jonkinlaista käyttäjän tehtäviin perustuvaa läpikäyntiä kokonaisuuden hahmottamiseksi. Yksittäisen sivun arvioinnissa arvioija voi keskittyä sivun sisäiseen toteutukseen, kun taas useamman sivun kokonaisuudet vaativat sivujen välisten suhteiden ja riippuvuuksien huomioimista.

Arvioidut sivut ja sivustot oli tallennettu arviointia varten todellisen sivun kokoisina kuvina, jotka avattiin www-selaimeen. Tällä pyrittiin toisaalta estämään arvioijien harhautuminen sivuilta pois arvioinnin aikana, toisaalta taas kuvien käyttö helpotti koeasetelmaan kuuluneen silmänliiketutkimuksen tallentamista ja analysointia. On huomattava, että www-sivun arviointi kuvien avulla esimerkiksi ilman toimivia linkkejä ja hiiren osoittimen muuttumista ei ole aivan luonnollista. Arvioidut sivut kuitenkin näyttivät www-selaimella todellisilta, ja arvioijille selitettiin tutkimuksen alussa kuvien luonne ja niiden tarkoitus.

#### 4.3.1.1 Tehtävä 1: Teknillisen korkeakoulun www-sivuston Ajankohtaista-sivu

Arvioinnin ensimmäisenä tehtävänä oli Teknillisen korkeakoulun www-sivustoon kuulunut Ajankohtaista-sivu (ks. Kuva 2). Sivulle oli koottu Teknillisen korkeakoulun toimintaan liittyviä uutisia, tiedotteita, tapahtumia sekä muita ajankohtaisia aiheita. Sivun oikeassa laidassa oli pitkä lista linkkejä, jotka johtivat ajankohtaisiin aiheisiin muualle TKK:n sivustolle, eri yksiköiden ja osastojen sivuille sekä muualle verkkoon. Sivun varsinainen oma sisältö esitettiin sisältöalueella, joka rajautui selaimen vasemman reunan ja edellä mainitun linkkilistan väliin, eli käytti valtaosan sivun pinta-alasta.

**Kuva 2.** Ensimmäisessä tehtävässä arvioitu Teknillisen korkeakoulun verkkosivuston Ajankohtaista-sivu.

Sivu valittiin arvioinnin kohteeksi kahdesta eri syystä. Ensinnäkin sivu oli erittäin pitkä ja huonosti jäsenneily. Sen hahmottaminen kokonaisuudessaan vaati vierittämistä ylhäältä alas sekä eri osien välisten suhteiden pohtimista. Toisekseen sivu oli elementteiltään hyvin sekava: sillä oli hyvin paljon pieniä yksittäisiä, mutta selvästi näkyviä käytettävyyso ongelmia, jotka johtuivat lähinnä sivun viimeistelemättömyydestä ja elementtien visuaalisesta epäyhätenäisyydestä sekä epäkonventionaalisuudesta. Nämä ongelmat olivat luonteeltaan sellaisia, että niitä oli periaatteessa helppo poimia, vaikka arviointikokemus olisikin vähäinen.

Todellisuudessa ehkä vakavin sivua koskeva ongelma oli se, että se ei vastannut tarkoitustaan kovin hyvin. Varsinaista sivun ajankohtaista sisältöä ei ollut helppo tarkastaa kovin nopeasti, koska sisältöä ei ollut esitetty selkeästi ja ymmärrettävästi, jäsenelty hyvin eikä erotettu sivun toissijaisesta sisällöstä. Näistä lähtökohdista oli mielenkiintoista tutkia, mihin asioihin eri arvioijat keskittyvät arvioinneissaan: kuinka tehokkaasti sivulta kerätään lukuisat pienet ja selvät ongelmat, kuinka paljon taas mennään tätä syvemmälle. Lisäksi sivun valintaa puolsi se, että ajankohtaisten aiheiden sivu on hyvin tyyppillinen osa monia www-palveluita, joten sen saattoi olettaa olevan tarkoitukseltaan hyvin tuttu kaikille arvioijille.

Ensimmäisen tehtävän arviointiin annettiin aikaa 15 minuuttia. Aika voi tuntua lyhyeltä, mutta kyseessä oli kuitenkin yksittäinen sivu, jota tuskin todellisessakaan arvioinnissa sellaisenaan tarkasteltaisi kovin paljon pidempään. Aika oli käytännössä riittävä sivun tarkastelemiseen kokonaisuudessaan eri näkökulmista, mutta ei kuitenkaan sallinut kaikkien yksittäisten linkkien tai tekstien pohtimista perin pohjin.

#### 4.3.1.2 Tehtävä 2: Elisan yksityisasiakkaiden liittymänhallintapalvelun sivu

Toisena tehtävä arvioijat arvioivat yksittäistä sivua, joka kuului Elisa Oyj:n asiakkailleen tarjoamaan www-palvelukokonaisuuteen (ks. Kuva 3). Palvelussa Elisan yksityisasiakas voi hallinnoida oman matkapuhelinliittymänsä tietoja sekä liittymäpalveluita. www-palvelu vaatii asiakasta kirjautumaan omilla tunnuksillaan, mutta kirjautuminen ei sisällynyt arviointiin. Kirjautuneen käyttäjän nimi tosin esitettiin myös arvioidulla sivulla.

The screenshot shows the Elisa customer portal interface. At the top, there is a navigation bar with the Elisa logo and 'Matkaviestintä' (Mobile Communications). Below this, there are tabs for 'yksityiset' (private) and 'yritykset' (business). The main content area is titled 'Elisan asiakkaan palvelut' (Elisa customer services) and includes a search bar and a 'Siirry' (Go) button. The page is divided into several sections:

- Asiakastiedot ja liittymät** (Customer information and subscriptions): This section contains links for 'Laskut ja Saldo' (Invoices and balance), 'Liittymän hankinta' (Subscription purchase), 'Liittymätyyppin vaihto' (Subscription type change), 'Palveluvalikon muokkaus' (Service menu modification), and 'Viestikeskus' (Message center).
- Liittymät** (Subscriptions): A table listing six subscriptions with columns for 'Liittymä' (Subscription), 'Käyttäjä' (User), and 'Omistaja' (Owner). Each row includes a 'Hallinnoi liittymää' (Manage subscription) link.
- Laskut ja saldo** (Invoices and balance): A section with a warning message and a table listing invoices with columns for 'Liittymä' (Subscription) and 'Asiakasnumero' (Customer number). Each row includes 'Katso lasku' (View invoice) and 'Tarkista saldo' (Check balance) links.

Kuva 3. Toisessa tehtävässä arvioitu Elisan yksityisasiakkaiden liittymien hallinnointisivu.

Arvioidulla sivulla asiakas voi nähdä sisältöalueella kootusti oman matkapuhelinliittymänsä tiedot sekä liittymää koskevat palvelut. Sivu tarjoaa mahdollisuuden siirtyä muokkaamaan eri tietoja toisille sivuille, joita arvioinnissa ei voinut nähdä. Sivun ylälaidassa on Elisan logo sekä www-palvelusta ja sen osasta kertova otsikko. Vasemmassa reunassa esitetään navigointivalikko, jonka avulla voi siirtyä www-palvelun muihin osiin. Arvioinnin kannalta

valikko antaa myös kuvan palvelun muusta sisällöstä, kun arviointi muuten rajoittuu vain yhteen sivuun.

Tehtävässä arvioitu sivu on ulkoisesti melko yksinkertainen ja pitkälti visuaalisesti viimeistelty. Vaikka kyseistä palvelua ei vielä arvioinnin aikana ollut julkaistu, oli se jo hiottu valmiiksi, eikä siitä ollut heti havaittavissa silmiinpistäviä käytettävyysongelmia. Pintaa syvemältä ongelmia kuitenkin löytyi: sivun rakenteessa ja käyttäjän luonnollisten tehtävien kautta. Tämä teki sivusta arvioinnin kohteena mielenkiintoisen; sivu on selvästi siistimpi kuin ensimmäisen tehtävän sivu, ja siltä löydettävissä olevat ongelmat ovat hyvin eriluonteisia kuin ensimmäisessä tehtävässä.

Merkittävä piirre toisen tehtävän arvioinnissa oli se, että sivu oli irrotettu kokonaisesta palvelusta. Arviointiin vaikutti siis puutteellinen konteksti: arvioijat eivät voineet tarkastella muita palvelun sivuja tai tutkia siirtymistä arvioidulta sivulta muille sivuille. Niin ikään sivun rooli kokonaisuuteen nähden jäi arvioijien pääteltäväksi, koska se ei ollut tarkastettavissa. Joka tapauksessa tutkimuksen kannalta tämä ei ollut varsinainen puute – ennemminkin se toi mielenkiintoisen asetelman arvioijien erojen tutkimiseen.

Toisen tehtävän arviointiin annettiin ensimmäisen tapaan 15 minuuttia. Molemmissa tehtävissä arvioitavana oli yksittäinen ja irrallinen sivu, joskin toisen tehtävän sivu oli selvästi ensimmäistä yksinkertaisempi ja pelkistetympi.

#### **4.3.1.3 Tehtävä 3: Elisan yritysasiakkaiden liittymien hallintapalvelu**

Kolmantena tehtävänä oli arvioida nelivaiheinen palvelu, jolla Elisan yritysasiakkaat voivat tilata yrityksensä matkapuhelinliittymiin lisäpalveluita. Palvelu etenee vaiheittain neljän erillisen sivun avulla, jotka muodostavat ohjatun läpikäytävän polun (ks. Kuva 4).

Ensimmäisenä sivuna on hakulomake, jolla haetaan hallinnoitavia matkapuhelinliittymiä kaikkien yrityksen liittymien joukosta erilaisin hakukriteerein. Toisella sivulla esitetään hakutulos; käyttäjä valitsee haun tuottamista tuloksista ne liittymät joita haluaa muokata. Kolmannessa vaiheessa annetaan tarvittavat asetukset valituille liittymille sekä muut tilausta koskevat tiedot. Viimeisessä vaiheessa puolestaan hyväksytään tehtävät toimenpiteet.

Tutkimuksessa käytettiin HTML-sivujen sijaan sivuista otettuja ruutukaappauksia – kuvia, jotka oli linkitetty toisiinsa niiden navigointipainikkeiden ja linkkien osalta, jotka todellisessa palvelussa toimintaa veivät eteenpäin. Toisin sanoen palvelu ei ollut täysin toiminnallinen, mutta etenemisen kannalta tärkeät navigointielementit oli kuitenkin linkitetty kuviin.

Toisen tehtävän sivun tapaan myös kolmannen tehtävän palvelu oli jo viilattu melko valmiiksi, ja sitä oli siistitty visuaalisesti. Näin ollen ensimmäisen tehtävän kaltaisia selvästi näkyviä yksittäisiä ongelmia ei tehtävässä ollut. Sitä vastoin toimintopolun keskeisimmät ongelmat liittyivät eri vaiheiden yhtenäisyyteen sekä siihen, miten loogisesti koko polku käyttäjän toiminnan kannalta etenee. Arviointitehtäväksi tämä valittiinkin juuri siksi, että yksittäisen sivun sijaan siinä on huomioitava liikkumista usean erillisen sivun välillä, sekä tilauksen etenemistä kokonaisuudessaan.

Koska edellisistä tehtävistä poiketen arvioinnin kohteena oli yksittäisen sivun sijaan useamman sivun sisältävä polku, annettiin kolmannen tehtävän arviointiin yhteensä 30 minuuttia.

elisa | Matkaviestintä

yksityiset yritykset

Liittymien hallinta - pääsivulle

## Vastaaja Pro:n tilaus

Hae yrityksesi liittymiä Vastaaja Pro:n tilausta varten.

**Palvelun hinta:** Hinta määräytyy yrityssopimuksen tai hinnaston mukaisesti.

**Lisätietoa**  
→ [Vastaaja Pro](#)

1 **Liittymien haku** 2 Liittymien valinta 3 Vastaaja Pro:n tilaus 4 Kuittaus

Perushaku Tarkennettu haku

Liittymänumerot:  Esim. 050 123456

Käyttäjien nimet:  Esim. järvinen, järvinen matti

Asiakasnumerot:  Valitse monta Ctrl-näppäin pohjassa.  
123948558 STOCKMANN OYJ/TAPIOLAN ST  
123948558 STOCKMANN OYJ/TAPIOLAN ST  
123948558 STOCKMANN OYJ/TAPIOLAN ST

Kustannuspaikat:  Esim. hallinto, A100, 123

Hakujoukko 2004 liittymää  
[Listaa kaikki](#)

- Voit antaa yhden tai useamman hakuvedon
- Haku toimii JA-operaattorilla
- Hakusanat erotetaan pilkulla (,)
- Nimen alulla voi hakea, mutta muuten hakutiedot pitää syöttää kokonaan

Palautte | Yhteystiedot | Käyttöehdot © Elisa Oyj 2004

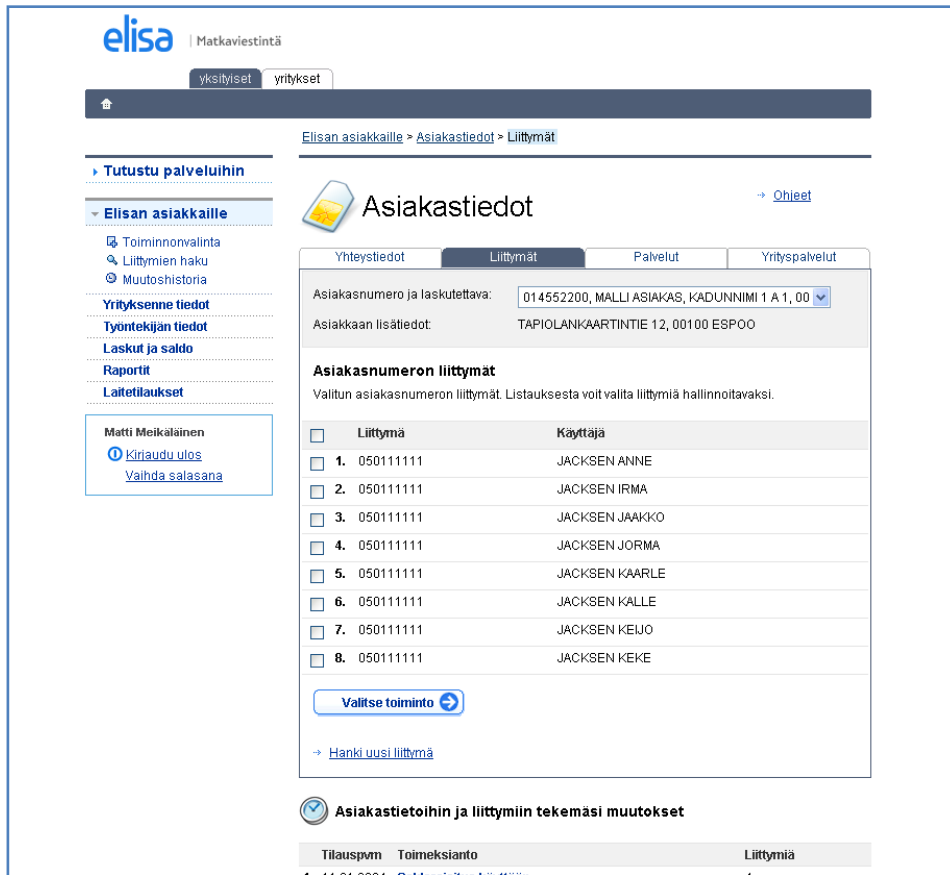
**Kuva 4.** Ensimmäinen vaihe kolmannessa tehtävässä arvioidusta Elisan yritysasiakkaiden liittymien hallinnointipalvelusta.

#### 4.3.1.4 Tehtävä 4: Elisan yritysasiakkaiden liittymien hallintapalvelu (välilehdet)

Neljäntenä ja viimeisenä tehtävänä oli arvioida Elisan yritysasiakkaiden käyttöön tarkoitettu sivu, jolla asiakas voi tarkastella yrityksen liittymien tietoja sekä hallinnoida niitä. Sivun sisältö on jaettu neljälle välilehdelle, joista kukin esittää tietyn kokonaisuuden liittymiä koskevista tiedoista. Eri välilehdillä esitetään toisistaan riippumatonta tietoa, eli välilehdet ovat itsenäisiä. (ks. Kuva 5.)

Tehtävässä arvioitu käyttöliittymä muistutti visuaalisesti tehtävien 2 ja 3 palveluita, ja oli myös samalle tasolle viimeistely. Käytettävyysoongelmia oli kuitenkin löydettävissä siitä, miten tieto on jaettu eri välilehdille. Lisäksi välilehdet eivät olleet sisällöltään selkeitä.

Viimeisen tehtävän arviointi poikkesi koeasetelmassa muista tehtävistä siten, että siinä arvioijien ei tarvinnut kirjata tekemiään havaintoja ylös, vaan niiden esittäminen suullisesti riitti. Tämän tarkoituksena oli toisaalta keventää viimeisen tehtävän arviointia, toisaalta myös tarkastella, mikä on ongelmien kirjaamisen vaikutus arviointiprosessiin ja sen etenemiseen. Koska havaintojen kirjaamiseen ei kulunut aikaa, annettiin viimeisen tehtävän arviointiin yksittäisten sivujen (tehtävät 1 ja 2) tapaan aikaa 15 minuuttia.



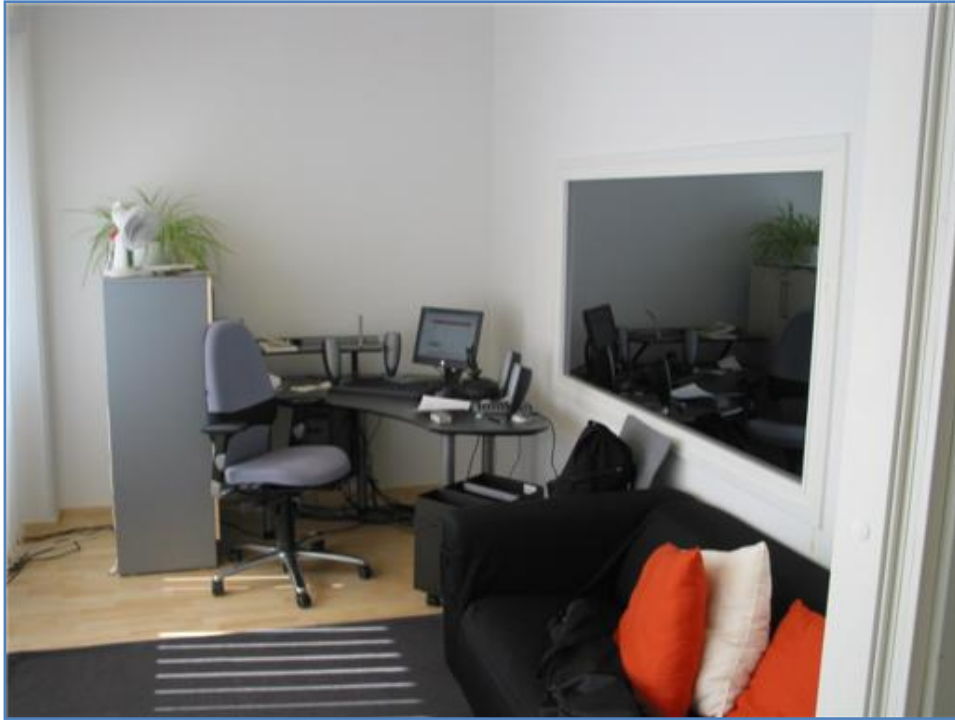
Kuva 5. Sivu neljännessä tehtävässä arvioidusta Elisan yritysasiakkaiden Asiakastiedot-kokonaisuudesta.

### 4.3.2 Arviointipaikka

Arvioinnit suoritettiin Adage Oy:n käytettävyysslaboratoriossa (ks. Kuva 6). Tila oli tyypillinen käytettävyysslaboratorio, joka vastaa rauhallista ja viihtyisää työhuonetta. Laboratoriossa oli työpöytä ja tietokone, jolla arviointi suoritettiin. Seinän toisella puolella oli käytettävyyssiestien seuraamista varten erillinen tarkkailuhuone, josta testitilaan näkee puoliläpäisevän lasin avulla. Tarkkailuhuonetta ei kuitenkaan tutkimuksessa käytetty, vaan kirjoittaja istui samassa tilassa arvioijan vieressä ohjaten tutkimuksen etenemistä.

Käytettävyysslaboratorio valittiin tutkimuksen suorituspaikaksi erityisesti tilassa olevien valmiiden tallennusmahdollisuuksien vuoksi. Äänen ja kuvaruutukaappauksen tallentaminen onnistui vaivatta laboratorion tietokoneella. Muutenkin tila oli suunniteltu erityisesti juuri tutkimusten järjestämiseen.

Laboratoriossa oli lisäksi valmiina katseenseurantalaitteisto, jota tutkimuksessa käytettiin. Katseenseurantalaite tarkoittaa tässä tapauksessa itse asiassa vain erikoisrakenteista tietokoneen näyttöä, joka kykenee rekisteröimään tietokoneen käyttäjän silmien liikkeitä (ks. Kuva 7). Fyysisen laitteen lisäksi katseenseurantaan tarvittiin tallennusohjelmisto, joka oli testikoneella valmiiksi asennettuna.



**Kuva 6.** Adage Oy:n käytettävyyslaboratorio, jossa arvioinnit suoritettiin.



**Kuva 7.** Tutkimuksessa silmänliikkeiden tallentamiseen käytetty näyttö.

### **4.3.3 Tutkimuksessa kerätty tieto**

Tutkimuksessa kerättiin arvioinneista tietoa usealla eri tavalla. Tietoa kerättiin noin viikkoa ennen arvioinnin aloittamista, arvioinnin aikana sekä välittömästi arvioinnin jälkeen.

#### **4.3.3.1 Taustatiedot**

Tutkimukseen osallistuvia arvioijia hankittaessa varmistuttiin siitä, että he soveltuivat kokemattomien tai kokeneiden arvioijien ryhmiin ennalta määriteltyjen kriteerien perusteella (ks. 4.2). Muut taustatiedot selvitettiin arviointitilanteen aluksi määrämuotoisen

alkukyselyn avulla. Kyselyllä selvitettiin arvioijien opintojen suuntautuminen sekä tarkempi kokemus käytettävyyden arviointimenetelmistä. Kyselyssä esitettiin seuraavat kysymykset:

- Nykyinen ammatti ja työnimike
- Oppilaitos ja linja tai suoritettujen opintojen suuntaus
- Kokemus asiantuntija-arvioinneista ajallisesti ja määrällisesti
- Kokemus käytettävyydesteistä ajallisesti ja määrällisesti.

Tutkimukseen osallistuneiden arvioijien taustatiedot esitetään kappaleessa 4.2 (ks. Taulukko 1 sivulla 26).

#### 4.3.3.2 Kuvaus asiantuntija-arvioinnista

Noin viikkoa ennen arvioinnin suorittamista osallistujilta pyydettiin vapaamuotoista kuvausta asiantuntija-arvioinnista ja sen suorittamisesta. Kysymys esitettiin muodossa:

- Kerro omin sanoin, mitä käytettävyyden asiantuntija-arviointi tarkoittaa ja miten se tehdään?

Kuvauksen avulla haluttiin selvittää, miten kokeneiden ja kokemattomien arvioijien näkemykset asiantuntija-arvioinnista eroavat toisistaan. Lisäksi kuvauksen avulla oli mahdollista tutkia, noudattaako arvioijien tutkimuksessa suorittama arviointi heidän omaa näkemystään menetelmästä. Kun lisäksi arvioinnin jälkeisessä haastattelussa arvioijia pyydettiin kuvailemaan arviointiaan jälkeensä, oli tutkimuksessa mahdollista muodostaa ketju: miten arviointi ajatellaan tehtävän, miten se todellisuudessa tehdään ja miten oma arviointi nähdään sen suorittamisen jälkeen.

#### 4.3.3.3 Havaintojen kirjaaminen

Arvioijia pyydettiin kirjaamaan ylös heidän havaitsemansa käytettävyysoingelmat. Tutkimuksessa arvioijille ei annettu tueksi heuristisia sääntöjä, listoja tai muita apuvälineitä. Näin haluttiin selvittää arvioijan omaan näkemykseen perustuvaa asiantuntija-arvioinnin tekemistä niin, että tietyt säännöt eivät ohjaisi tai rajaisi sitä. Todellisuudessa arviointi tehdään usein omaan kokemukseen nojautuen (Nielsen 1993; Riihiahho 2000). Heuristiikoista on todettu olevan hyötyä arvioinnin tekemisessä niin kokemattomille (Sears 1997) kuin kokeneillekin arvioijille (Bastien & Scapin 1995). Toisaalta sääntöjen merkitystä kokeneille arvioijille on myös kyseenalaistettu (Karat ym. 1992). Tässä tutkimuksessa haluttiin keskittyä arvioijan oman kokemuksen vaikutukseen, joten säännöt ja muut apuvälineet jätettiin arvioinnista pois.

Käytettävyysoingelman käsitettä ei tietoisesti määritelty, vaan tämä jätettiin tutkimuksen tulosten asiaksi: mitä arvioijat pitävät käytettävyysoingelmina, millaisia kokonaisuuksia he muodostavat ja mitä osa-alueita he arvioinnissaan käsittelevät. Käytettävyysoingelman käsitettä ei ole tarkasti määritelty, mikä oli nähtävissä yhdeksi mahdolliseksi arvioijien välisiä eroja tuottavaksi tekijäksi (ks. 2.4.1 ja 2.4.3). Esimerkiksi Jeffries ym. (1991) määrittivät tutkimuksessaan käytettävyysoingelman olevan mikä tahansa käyttäjää häiritsevä asia, oli se sitten rakenteellinen ongelma tai väärin kirjoitettu sana.

Havainnot pyydettiin kuvaamaan otsikkotasolla, mikä käytännössä tarkoitti yhtä virkettä. Usein havaintoja on tarpeellista perustella tai kuvata tarkemmin, mutta tutkimuksessa tarkempien kuvausten kirjoittaminen ei ollut ajallisesti mahdollista. Perustelut pyydettiin antamaan suullisesti arvioinnin aikana. Ongelmakuvauksilla on merkittävä rooli tehtyjen havaintojen esittämisessä ja välittämisessä eteenpäin. Esimerkiksi Capra (2006) on nostanut esiin kuvausten laadun tärkeyden. Tässä tutkimuksessa kuitenkin tärkeintä oli saada

kuvaukset ylös siten, että ongelmia voitiin kategorisoida ja analysoida myöhemmin. Lisäksi kiinnostus kohdistui siihen, muotoilevatko arvioijat ongelmien otsikoita ylipäänsä esitettävään muotoon, vai jäävätkö ne omien muistiinpanojen tasolle.

Tehdyt havainnot kirjattiin ylös erillisellä kannettavalla tietokoneella, joka sijoitettiin varsinaisen arviointikoneen viereen samalle pöydälle. Tämän järjestelyn avulla arvioitava palvelu saattoi olla koko arvioinnin ajan näkyvillä varsinaisen testikoneen näytöllä. Neliosaisen arvioinnin viimeisessä tehtävässä havainnot pyydettiin esittämään vain suullisesti; kirjoittamiselle ei olisi ollut juuri aikaa ja toisaalta näin saatiin myös selvitettyä, mikä on havaintojen kirjaamisen vaikutus arviointiprosessiin ja arvioinnin tulosten laatuun.

Havaintojen vakavuusluokitusta ei pyydetty tekemään arvioinnin aikana. Tämän tarkoituksena oli selvittää, kuuluuko vakavuusluokituksen antaminen arvioijien luonnolliseen arviointiprosessiin. Arvioinnin päätteeksi arvioijia pyydettiin kuitenkin luokittelemaan havaintonsa kolmiportaisella vakavuusasteikolla, jossa 1 tarkoitti lievintä ja 3 vakavinta käytettävyysongelmaa. Skaala esitettiin vain numeroina, jotta sanallisten vakavuustasojen erilainen tulkinta ei vaikuttaisi luokitteluun. Lisäksi skaala laadittiin tarkoituksella varsin kapeaksi, jotta luokitus olisi kevyempi tehdä ja toisaalta myös erot tuloksissa olisivat selkeämpiä. Joissakin tutkimuksissa vakavuusluokituksen muodostaminen on määritelty tarkemmin. Esimerkiksi Jeffries ym. (1991) määrittelivät tutkimuksessaan ongelman vakavuuden sen vaikutuksen ja yleisyyden suhteen, sekä sen perusteella, kuinka moni käyttäjä siihen käytön aikana törmää. Tämän kaltaista määrittelyä kuitenkin vältettiin tietoisesti tässä tutkimuksessa: vakavuusluokitukseen ei annettu tarkempia ohjeita, vaan se jätettiin arvioijien itsensä pääteltäväksi.

Tehdyille havainnoille ei pyydetty antamaan arvioinnin aikana kehitysehdotuksia, sillä tutkimuksessa haluttiin selvittää, nousevatko korjausehdotukset luonnostaan esiin tehtyjen havaintojen rinnalla vai onko niiden tuottaminen selkeästi oma prosessinsa. Lisäksi käytettävissä olleen ajan rajallisuus ei oikein olisi antanut mahdollisuutta järjestelmälliselle korjausehdotusten tuottamiselle. Kunkin arviointitehtävän päätteeksi arvioijia kuitenkin pyydettiin mainitsemaan kolme keskeisintä korjausta, minkä tarkoituksena oli selvittää, miten arvioijat priorisoivat kehitysehdotuksia ja toisaalta sitä, miten korjausehdotukset vastaavat arvioinnin aikana tehtyjä havaintoja. Korjausehdotukset ovat arvioinnin kannalta mielenkiintoisia siksi, että pelkkien löydettyjen käytettävyysongelmien sijaan niitä voidaan pitää tärkeämpänä arvioinnin lopputuloksena (esim. Karat 1997). Toisaalta taas arviointimenetelmät eivät tarjoa systemaattista keinoa korjausehdotusten laatimiseen (mm. Nielsen & Molich 1990).

Arviointitehtävän päätteeksi korjausehdotusten lisäksi arvioijia pyydettiin antamaan kouluarvosana (asteikolla 4-10) juuri arvioidulle sivulle tai kokonaisuudelle. Arvosanojen tarkoituksena oli selvittää yleisellä tasolla arvioijien subjektiivinen näkemys arvioitujen osien käytettävyydestä, jota voitaisiin verrata varsinaisiin arvioinnin tuloksiin. Oletuksena oli, että arvioinnin tulosten ja arvosanan välillä olisi korrelaatio.

Arvioijien otsikkotasolla kirjaamien havaintojen perusteella kirjoittaja selvitti tehtyjen havaintojen määrää ja laatua, havaintojen käsittelemiä osa-alueita, havaintojen abstraktiotasoa sekä havaintojen vakavuusluokitusta. Kirjattujen havaintojen analysoinnista kerrotaan tarkemmin luvussa 4.4.2.



#### 4.3.3.4 Ääneen ajattelu ja arvioinnin seuraaminen

Arvioijia pyydettiin ajattelemaan ääneen arviointia suorittaessaan. Ääneen ajattelun avulla voitiin tulkita luotettavammin arvioijan ajatteluprosessia: mitä asioita ja mitä näkökulmia hän pohtii arvioinnin aikana, sekä miten hän arvioinnissaan etenee. Ääneen ajattelu toi tärkeän komponentin kokonaisuuden tarkasteluun, johon sen lisäksi kuuluivat toiminnan seuraaminen arvioinnin aikana, arvioinnin tulosten tarkastelu sekä katseenseuranta. On huomattava, että ääneen ajattelu saattaisi tehdä arviointitilanteesta epäluontevan, mutta ääneen ajattelu oli kuitenkin kaikille koehenkilöille tuttua, koska sitä käytetään usein käytettävyytsteissä käyttäjän toiminnan tulkitsemiseen.

Arvioijia pyydettiin perustelemaan tehdyt havainnot sekä tarkemmat havaintojen kuvaukset suullisesti tutkimuksen aikana, koska aikaa perustelujen kirjoittamiseen ei olisi ollut. Ääneen ajattelun avulla saatiin lisäksi kerättyä arvioinnin seuraamisen tueksi tietoa arvioijien arviointiprosessista ja arvioinnin aikana tapahtuvasta ajatustyöstä. Lisäksi ääneen ajattelun avulla saatiin esiin havainnot, joita arvioijat eivät syystä tai toisesta kirjanneet ylös, vaikka pohtivatkin niitä ääneen. Arvioijia ei pyydetty kirjaamaan ylös havaintoihin liittyviä korjausehdotuksia, mutta mikäli ehdotuksia oma-aloitteisesti esitettiin suullisesti, ne saatiin kerättyä ääneen ajattelun avulla.

Havaintojen tekemisessä ja kirjaamisessa kiinnostavaa oli myös arvioijien varmuus tekemistään havainnoista ja mahdollisista korjausehdotuksista. Varmuudesta saatiin kerättyä tietoa ääneen ajattelun ja sen analysoinnin kautta.

Oma-aloitteisen ääneen ajattelun lisäksi arvioijille esitettiin kysymyksiä arvioinnin aikana tilannekohtaisesti. Kysymykset liittyivät tilanteisiin, joissa arvioijat selkeästi pohtivat jotain asiaa tarkemmin. Arvioijien itsensä esittämiin kysymyksiin ei kuitenkaan vastattu paitsi tutkimuksen ohjaamisen osalta. Arvioijien annettiin tehdä arviointiaan itsenäisesti puuttumatta siihen mitenkään. Tutkimuksessa tallennettiin ääni, jotta ääneen ajattelua voitiin analysoida jälkepäin tallenteiden avulla.

Ääneen ajattelun lisäksi arvioinnin aikana seurattiin arviointiprosessia ja muita tilanteessa havaittavia seikkoja, koska tallenteelle päätyi äänen lisäksi vain kuvaruutukaappaus testikoneelta. Kirjoittaja teki käsin muistiinpanoja arvioinnin aikana seurattavista asioista.

#### 4.3.3.5 Silmänliiketutkimus

Arvioinnin aikana käytettiin käytettävyyslaboratorion katseenseurantalaitteistoa, eli arvioijan silmänliikkeet tallennettiin arvioinnin aikana. Silmänliiketutkimuksen tarkoituksena oli kerätä tietoa siitä, miten arvioijien katse ja havaitseminen etenee arvioinnin kuluessa. Silmänliiketutkimusta käytettiin sekä ryhmäkohtaiseen että yksilökohtaiseen tarkasteluun. Toisaalta tarkoituksena oli vertailla eri ryhmien välisiä eroja arvioinnin etenemisessä tutkittavan käyttöliittymän alueiden ja elementtien suhteen, toisaalta tuoda yksittäisen arvioijan koetallenteen analysointiin lisätietoa hänen katseensa kohdistumisen avulla.

Ryhmäkohtainen tarkastelu vaati ryhmän jäsenten silmänliikedatan yhdistämistä ja yhdistetyn datan analysointia kootusti. Ryhmien vertailu onnistui siis vertailemalla molempien ryhmien koottuja tietoja keskenään. Ryhmäkohtaisesti selvitettiin ja vertailtiin eri ryhmien havaitsemisen etenemistä ja erilaisten kokonaisuuksien tunnistamista arvioinnin aikana. Silmänliiketutkimuksen tulokset kertovat nopeasti, jos joitakin elementtejä tai osia ei järjestelmässä ole havaittu, tai mihin huomiota on kiinnitetty eniten. Yksittäisten arvioijien osalta katsepolun seuraaminen videolta tukee kuvaruutukaappauksen ja ääneen ajattelun

analysointia. Sen avulla voidaan havaita, miten arvioija käy arvioitavaa sivua läpi ja missä suhteessa katseen eteneminen ja kohdistuminen näkyy ääneen ajatteluun ja havaintojen kirjaamiseen verrattuna.

Silmänliiketutkimuksen avulla kerättiin tietoa seuraavista asioista:

- Mitä elementtejä ja asioita arvioijat sivuilta havaitsivat, mitä he taas eivät havaitsivat?
- Katsooko arvioija kaikkia palvelun osia vai keskittyykö hän joihinkin asioihin?
- Missä järjestyksessä ja miten kattavasti arvioitavan sivun havaitseminen etenee ajan kuluessa?
- Onko elementtien ja asioiden havaitsemisessa ja havaitsemisen etenemisessä eroja ryhmien välillä?

Katseenseurantalaitteen käyttö näkyi arvioijille käytännössä vain erikoisrakenteisena näyttölaitteena (ks. Kuva 7 sivulla 33), joten se ei aiheuttanut suurta eroa tyypilliseen tietokoneen avulla tehtävään arviointiin. Tavallisesta tietokoneesta ja arviointitilanteesta poikkeavaa oli vain istunnon aluksi tehty laitteen kalibrointi sekä istumisasento, johon oli kiinnitettävä huomiota vähän enemmän huomiota. Laitteen toiminta perustuu henkilön pupillien tunnistamiseen ja rekisteröimiseen, mikä on yksilöllistä. Tästä johtuen laite oli kalibroitava jokaiselle laitteen käyttäjälle erikseen. Käytännössä kalibrointi tarkoitti noin minuutin aikana eri puolille näyttöä yksitellen ilmestyvien pisteiden seuraamista katseella. Näytön alareunassa olevat anturit rekisteröivät pupilleista heijastuvaa infrapunavaloa, minkä vuoksi kalibroinnin lisäksi henkilön silmien oli pysyttävä jokseenkin samalla etäisyydellä ja korkeudella näyttöön nähden koko arvioinnin ajan. Kullekin arvioijalle sopiva ja luonteva asento etsittiin heti arvioinnin alussa jo ennen kalibrointia, joten ei tämä kuitenkaan johtanut epäluonnolliseen tai -miellyttävään työskentelyasentoon. Ajoittain arvioinnin aikana ohjaajan oli kuitenkin pyydettävä arvioijaa korjaamaan hieman asentoaan, jos ryhti tai istuma-asento muuttui ratkaisevasti. Kalibroinnin ja istuma-asennon lisäksi katseenseurantalaitteen käyttö ei käytännössä vaikuttanut tutkimustilanteeseen millään muulla tavalla.

Silmänliiketutkimus on käytettävyytutkimuksessa suosittu menetelmä käyttäjän toiminnan seuraamiseen (Salvucci & Anderson 2001). Käytettävyyden asiantuntija-arvioinnin tutkimiseen sitä ei aiemmissa tutkimuksissa juuri ole sovellettu. Tästä johtuen silmänliiketutkimus on tämän tutkimuksen kannalta kiinnostava osa. Silmänliiketutkimus tuo kuitenkin tutkimukselle myös selviä haasteita: katseenseurantalaitteen käyttöön liittyy teknisiä haasteita, lisäksi tiedonkeruu ja analysointi on työlästä ja vaativaa (Laarni 2004). Vaikka silmänliikedata tallennettiin kaikilta arvioijilta kaikissa arviointitehtävissä, sen tarkastelu rajattiin analysointivaiheessa varsin suppeaksi (ks. 4.4.3). Syynä tähän oli juuri analyysin työläys, mutta toisaalta myös se, että mahdolliset muuttujat haluttiin pitää vähäisinä, kun kyseessä oli käytännössä uusi silmänliiketutkimuksen soveltamiskohde.

#### 4.3.3.6 Haastattelu

Arvioinnin aikana arvioijille esitettiin tarvittaessa ja harkiten arvioijan toimintaan liittyviä lyhyitä kysymyksiä, joilla pyrittiin täsmentämään arvioijien toimintaa ja havaintoja. Lisäksi arvioinnin päätteeksi suoritettiin varsinainen loppuhaastattelu. Laadullisen haastattelun tehtävänä oli suhteuttaa koeasetelma (ekologinen validiteetti) sekä selvittää arvioijan omia näkemyksiä tehdystä arvioinnista ja tukea siten muilla tavoilla kerättyä tietoa. Haastattelun avulla oli myös mahdollista verrata arvioijan omaa käsitystä arvioinnistaan toteutuneeseen arviointitapaan sekä asiantuntija-arvioinnin kuvaukseen, joka arvioijalta pyydettiin ennen tutkimusta (ks. 4.3.3.2).

Haastattelussa esitettiin seuraavat kysymykset:

- Kuvaile käyttämäsi arviointitapaa; miten arviointi eteni?
- Miten arviointi erosi tavanomaisesta arvioinnista?
- Miten kattavasti uskot löytäneesi ongelmat tuotteista?
- Miten kattavasti uskot löytäneesi erityisesti vakavat ongelmat?
- Miten rajoitettu aika vaikutti arviointiin?
- Miten arvioitujen osien kontekstin puute vaikutti arviointiin?
- Mitä tarkempia menetelmiä käytit arvioinnissa?
- Mitkä menetelmät soveltuisivat tällaisille tuotteille?

#### 4.3.3.7 Eettiset näkökulmat

Tutkimukseen osallistuvat arvioijat joutuivat koetilanteeseen, jossa heidän toimintaansa tarkkailtiin ja analysoitiin. Arvioijat tiesivät, että huomio ja kiinnostus tutkimuksessa painottui juuri heidän kykyynsä toimia arvioijina, mikä lisäsi merkittävästi koetilanteeseen liittyvää jännitystä. Tämä oli huomioitava koetilanteen johdatuksessa, ohjeistuksessa sekä kerätyn datan käsittelyssä. Arvioijille luotiin ilmapiiri, jossa he eivät olisi tunteneet oloaan vaivaantuneeksi tai epämiellyttäväksi kokeen aikana. Lisäksi heille korostettiin, että arvioijia käsitellään tutkimuksessa ryhmänä, ei yksittäisinä henkilöinä.

Tutkimuksessa kerätty tieto käsiteltiin luottamuksellisesti ja niin, että yksittäiset arvioijat ja heidän työnsä eivät ole myöhemmin tuloksista tunnistettavissa. Pääosin tuloksia käsiteltiin ryhmänä, joihin arvioijat oli jaettu heidän kokemuksensa perusteella. Niissäkin tapauksissa, joissa tuloksia oli tarpeen nostaa esiin arvioijakohtaisesti, samaan yhteyteen ei liitetty mitään sellaisia tietoja, joiden avulla henkilö olisi ollut tunnistettavissa. Arvioijista ei jäänyt tutkimuksen julkisen osan tuloksiin mitään tietoja, joista heidät voisi tunnistaa.

Eettiseltä kannalta tilannetta helpotti se, että kaikki arvioijat olivat vapaaehtoisia ja halukkaita osallistumaan tutkimukseen. Vapaaehtoisuudesta huolimatta koetilanne oli monille jännittävä ja siihen tuli siksi suhtautua asianmukaisesti.

## 4.4 Kerätyn tiedon analysointitavat

Tutkimuksessa kerätty tieto oli luonteeltaan sekä laadullista että määrällistä. Laadullinen tieto käsitti pääasiassa arviointitapaan ja arviointiprosessiin liittyvät tiedot, määrällinen puolestaan enimmäkseen arvioinnin tuloksiin, eli tehtyihin havaintoihin liittyvät tiedot.

### 4.4.1 Ääneen ajattelu ja arvioijien antamat kuvaukset

Arviointitilanteet oli tallennettu, jolloin ääneen ajattelun avulla kerätty tieto oli käytössä tallenteiden avulla. Tallenteet olivat videomuodossa, ja äänen lisäksi niillä näkyi arviointikoneen kuvaruutukaappaus sekä katseenseurantalaitteen tallentamat arvioijien katsepolut (ts. miten katse siirtyy ja mihin se kiinnittyy).

Tallenteet litteroitiin kokonaisuudessaan siten, että arvioijan ääneen ajattelu ja mahdollinen dialogi tutkimuksen ohjaajan kanssa kirjoitettiin ylös sanasta sanaan. Erityistä formaattia litteroinnille ei kuitenkaan käytetty, vaan puhe kirjoitettiin sellaisenaan tekstinkäsittelyohjelmalla. Litteroinnin aikana tarkkailtiin myös tallennettua kuvaruutukaappausta sekä silmänliikkeitä, joita molempia käytettiin litteroinnin tukena. Muistiinpanoina kirjoitettiin ylös myös sellaiset olennaiset seikat, jotka eivät pelkän ääneen

ajattelun avulla olisi tulleet esiin (esim. arvioija puhuu jostakin käyttöliittymän osasta, mutta ei mainitse eksplisiittisesti, mistä on kyse).

Kaikkien arvioijien tallenteiden litteroinnin jälkeen litterointeja analysoimalla ja läpikäymällä niistä kerättiin paperilapuille tutkimuksen kannalta mielenkiintoisia havaintoja. Ylös kirjattiin kaikki jollakin tavalla olennaiset ja kiinnostavat tilanteet, kuten esimerkiksi:

- Arvioija kertoo ääneen, kuinka etenee arvioinnissaan.
- Arvioija perustelee kirjaamiaan havaintoja, tai sitä, miksi jättää jonkin kirjaamatta.
- Arvioija antaa korjausehdotuksia tekemilleen havainnoille.
- Arvioijan työskentelyssä näkyvät erityisiä tekniikoita tai menetelmiä.
- Arvioija osoittaa epävarmuutta, varmuutta, virhetulkintoja tai kykyä erityisen analyttisiin johtopäätöksiin.

Ääneen ajattelun lisäksi samalla tavalla käsiteltiin myös muu laadullinen tieto; arvioijien omin sanoin ennen arviointia antamat kuvaukset asiantuntija-arvioinnista sekä loppuhaastattelun vastaukset, joissa arvioijat kertoivat arvioinnin jälkeen arvioinnin tekemisestä ja sen etenemisestä. Lisäksi arvioijien kuvauksia arvioinnista verrattiin loppuhaastattelun vastauksiin sekä kirjoittajan näkemykseen todellisesta arvioinnin etenemisestä. Tarkoituksena oli tarkastella, miten hyvin arvioijien kuvaukset vastaavat heidän tekemiään arviointeja, sekä sitä, miten todenmukainen kuva heillä arvioinnin päätteeksi oli omasta työskentelystään.

Kaikki kiinnostavat havainnot edellä mainituista asioista kerättiin yksitellen värillisille tarralapuille. Havaintojen jäsentelyyn ja luokitteluun käytettiin affiniteettidiagrammia. Yksittäisistä tarralapuista muodostettiin isolle seinälle ryhmiä niiden yhteenkuuluvuuden ja keskinäisten suhteiden perusteella. Lappuja siirrettiin useaan otteeseen, niistä muodostuneita ryhmiä yhdisteltiin ja joitakin ryhmiä jaettiin erilleen. Lappujen järjestelyn lähtökohtina käytettiin useita eri näkökulmia, kuten arviointitapaa, arvioinnissa tehtyjen havaintojen muodostumista, arvioinnissa käytettyjä tekniikoita sekä joitakin muita arviointia koskevia piirteitä.

Yksittäisten havaintojen ryhmittelyn lopputuloksena oli useita lappuryhmiä, joista kukin käsitteli tiettyä teemaa. Ryhmät nimettiin näiden teemojen perusteella. Tämän jälkeen kustakin ryhmästä muodostettiin kokonaisnäkemys: affiniteettidiagrammista johdettiin havaintokokonaisuuksia, ilmiöitä sekä niiden perusteella johtopäätöksiä, jotka on esitetty tuloksina kappaleessa 5. Lopullinen affiniteettidiagrammi on esitetty seuraavassa kuvassa (ks. Kuva 8).



**Kuva 8.** Havaintojen ryhmittelyn perusteella muodostettu kaksiosainen affiniteettidiagrammi.

#### 4.4.2 Arvioinnissa kirjattujen havaintojen tarkastelu

Arvioijien otsikkotasolla (ilman perusteluja ja tarkempia kuvauksia) kirjaamia havaintoja tarkasteltiin määrällisesti useista eri näkökulmista. Syynä tähän oli se, että monissa aiemmissa tutkimuksissa oli pureuduttu pääsääntöisesti pelkkien ongelmien absoluuttisten määrien vertailuun analysoimatta havaintojen luonnetta ja kokoa tarkemmin (mm. Nielsen 1992; Desurvire ym. 1991; Desurvire ym. 1992). Muiden muassa Sears (1997) nosti esille, että kaikkien ongelmien määrän hahmottaminen on hankalaa, ja ongelma pitäisi määritellä tarkemmin. Myös Hertzum ja Jacobsen (2003) painottavat ongelmakriteeristön tärkeyttä, mutta myös sanovat sen laatimisen olevan vaikeaa, sillä ongelmat liittyvät aina tuotteeseen, käyttäjän tehtäviin sekä arvioinnin tavoitteisiin. Ongelmien määrittely nousee merkitykselliseksi myös siksi, että arvioijien tuloksissa on tyypillisesti suuria yksilöllisiä eroja (ks. 2.4.3).

Arviointimenetelmien aikaisimmissa tutkimuksissa käytettiin yksinkertaisia sääntölistoihin ja ongelmien vakavuusluokitukseen perustuvia luokitteluperiaatteita (mm. Nielsen & Molich 1990; Karat 1994). Tämän jälkeen havaintojen analysointiin on tutkimuksissa käytetty myös joitakin ongelmien osa-alueisiin perustuvia luokittelutapoja. Slavkovic ja Cross (1999) kategorisoivat ongelmia sovelluksen osion perusteella. Myöhemmin esimerkiksi Hvannberg ja Law (2003) kehittivät ohjelmistokehityksen menetelmiin ja virheiden tarkasteluun perustuvan määrittely- ja luokittelumenetelmän käytettävyysongelmille. Chattrachart ym. (2003) muodostivat puolestaan ongelmaprofiilit, jotka perustuvat käyttöliittymän eri osa-alueisiin (sisältö, navigointi, grafiikka, toimintojen tehokkuus, muotoilu ja layout, kieli, apu ja virheilmoitukset).

Keenan ym. (1999) totesivat, että aiemmin on tutkittu ongelmien vakavuusluokitusta ja korjauskustannuksia, jotka auttavat tuotteiden kehittäjiä tuotteiden parantamisessa, mutta ne eivät tue ongelmien korkean tason globaalia analyysia. Toisin sanoen ongelmien

luokitteluun, vertailuun ja analysointiin ei ole olemassa selvää kehystä. Tämän vuoksi he kehittivät taksonomian käytettävyysohjelmien käsittelyyn. Globaali tarkastelu on tärkeää myös siksi, että usein kokonaisuus, joka arvioinnissa tulisi huomioida, on laajempi kuin yksi näkökulma, jota arvioinnissa monesti kerrallaan tarkastellaan (Dumas & Redish 1994). Artikkelissa (Keenan ym. 1999) esitetty taksonomia perustuu käytettävyysohjelmien kategorisointiin toisaalta artefaktien, toisaalta käyttötilanteen tehtävien perusteella. Taksonomia auttaa tarkastelemaan eri osa-alueita, löytämään korjauksia laajoille ongelmajoukoille sekä tarkastelemaan ongelmia globaalisti.

Edellä mainittu taksonomia, kuten muutkaan edellisten tutkimusten luokittelutavat eivät kuitenkaan tarjoa keinoa ongelmien abstraktiotasojen tarkasteluun: mitä tarkoittaa yksi käytettävyysohmelmä, kuinka yksityiskohtaisesta tai laajasta kokonaisuudesta se voi muodostua? Tämä oli kiinnostava ja merkittävä näkökulma tämän tutkimuksen kannalta, jotta arvioijien kirjaamia havaintoja voitiin ylipäänsä määrällisesti verrata.

Tässä tutkimuksessa havaintojen analysointi aloitettiin laskemalla arvioijien tekemien havaintojen absoluuttiset kokonaismäärät, joita vertailtiin arviointitehtävittäin kokoneiden ja kokemattomien ryhmien välillä. Tämän jälkeen havaintoja tarkasteltiin arvioijien niille tekemän vakavuusluokituksen kautta: erikseen analysoitiin, miten havainnot erosivat ryhmien välillä vakavuusasteittain, sekä miten erot vakavuusasteissa näkyivät eri arviointitehtävien välillä.

Seuraavaksi arvioijien tekemiin havaintoihin pureuduttiin tarkemmin niiden tyyppien ja laajuuksien luokittelun avulla. Kirjoittaja laati analysointivaiheessa kategoriat yksittäisten havaintojen luokittelua varten sekä tarkasteltaville käytettävyyden osa-alueille (eri näkökulmat) että havaintojen abstraktiotasoille. Tämän luokittelun tarkoituksena oli saada arvioijien tekemistä havainnoista keskenään vertailukelpoisia sekä analysoida tarkemmin, millaisia ongelmakokonaisuuksia arvioijat ja eri ryhmät todellisuudessa tarkastelevat. Hypoteesina oli, että varsinkin havaintojen abstraktiotasot saattavat vaihdella yksilöllisesti<sup>8</sup>, jolloin pelkkien absoluuttisten määrien vertailu olisi ollut ongelmallista ja harhaanjohtavaa. Käytettävyysohjelmien abstraktiotasojen kategorisointiin analysointivaiheessa käytettiin seuraavia kirjoittajan laatimia luokkia:

- **Yksittäiset havainnot:** tarkat, yhteen elementtiin tai alueeseen liittyvät havainnot ja muut yksityiskohdat
- **Keskisuuret ongelmat:** useampaan sivun tai palvelun yksittäiseen asiaan tai niiden suhteisiin liittyvät ongelmat
- **Suuret ongelmakokonaisuudet:** koko sivustoa, palvelua tai sen luonnetta koskevat ongelmat
- **Useammasta yksittäisestä havainnosta muodostetut kokonaisuudet:** yhdeksi ongelmaksi kirjatut, useammasta erillisestä havainnosta koostuvat kokonaisuudet.

Havaintojen analysointiin osa-alueittain käytettiin kategorisointia, joka on samankaltainen edellä mainitun taksonomian (Keenan ym. 1999) artefaktiluokittelun kanssa. Kategorisointi muokattiin kuitenkin paremmin tämän tutkimuksen tarpeita vastaavaksi. Taksonomia on suunnattu tuotteiden kehittämisen ja ongelmien korjaamisen tueksi, mutta tässä tutkimuksessa tarkoituksena oli saada eri havainnoista keskenään vertailukelpoisia suorituksen tutkimisen kannalta. Tästä syystä kategorisointi tehtiin hieman eri tavalla. Tässä

---

<sup>8</sup> Esimerkiksi yksi arvioija saattaa kirjata erillisinä havaintoina useita samaan kokonaisuuteen liittyviä yksityiskohtia, kun toinen taas mahdollisesti kirjaa ylös havaintoja yhden laajemman ongelmakokonaisuuden, joka kattaa nämä kaikki yksityiskohdat.

tutkimuksessa käytettiin havaintojen osa-alueiden tarkastelussa seuraavaa kirjoittajan laatimaa luokitusta:

- **Konsepti:** arvioitavan sivun tai koko palvelun tarkoitus, ja sen suhde toteutukseen
- **Tekstisisältö:** varsinainen sisältö, termistö ja kieliasu
- **Rakenne:** sivun asettelu (layout), sivun eri osat ja niiden keskinäiset suhteet
- **Toiminnot:** käyttäjälle tarjottavat yksittäiset toiminnot ja käyttöliittymäkontrollit
- **Navigointi:** liikkuminen palvelussa, linkit ja etenemispainikkeet
- **Visuaaliset huomiot:** visuaaliset elementit, värit, muodot, kuvat.

Kirjattujen havaintojen kanssa analysoitiin myös korjausehdotuksia, jotka arvioijat antoivat kunkin arviointitehtävän päätteeksi. Arvioijia pyydettiin esittämään kolme heidän mielestään keskeisintä korjattavaa asiaa juuri arvioidusta tuotteesta. Korjausehdotuksia analysoitiin vertaamalla niitä yksitellen kyseisen arvioijan kirjaamiin havaintoihin: kuinka hyvin korjausehdotukset vastasivat arvioijien vakavimmiksi määrittelemiä havaintoja, sekä vastasivatko ne yksittäisiä havaintoja vai korjasivatko ne useamman havainnon käsittelemää laajempaa kokonaisuutta.

Lisäksi kirjattujen havaintojen yhteydessä analysoitiin myös arvioijien antamia kouluarvosanoja kunkin arviointitehtävän kohteena olleelle kokonaisuudelle. Arvosanoja analysoitiin vertaamalla eri ryhmien arvosanojen keskiarvoja kussakin tehtävässä, millä selvitettiin sitä, näkevätkö ryhmät eri osien välisen paremmuuden kokonaisuutena eri tavoin. Lisäksi arvosanoja verrattiin ryhmien keskiarvojen avulla arvioijien tekemään havaintojen vakavuusluokitusten keskiarvoihin arviointitehtävittäin. Havaintoja painotettiin vakavuusasteiden perusteella luoduilla kertoimilla siten, että arvioitu kokonaisuus sai sitä pienemmän vertailuluvun, mitä enemmän ongelmia ja vakavia ongelmia siihen liittyi. Tämän vertailun tarkoituksena oli selvittää korreloivatko yksittäiset havainnot ja niiden vakavuudet arvioijille muodostuvan kokonaiskuvan kanssa.

#### 4.4.3 Silmänliiketutkimuksen data

Tutkimuksessa tallennettiin katseenseurantalaitteen avulla arvioijien silmänliikedata kaikissa neljässä arviointitehtävässä. Tarkempaan analyysiin otettiin kuitenkin mukaan vain ensimmäisen ja toisen tehtävän data. Silmänliiketutkimuksen data ja sen analysointi haluttiin pitää rajattuna toisaalta tutkimuksen lukuisten muiden näkökulmien ohella, toisaalta silmänliiketutkimuksen uuden sovelluskohteen vuoksi.

Silmänliiketutkimusta ei aiemmin ole sovellettu kokeneiden ja kokemattomien arvioijien työskentelyssä, joten tässä tutkimuksessa haluttiin saada tällä lähestymistavalla ensimmäisiä – mahdollisimman selviä tuloksia mahdollisimman vähillä muuttujilla. Silmänliiketutkimuksen yhtenä keskeisenä haasteena on työläs ja vaikea analyysi (Salvucci & Anderson 2001, Laarni 2004). Erityisesti tarkasteltavien näkymien suuri määrä on erityinen haaste tulosten analysoimiseksi (Goldberg, Stimson, Lewenstein, Scott & Wichansky 2002). Goldberg ym. käyttivät tutkimuksessaan nimenomaan useita näkymiä, joiden välillä käyttäjä sai liikkua vapaasti.

Tässä työssä silmänliikeanalyysi rajattiinkin kahteen ensimmäiseen tehtävään siksi, että niissä arvioitiin yksittäisiä sivuja, eikä analyysissä tarvinnut huomioida liikkumista sivujen välillä ja sen tuomaa kompleksisuutta. Käytännössä silmänliiketutkimuksen analyysi on usein epäformaalia (Salvucci & Anderson 2001). Salvucci ja Anderson (2001) kehottavat käyttämään formaaleja analyysimenetelmiä ja useiden eri tietojen käyttämistä silmänliikedatan ohella analyysissä. Tässä työssä kuitenkin päätarkoitus oli yksinkertaisesti

selvittää, onko eri arvioijaryhmien välillä eroja silmänliikkeissä, ei niinkään tulkita mahdollisia eroja ja niiden syitä tarkemmin. Tästä syystä formaalia analyysiä ei tässä työssä käytetty.

Silmänliikeanalyysi perustuu tarkkoihin pisteisiin, joihin ihminen kohdistaa katseensa tiettyä kohdetta tarkastellessaan. Ihmisen tarkan näön alue on hyvin kapea, se kattaa vain noin kahden asteen kokoisen sektorin katseen suuntaan. Tarkan näön ulkopuolisilta alueilta ihminen voi havaita muutosta, muotoja ja liikettä, mutta tarkkaa havaitsemista siellä ei tapahdu. Tästä johtuen katsetta on siirrettävä kohteesta paikasta toiseen yksityiskohtien havaitsemiseksi. Ihminen kerää visuaalista tietoa *fiksaatioiden* aikana, joilla tarkoitetaan katseen kiinnittymistä tiettyyn pisteeseen tyypillisesti joidenkin sekunnin kymmenyksien ajaksi. Nopeita siirtymiä fiksaatioiden välillä kutsutaan puolestaan *sakkadeiksi*. Sakkadien aikana tiedon rekisteröintiä ei juuri tapahdu. Silmänliikkeet voivat ohjautua käytännössä kahdella eri tavalla: ihmisen etsiessä kohteesta tiettyä asiaa (*top-down*) sekä kohteen elementtien ja yksityiskohtien ohjatessa katsetta (*bottom-up*) (Goldberg ym. 2002).

Yksinkertaisesti silmänliikkeitä tulkitaan siten, että se, mitä kohteessa parhaillaan katsotaan, vastaa sitä, mitä sillä hetkellä myös ajatellaan. Analysoinnissa olisi kuitenkin aina otettava huomioon koehenkilön tavoite, toisin sanoen se, mitä olisi tarkoitus havaita ja tarkastella. Katseen kiinnittyminen tiettyyn pisteeseen voi tarkoittaa kohteen olevan mielenkiintoinen, poikkeava tai tärkeä tehtävän kannalta (Loftus & Mackworth 1978). Pitkä fiksaatio merkitsee sitä, että henkilön on tulkittava näkemäänsä tarkoin (Goldberg & Kotval 1998). Tämä voi kertoa siitä, että henkilöllä on hankaluuksia tulkita näkemäänsä, mutta toisaalta myös sitä, että henkilö on ymmärtänyt näkemänsä ja pohtii sitä tarkemmin.

Tässä tutkimuksessa silmänliikedatan analysointiin ryhmiä vertailtaessa käytettiin lämpökartoista (*hot spot*) muodostettuja kuvasarjoja. Lämpökartat esittävät painotetusti fiksaatioiden osumisen sivulle: mitä enemmän tiettyä pistettä on katsottu, sitä punaisempana piste esitetään. Kohdat, joihin katsetta ei ole osunut lainkaan, näkyvät harmaina. Keskittyminen fiksaatioiden tarkasteluun ryhmiä vertailtaessa oli tässä tutkimuksessa järkevää tutkimuksen rajaamisen vuoksi. Lisäksi ryhmien vertailu katsepolkujen (fiksaatioiden lisäksi myös sakkadien) tarkastelun avulla olisi vaatinut datan tarkastelua arvioijakohtaisesti, mikä ei olisi ollut tämän tutkimuksen kannalta mielekästä. Lämpökarttojen avulla oli lisäksi mahdollista yhdistää usean arvioijan datat ryhmäksi, ja siten vertailla kahta eri ryhmää keskenään ryhmien kokonaisuuden avulla. Ne toivat analyysiin myös havainnollisen visuaalisen esitystavan, kun tarkasteltavasta tuotteesta ei ole etukäteen määritelty erityisiä kiinnostuskohteita (*area of interest*), joihin katseen osumista voitaisiin analysoida puhtaasti määrällisesti. Katsepolkuja käytettiin kuitenkin yksittäisten arvioijien videotallenteiden analysoinnissa arvioijien toiminnan seuraamisen ja ääneen ajattelun tukena. Katsepolut näkyivät kokonaisuudessaan videotallenteilla arvioinnin alusta loppuun.

Tutkimuksen koehenkilöille ei annettu tarkempia ohjeita arvioinnissa etenemiseen, vaan havaitseminen ja eteneminen perustuivat heidän yksilöllisiin tapoihinsa ja prosesseihinsa. Koska tarkempia tavoitteita ei etukäteen ollut, ohjasi tutkittava tuote ainakin aluksi heidän havaitsemistaan (*bottom-up*). Tämä on silmänliiketutkimuksen kannalta ratkaisevaa siksi, että toimintaa ohjaavien tehtävien osuus katseen ohjautumisessa on erittäin suuri (Goldberg ym. 2002). On toki mahdollista, että he mielessään loivat arvioinnin aikana itselleen tavoitteita ja tehtäviä, jolloin havaitseminen perustui tiettyjen asioiden etsimiseen (*top-down*). Joka tapauksessa aiemmin on todettu, että yhden sivun sisällä liikkuminen johtaa katseen ohjautuvaan hakuun (Goldberg ym. 2002). Tutkimuksessa ei myöskään käytetty hypoteeseja fiksaatioiden mahdollisesta jakaantumisesta, joten mielekkäämpää oli silmänliiketutkimuksen osalta edetä analyysissä kerätystä datasta lähtien, jolloin lämpökartat tarjosivat luontevan mahdollisuuden tulkita dataa visuaalisesti.



Fiksaatioita voidaan tarkastella lämpökarttojen avulla kolmella eri tavalla:

- Absoluuttinen kesto: kokonaisaika, jonka henkilöt ovat käyttäneet eri pisteiden katsomiseen
- Suhteellinen kesto: aika, jonka henkilöt ovat käyttäneet eri alueiden katsomiseen suhteessa koko kohteeseen
- Fiksaatioiden määrä: katseen jakautuminen eri pisteisiin riippumatta niiden kestosta (asettaa kaikki fiksaatiot samanarvoisiksi niiden pituuksista riippumatta).

Tässä tutkimuksessa tarkasteluun valittiin ensisijaisesti fiksaatioiden määrä, jolla saatiin parhaiten tarkasteltua sitä, mitä asioita kohteesta on ylipäänsä katsottu ja mitä ei. Toisena analysointimuotona käytettiin fiksaatioiden suhteellista kestoa. Tämä oli järkevää siksi, että eri arvioijat käyttivät arviointiin hieman eripituisia aikoja annetusta maksimijasta huolimatta. Lisäksi suhteellinen tapa oli hyödyllinen ajan kuluessa pidemmälle, sillä fiksaatioiden määrä sivulla kasvoi väistämättä hyvin suureksi.

Analysissä tarkasteltiin lämpökarttoja arvioinnin eri ajan hetkillä. Kullakin hetkellä tarkasteltiin rinnakkain kahta lämpökarttaa: kokoneiden arvioijien (6 arvioijaa) yhdistetty silmänliikedata sekä vastaavasti kokemattomien arvioijien (6 arvioijaa) yhdistetty data. Tarkasteluun otettiin lämpökartat 2, 4, 8, 15 ja 30 sekunnin sekä 1, 2, 3, 4 ja 10 minuutin kohdalla. Varhaisimmat lämpökartat kertovat pääosin katseen tiedostamattomasta ohjautumisesta ja ensivaikutelmasta. On huomattava, että katseen ohjautumiseen sivun alussa saattaa vaikuttaa myös arvioijan itsensä asettama päämäärä ja tavoite, jolloin tilanne ei vastaa esimerkiksi todellisen käyttäjän saapumista sivulle. Myöhäisemmät lämpökartat puolestaan kertovat enemmän siitä, miten katseen kohdistuminen jakaantuu arvioinnin edetessä, kun arvioijat käyvät läpi ja tarkastelevat sivun eri osia enemmän tietoisesti.

Silmänliiketutkimuksen dataa analysoitaessa on kuitenkin muistettava, että vaikka katseen kohdistuminen tiettyyn pisteeseen tarkoittaa kyseisen kohdan havaitsemista, se ei kerro kuitenkaan luotettavasti kyseisen kohdan tulkitsemisesta tai ymmärtämisestä arvioinnin aikana. Tässä analyysissä pitäydettiin siinä, mitä elementtejä ja asioita arvioijat missäkin vaiheessa katsovat, eikä tämän pohjalta pyritty tekemään laajempia johtopäätöksiä tulkitsemisesta.

## 5 VERTAILUN TULOKSET

---

Tässä luvussa esitellään tutkimuksen empiirisestä osuudesta saadut tulokset. Ensin tarkastellaan määrällisesti arvioinnin aikana kirjattuja käytettävyyso ongelmia, sen jälkeen laadullisesti arviointiprosessia, ja lopuksi silmänliiketutkimuksen tuloksia.

### 5.1 Arvioinnissa tehdyt havainnot

Tässä kappaleessa esitellään määrällisen analyysin tulokset arvioijien kirjaamista käytettävyyso ongelmista. Ongelmia tarkastellaan useasta eri näkökulmasta; kirjattujen ongelmien kokonaismääränä, ongelmien vakavuusluokituksen kautta, ongelmia koskevien osa-alueiden kannalta sekä ongelmien abstraktiotasojen kannalta. Ongelmien vakavuusluokitus on kunkin arvioijan itse antama. Muut mainitut luokittelutavat ovat kirjoittajan analyysivaiheessa laatimia.

On huomattava, että kirjatut käytettävyyso ongelmat eivät täysin vastaa kaikkia arvioinnin aikana havaittuja ongelmia, sillä eri syistä molemmilla ryhmillä joitakin ääneen mainittuja ongelmia jäi kokonaan kirjaamatta ylös. Tästä johtuen tässä kappaleessa käsitellään vain kirjattuja ongelmia puuttumatta muihin havaittuihin ongelmiin. Ratkaisu on perusteltu myös käytännön kannalta: kirjaamatta jääneet ongelmat eivät välttämättä päädy arviointiraporttiin eivätkä arvioinnin tilaajan tietoon.

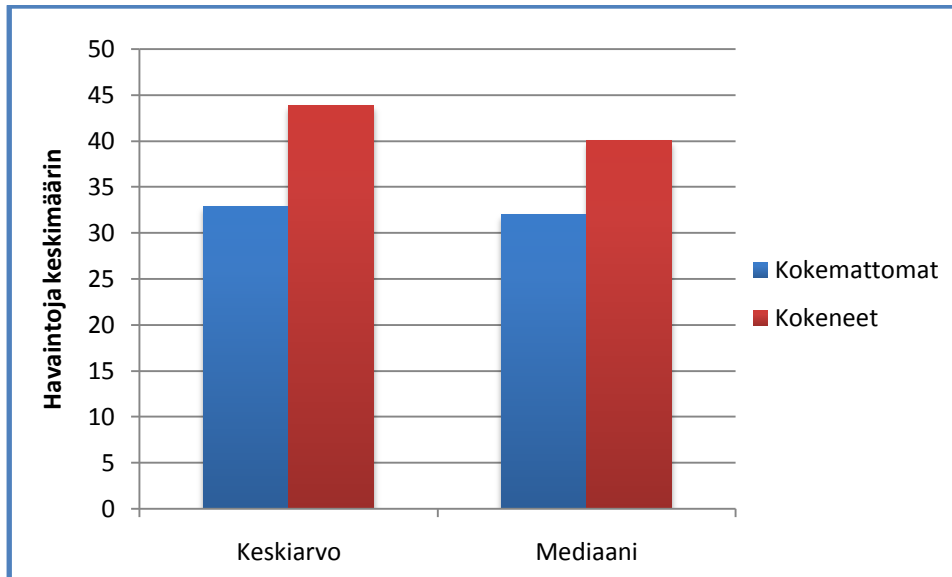
Kokeneiden ja kokemattomien ryhmiin kuului molempiin seitsemän arvioijaa, kun varsinaisten arvioijien lisäksi mukaan laskettiin myös pilottiarvioijien kirjaamat havainnot<sup>9</sup> (määrällisessä analyysissä N=7 ryhmää kohden). Ongelmien vakavuusluokitus jäi puuttumaan yhdeltä pilottiarvioijalta, minkä vuoksi myös toisesta ryhmästä jätettiin pilottiarvioijan vakavuusluokitus huomioimatta. Tästä johtuen vakavuusluokitusta tarkasteltaessa N=6 molempia ryhmiä kohden.

#### 5.1.1 Kirjattujen havaintojen absoluuttinen määrä

Arvioinnin eri osat yhteen laskettuna kirjattujen havaintojen määrä vaihteli aloittelevilla arvioijilla välillä 14–54, kokeneilla asiantuntijoilla puolestaan välillä 24–76. Kokonaisuudessaan kokeneet asiantuntijat kirjasivat enemmän ongelmia (307) kuin kokemattomat (230). Yksi kokenut arvioija kirjasi keskimäärin 43,9 ongelmaa (mediaani 40), kokematon puolestaan 32,9 (mediaani 32). (ks. Kuva 9.)

---

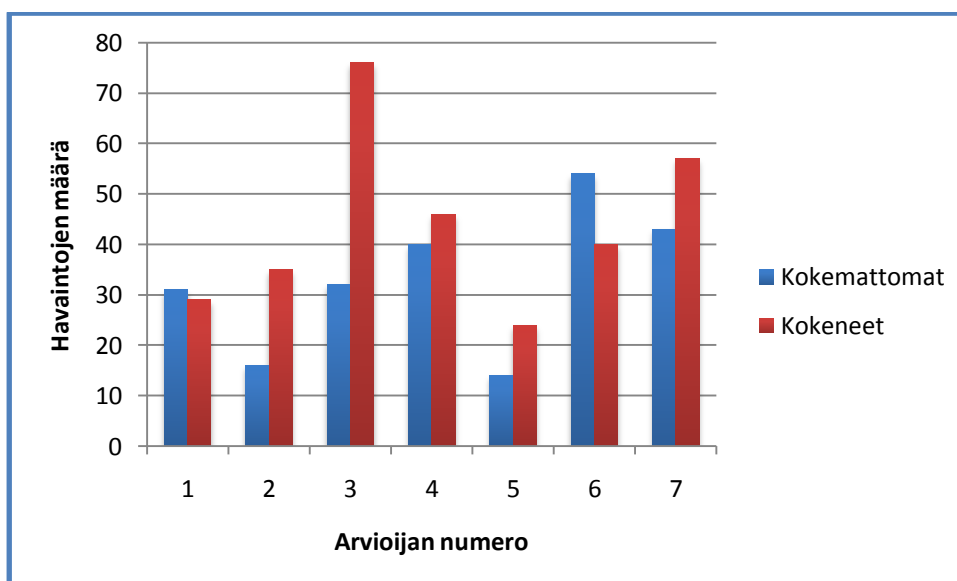
<sup>9</sup> Vakavuusluokitusta ja korjausehdotusten antamista lukuun ottamatta koeasetelmaan ei tullut pilottiarviointien jälkeen muutoksia, jotka olisivat vaikuttaneet arviointiin, tai erityisesti havaintojen kirjaamiseen.



Kuva 9. Arvioinnissa kirjatut ongelmat keskimäärin ryhmäkohtaisesti esitettynä.

Ero kirjattujen ongelmien absoluuttisessa määrässä on selvä, mutta ei kuitenkaan tilastollisesti merkitsevä. Aiempien tutkimusten perusteella (mm. Nielsen 1992; Desurvire 1994) eron olisi voinut olettaa suuremmaksi; esimerkiksi Nielsenin tutkimuksessa ero oli tilastollisesti merkitsevä. Lisäksi on huomattava, että eroon vaikuttaa ainakin osittain ongelmien vaihteleva abstraktiotaso; pienempi määrä abstraktiotasoltaan suurempia ongelmakokonaisuuksia voi olla käytännössä arvokkaampi kuin suurempi joukko yksityiskohtia (ks. 5.1.4).

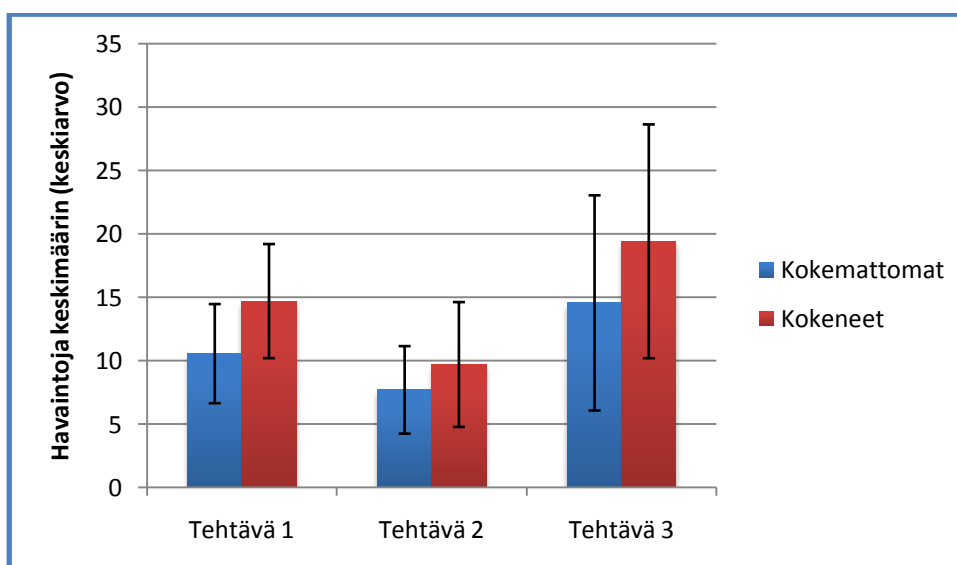
Asiantuntija-arviointiin suositellaan useita asiantuntijoita, koska ongelmien löytämisessä on yksilöllisiä eroja (mm. Nielsen & Molich 1990). Sama on nähtävissä myös tämän tutkimuksen perusteella. Yksittäisten arvioijien kirjaamien ongelmien tarkastelu paljastaa määrissä suuria eroja. Yksilölliset erot näkyvät arvioijakohtaisesti oheisessa kaaviossa (ks. Kuva 10) sekä keskihajonnan avulla ryhmäkohtaisissa tuloksissa (ks. Kuva 11). Hieman yllättävää on, että yksilölliset erot ovat kokeneiden arvioijien välillä kokemattomiin verrattuna suurempia, eli toisin sanoen kokemattomat arvioijat ovat arvioinnissaan ongelmien määrällä mitattuna ryhmän sisällä tasaisempia.



Kuva 10. Kirjatut ongelmat arvioijakohtaisesti.

Muodostettaessa yksittäisistä arvioijista esimerkiksi kolmen arvioijan ryhmiä havaitaan, että jokin kolmen kokemattoman arvioijan joukko saattaa päästä yhtä hyvään tai parempaan tulokseen kuin jokin tietty kolmen kokeneen ryhmä. Satunnaisesti valituissa ryhmissä tämä on kuitenkin epätodennäköistä. Yksilölliset erot näkyvät kokonaismäärän lisäksi myös ongelmien osa-alueiden (ks. 5.1.3) ja abstraktiotasojen (ks. 5.1.4) tarkastelussa.

Tarkasteltaessa kirjattujen ongelmien määriä ryhmien välillä tehtäväkohtaisesti havaitaan eron näkyvän kokeneiden arvioijien eduksi jokaisen arvioidun tehtävän kohdalla. Tosin minkään tehtävän kohdalla ero ei ole tilastollisesti merkitsevä. Tulos on kuitenkin uskottava, sillä molemmille ryhmille trendi on yhtenäinen: useamman sivun kokonaisuudesta (tehtävä 3) löytyy eniten ongelmia, kahden yksittäisen sivun välillä taas tehtävässä 1 epäyhtenäisestä ja sekavasta sivusta enemmän kuin tehtävässä 2. Suhteellisesti ero on suurimmillaan juuri tehtävässä 1, mikä on ehkä hieman yllättävää, sillä selvien ja näkyvien ongelmien keräämisen voisi ajatella olevan helppoa kokemattomallekin arvioijalle. Kokeneet arvioijat kuitenkin pääsivät arvioinnissaan selvästi näkyviä ongelmia syvemmälle, minkä vuoksi kokonaismääräkin oli suurempi. Sama selitys pätee jossain määrin toki myös muihin arviointitehtäviin. (ks. Kuva 11.)



**Kuva 11.** Ongelmien määrä tehtävittäin yhtä arvioijaa kohden (keskiarvo). Kaavio esittää keskihajonnat pystyjanojen avulla.

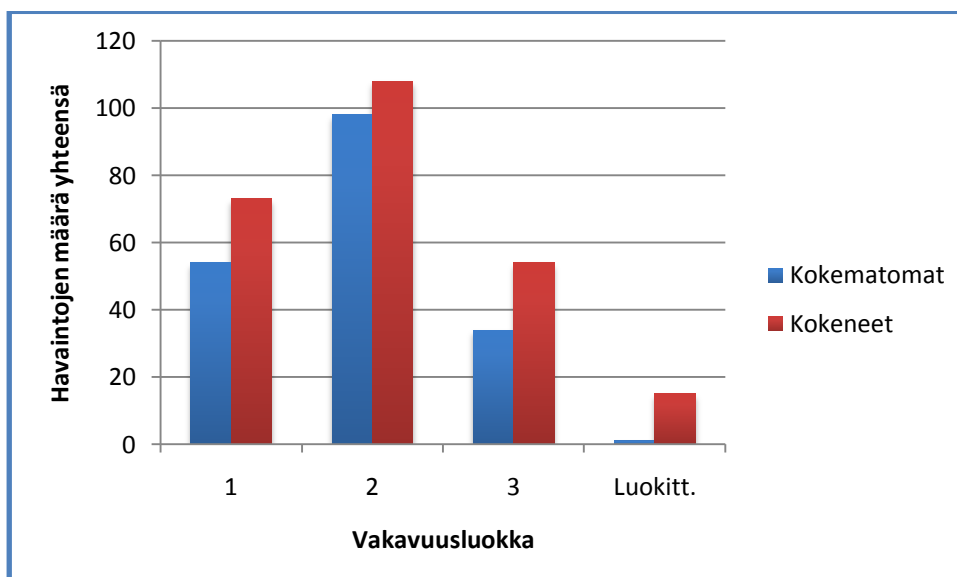
### 5.1.2 Havaintojen vakavuusluokitus

Arvioijia pyydettiin luokittelemaan kirjaamansa ongelmat kolmiportaisella asteikolla (1–3). Vaikka sanallisia kuvauksia luokille ei annettu, käytännössä voidaan kuitenkin puhua lievistä (1), merkittävistä (2) ja vakavista (3) ongelmista erittelemättä näiden merkitystä tarkemmin. Arvioijille annettiin myös lupa jättää luokitus antamatta niille ongelmille, joiden vakavuutta ei tarjolla olevan tiedon avulla voitu tai osattu määrittää.

Vakavuusluokituksia tarkastelemalla havaitaan molemmille ryhmille yhtenäinen trendi: merkittäviä ongelmia on kirjattu eniten, lieviä toiseksi eniten ja vakavia vähiten (ks. Kuva 12). Kaikissa vakavuusluokissa kokeneiden arvioijien kirjaamia ongelmia on enemmän kuin kokemattomilla. Suhteellisesti suurin ero on nähtävissä vakavien ongelmien kohdalla. Aiemmissa tutkimuksissa on todettu, että vakavimmat ongelmat löytyvät arvioinnissa varmimmin (Nielsen 1992; Jacobsen ym. 1998). Toisaalta Desurvire (1994) on havainnut, että kokemattomat eivät löydä vakavia ongelmia yhtä tehokkaasti. Tässä tutkimuksessa saatu tulos kertoo, että kokemattomilla arvioijilla oli vaikeuksia löytää juuri vakavampia ongelmia.

On huomattava, että ilmiö voi toki johtua myös siitä, että kokemattomat arvioijat eivät mahdollisesti luokittelle tekemiään havaintoja yhtä herkästi vakaviksi ongelmiksi. Vaikka havaintojen osa-alueilla ja abstraktiotasoilla ei olekaan suoraa yhteyttä vakavuusluokituksiin, tukee niiden tarkastelu kuitenkin sitä, että laajemmat ja vaikuttavammat ongelmat jäisivät kokemattomilta arvioijilta kirjaamatta selvästi kokeneita useammin (ks. 5.1.3 ja 5.1.4).

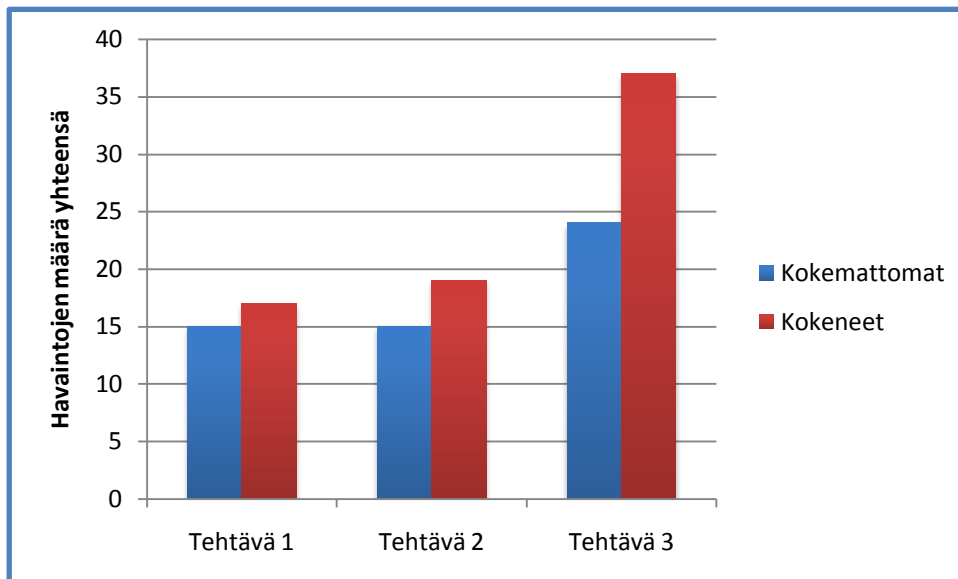
Lisäksi ääneen ajattelun avulla kerätyn laadullisen aineiston perusteella on havaittavissa se, että kokemattomat arvioijat saattavat jättää joitakin tekemiään keskeisiä havaintoja kirjaamatta epävarmoina siitä, ovatko nämä ylipäänsä ongelmia. Itse asiassa kokemattomien arvioijien ryhmän tuloksissa luokittelemattomia ongelmia oli vain yksi, kun vastaavasti kokeneilla arvioijilla niitä oli kaikkiaan 15. Kokeneet arvioijat jättivät vakavuuden määrittelemättä kahdesta eri syystä: joko he eivät olleet varmoja ongelman varsinaisesta vaikutuksesta käyttäjän toimintaan ja sitä kautta todellisesta vakavuudesta tai sitten he eivät puutteellisesta kontekstista johtuen tienneet, onko kyseessä ylipäänsä ongelma. Ensimmäisessä mainitussa tapauksessa kokemattomat arvioijat puolestaan sijoittivat ongelman johonkin luokkaan, jälkimmäisessä taas jättivät herkemmin koko ongelman kirjaamatta.



Kuva 12. Kirjattujen ongelmien määrä yhteensä vakavuusluokittain.

Vakavuusluokitusten tarkastelu tehtäväkohtaisesti paljastaa jonkin verran arvioitavan tuotteen luonteen vaikutuksesta eri vakavuusasteisten ongelmien tunnistamiseen. Toisena arvioidusta sivusta kokemattomat arvioijat saivat kerättyä lieviä ja merkittäviä ongelmia lähes saman verran kuin kokeneet, mutta vakavien ongelmien kohdalla syntyi selvä ero.

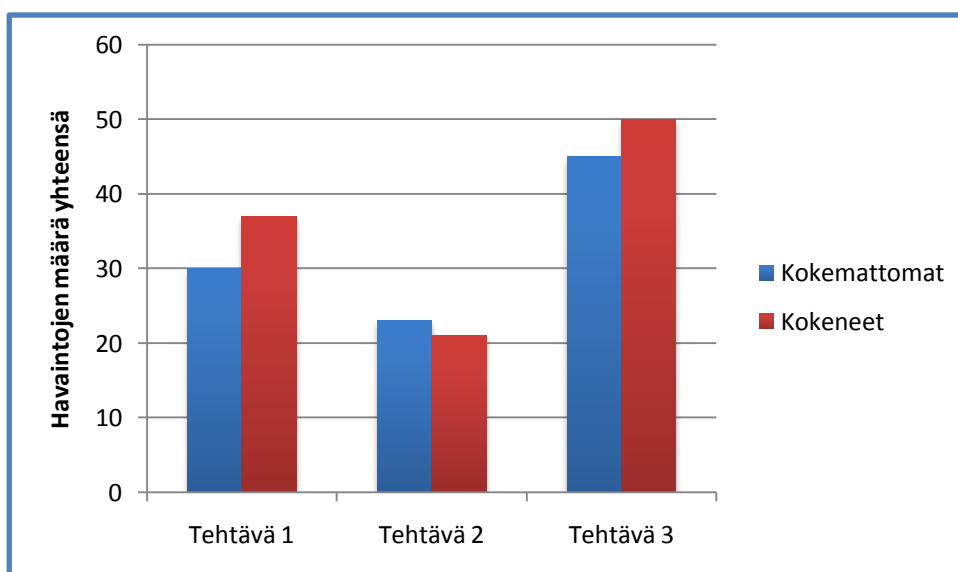
Lievien ongelmien kohdalla erot eivät ole suuria – itse asiassa kahden yksittäisen sivun välillä (tehtävät 1 ja 2) eroa lievien ongelmien määrässä ei juuri syntynyt. Myös eri arviointiryhmien väliset erot näissä tehtävissä olivat melko pieniä, vaikka kokeneet kirjasivatkin molemmissa kokemattomia enemmän havaintoja. Ainoastaan kolmannen arviointitehtävän kokonaisen tilausprosessin kohdalla kokeneet kirjasivat lieviä ongelmia puolet kokemattomia enemmän. (ks. Kuva 13.)



**Kuva 13.** Lievien ongelmien määrä tehtäväkohtaisesti.

Merkittävien ongelmien osalta huomioitavaa on se, että toisen tehtävän yksittäisen sivun kohdalla kokemattomat arvioijat kirjasivat havaintoja aavistuksen kokeneita enemmän, vaikka ero onkin käytännössä olematon. Molemmat ryhmät kirjasivat merkittäviä ongelmia ensimmäisessä tehtävässä toista tehtävää enemmän. Kiinnostavaa on, että kokeneet arvioijat kirjasivat havaintoja ensimmäisessä tehtävässä lähes kaksinkertaisesti toiseen verrattuna. (ks. Kuva 14.)

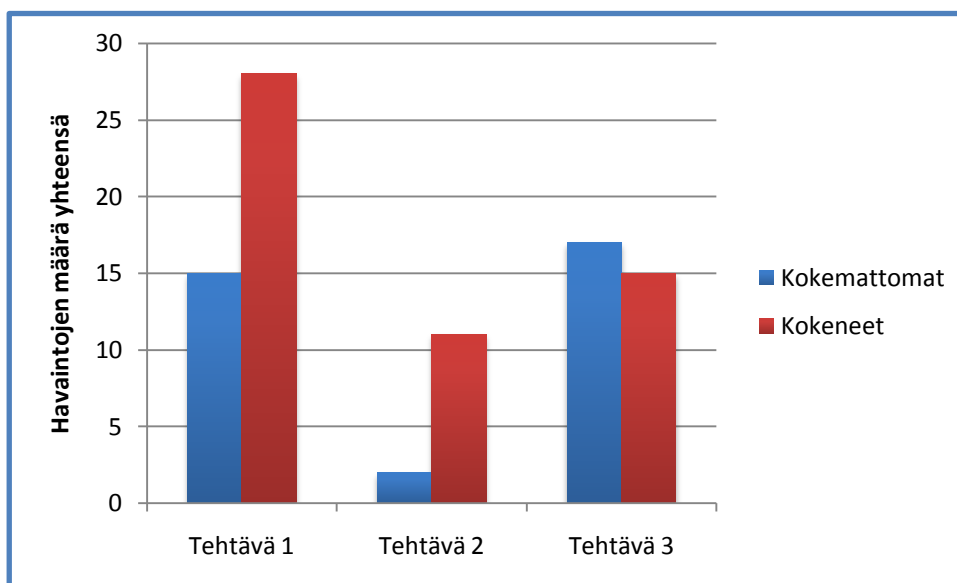
Merkittävien ongelmien määrän pohjalta johtopäätösten tekemistä hieman hankaloittaa se, että kyseinen luokka on käytetyn vakavuusluokituksen puolivälissä, johon arvioijan on tavallaan turvallista sijoittaa ongelma ollessaan epävarma. On toki mahdollista, että kokeneet arvioijat uskalsivat varmuudestaan johtuen käyttää skaalaa vapaammin, mikä voisi selittää ryhmien eroja merkittävien ongelmien määrässä kahden ensimmäisen tehtävän välillä. Joka tapauksessa tehtävän 2 yksittäinen sivu oli selkeästi ensimmäistä parempi, jollaisena sitä myös arvioijat pitivät. Tämä näkyi etenkin vakavien ongelmien, mutta kokeneilla myös merkittävien ongelmien määrässä.



**Kuva 14.** Merkittävien ongelmien määrä tehtäväkohtaisesti.

Vakavia ongelmia molemmat ryhmät kirjasivat suunnilleen saman verran viimeisen tehtävän tilausprosessin arvioinnissa – kokemattomat arvioijat itse asiassa jopa muutaman enemmän. Yksittäisistä sivuista (tehtävät 1 ja 2) vakavien ongelmien löytäminen oli kuitenkin kokemattomille selvästi vaikeampaa (ks. Kuva 15). Ilmiöön voisi vaikuttaa myös se, että kokemattomat eivät mahdollisesti luokittelisi ongelmia vakaviksi yhtä herkästi kuin kokeneet. Tähän kysymykseen kuitenkin tämä tutkimus ei vastaa.

Tehtävän 2 arvioinnissa kokemattomat arvioijat kirjasivat yhteensä vain kaksi vakavaa ongelmaa, kun kokeneet puolestaan 11. Ero on selvästi havaittava ja aivan tilastollisen merkitsevyyden rajalla. Tehtävän 1 osalta vastaava ero oli myös selvä; 15 vakavaa ongelmaa kokemattomille ja 28 kokeneille. Kokemattomien arvioijien kirjaamien vakavien ongelmien lähes olematon määrä tehtävässä 2 johtunee siitä, ettei heillä riittänyt kykyä pureutua syvemmälle sinänsä melko hyvin suunnitellun sivun arvioinnissa. Näkyvien ja selkeästi esiin nousevien ongelmien puute sai monet kokemattomat hyytymään arvioinnissaan, etenkin sirpaleisen ja sekavan ensimmäisen sivun jälkeen, josta ongelmia oli helppo kerätä ylös (ks. myös 5.2.1.2).



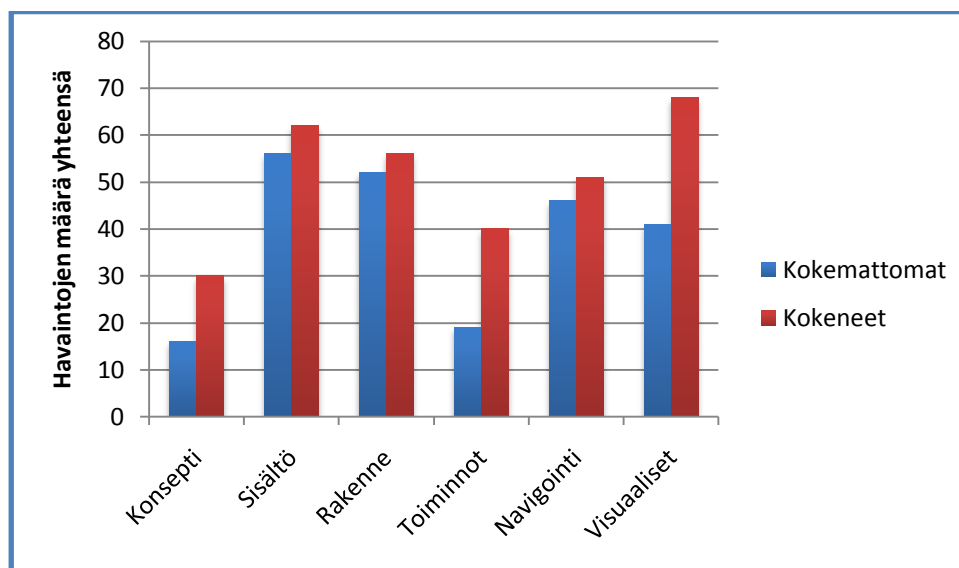
Kuva 15. Vakavien ongelmien määrä tehtäväkohtaisesti.

Vakavuusluokituksen osalta yksilölliset erot olivat molemmilla ryhmillä suurimmat lievissä ongelmissa. Tulosten perusteella näyttääkin siltä, että noin puolella arvioijista on tapana priorisoida ongelmia, toisin sanoen jättää vakavuudeltaan vähäisempiä ongelmia kirjaamatta (tai koota niitä suuremmiksi kokonaisuuksiksi, ks. 5.1.4). Osa arvioijista taas kirjaa niitä järjestelmällisesti yksitellen. Toisaalta tämä selittyy myös sillä, että eri asiantuntijat ovat herkempiä havaitsemaan ongelmia eri osa-alueilta, ja esimerkiksi yksittäiset visuaalisen viimeistelyn puutteet saattavat olla sellaisinaan vaikutukseltaan vähäisempiä.

### 5.1.3 Arvioinnin osa-alueet ja näkökulmat

Arvioijien kirjaamat ongelmat analysoitiin tarkemmin jakamalla ne eri luokkiin arvioinnin osa-alueiden ja näkökulmien perusteella. Kirjoittaja luokitteli ongelmat analysointivaiheessa, eli arvioija ei ohjattu luokittelemaan tai ryhmittelemään havaintojaan mitenkään. Käytetty luokittelu on esitelty kappaleessa 4.4.2.

Kokeneet arvioijat kirjasivat kaikilla osa-alueilla ongelmia absoluuttisena määränä kokemattomia enemmän, mutta tekstisisältöä, rakennetta ja navigointia koskevien ongelmien määrässä erot olivat pieniä. Kokemattomat löysivät näillä osa-alueilla 90 – 93 % kokeneiden vastaavista määristä. Muilla kolmella osa-alueella erot olivat hyvin selviä. Konseptia koskevia ongelmia kokemattomat arvioijat kirjasivat 53 %, toimintoja koskevia ongelmia 48 % ja visuaalisia huomioita 60 % kokeneiden arvioijien vastaavista määristä. Erot ongelmien määrässä eivät ole tilastollisesti merkitseviä, tosin toimintoja koskevien ongelmien kohdalla tämä on erittäin lähellä. (ks. Kuva 16.)



Kuva 16. Ongelmien määrä yhteensä osa-alueittain eriteltynä.

Koska sisältöön, rakenteeseen ja navigointiin liittyvien ongelmien määrässä kokemattomat pääsivät lähimmäksi kokeneita, voidaan heidän ajatella olevan tehokkaimmillaan näillä osa-alueilla. Konseptin arviointi oli kokemattomille arvioijille erityisen hankalaa arvioitaessa yksittäisiä sivuja tehtävissä 1 ja 2. Kokemattomat kirjasivat näissä tehtävissä konseptia koskevia ongelmia yhteensä vain muutaman. Tässä yhteydessä on syytä huomata myös kappaleessa 2.1.1 esitetty ero käytettävyyden määritelmässä. Kirjattaviin ongelmiin voi toki vaikuttaa se, pitääkö arvioija ylipäänsä esimerkiksi konseptia käytettävyyden arviointiin kuuluvana asiana; ISO:n määritelmään (ISO 9241-11) se kuuluu, mutta Nielsenin määritelmään taas ei (Nielsen 1993).

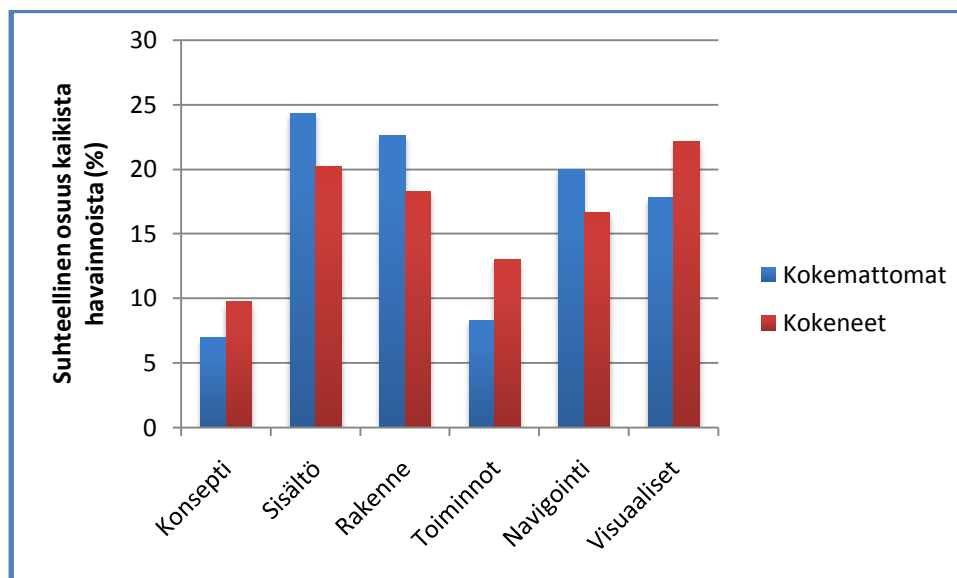
Toimintojen kohdalla taas kolmannen tehtävän monivaiheisen prosessin arviointi aiheutti suurimmat erot, koska prosessissa korostui toimintojen merkitys. Suurehkoa eroa visuaalisissa huomioissa voidaan pitää hieman yllättävänä, sillä usein visuaaliset virheet ja ratkaisut osuvat helposti arvioijan silmään, jolloin kokemuksen vaikutuksen voisi ajatella pienemmäksi.

Toinen mielenkiintoinen lähestymistapa osa-alueisiin on ryhmien kirjaamien havaintojen kokonaismääriin vertaamalla lasketut suhteelliset määrät. Suhteellinen tarkastelu paljastaa, kuinka kaikki ryhmän kirjaamat havainnot jakautuvat eri osa-alueiden välille. Koska ryhmien löytämien ongelmien kokonaismäärissä oli eroja, kertoo tämä tarkastelutapa ryhmien tekemistä arvioinneista enemmän kuin kokonaismäärät.

Suhteelliset määrät paljastavat, että vaihtelu eri painopisteiden kesken oli kokemattomilla arvioijilla huomattavasti suurempaa kuin kokeneilla. Kokemattomien arvioijien kirjaamista havainnoista noin seitsemän prosenttia koski konseptia, kun taas sisältöön liittyi lähes neljännes. Neljän osa-alueen (sisältö, rakenne, navigointi ja visuaaliset huomioid) suhteelliset



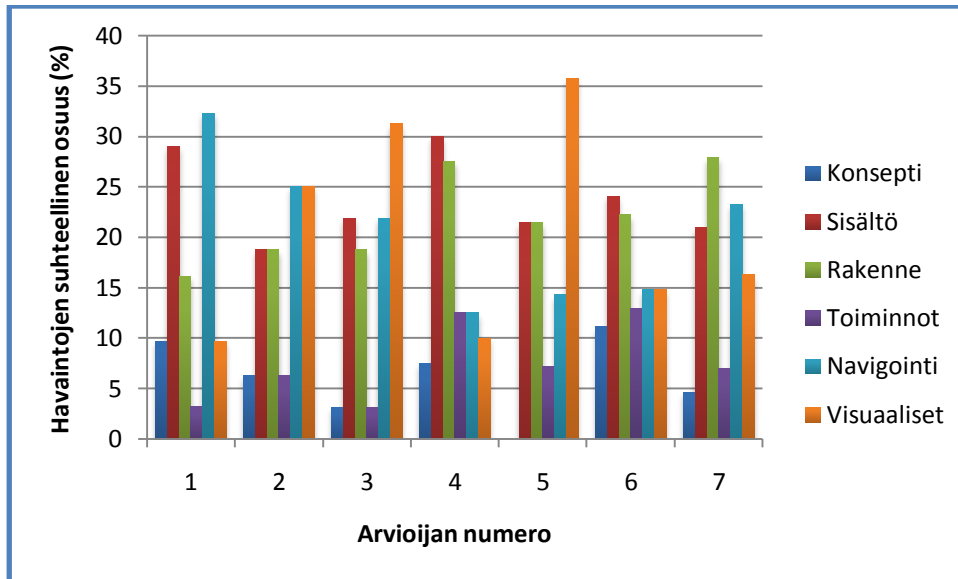
osuudet erottuivat huomattavasti kahdesta muusta. Kokoneiden arvioijien kirjaamien havaintojen jakautuminen suhteellisesti eri osa-alueisiin on selvästi tasaisempaa; vaihtelu mahtui kokonaisuudessaan välille 10 – 22 %. (ks. Kuva 17.)



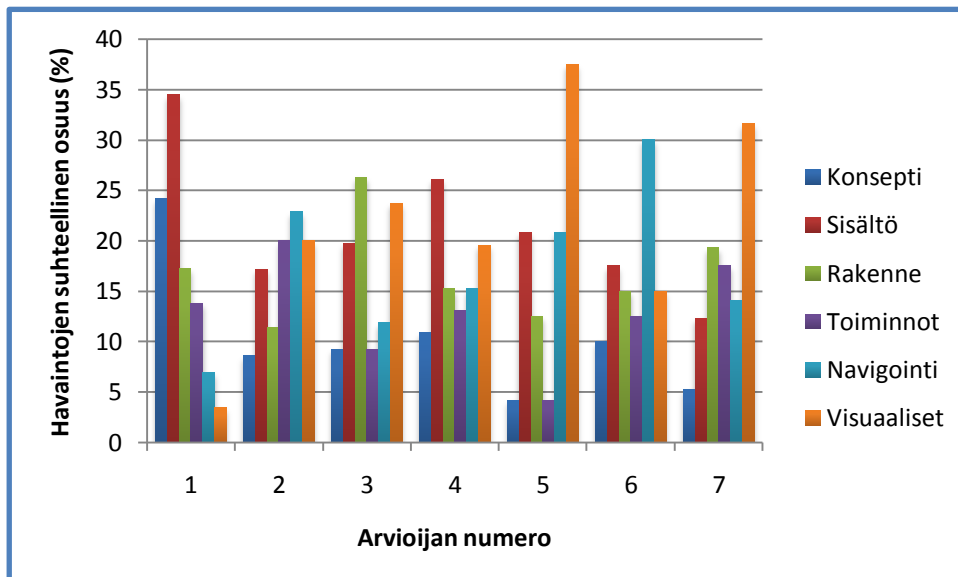
**Kuva 17.** Eri osa-alueita koskevien ongelmien suhteellinen osuus kaikista havainnoista.

On kuitenkin huomattava, että tutkimuksessa ero syntyi kahden seitsenhenkisen ryhmän välille, ja ryhmien sisällä yksilöllistä vaihtelua oli melkoisesti (ks. Kuva 18 ja Kuva 19). Suurista henkilökohtaisista eroista johtuen ei ole itsestään selvää, että kokoneet arvioijien tulokset ovat kattavampia eri osa-alueiden huomioimisella mitattuna, jos arviointiin osallistuvia henkilöitä on vähemmän. Esimerkiksi joillakin tarkkaan valituilla kolmen arvioijan yhdistelmillä kokemattomat olisivat päässeet osa-alueet huomioiden vähintään yhtä kattavaan lopputulokseen, kun lasketaan vain kunkin osa-alueen ongelmien määrät erittelemättä mahdollisia päällekkäisyyksiä tai ristiriitaisuuksia. Joidenkin tiettyjen kolmen arvioijan yhdistelmien kohdalla kokemattomat pääsivät kokoneita parempaan tulokseen jopa niillä osa-alueilla, jotka heille koko ryhmänä osoittautuivat heikommiksi (konsepti, toiminnot, visuaaliset huomiot). Joillakin kolmen kokoneen asiantuntijan yhdistelmillä tietyt osa-alueet korostuivat merkittävästi ja jotkut jäivät puolestaan toisarvoiseen asemaan. Edellä mainitut tilanteet ovat toki kaikki yhdistelmät huomioiden harvinaisia, mutta silti seikka korostaa entisestään useamman asiantuntijan tarvetta arviointiprojektissa riittävän kattavuuden varmistamiseksi.

Kun lasketaan jokaisen arvioijan eri osa-alueita koskevat havainnot suhteessa kaikkiin kyseisen arvioijan kirjaamiin ongelmiin, saadaan selville henkilökohtaiset painopisteet osa-alueittain ongelmien kokonaisuudesta riippumatta. Poimittaessa kustakin osa-alueesta suurin henkilökohtainen osuus havaitaan, että molemmilla ryhmillä kuuden tarkastellun osa-alueen suurin painotus osuu viidelle eri arvioijalle seitsemästä (ks. Kuva 18 ja Kuva 19). Kokemuksesta riippumatta siis ryhmien sisällä on havaittavissa eräänlaista erikoistumista eri osa-alueisiin.



**Kuva 18.** Kokemattomien arvioijien kirjaamien ongelmien suhteellinen jakautuminen henkilökohtaisesti eri osa-alueisiin.



**Kuva 19.** Kokeneiden arvioijien kirjaamien ongelmien suhteellinen jakautuminen henkilökohtaisesti eri osa-alueisiin.

Edellä mainittu tarkastelu on kuitenkin tehty vain suhteellisesti kunkin arvioijan havaintojen jakautumisesta osa-alueittain, eikä tällöin huomioida arvioijan tekemien havaintojen määrää suhteessa muihin arvioijiin. Tämä ulottuvuus saadaan henkilökohtaisten painopisteiden rinnalle suhteuttamalla kunkin arvioijan tiettyä osa-alueetta koskevat havainnot koko ryhmän kyseisestä osa-alueesta tekemiin havaintoihin. Näin tarkasteltuna havaitaan, että kuuden eri osa-alueen suurimmat henkilökohtaiset osuudet jakautuvat kokemattomilla arvioijilla kolmen ja kokeneilla arvioijilla neljän eri arvioijan kesken. Tämä kertoo siitä, että kokemattomien arvioijien joukossa joku arvioijista voi nousta usealla osa-alueella selvästi muita parempaan tulokseen ja erottua kattavamman arvioinnin ansiosta muusta ryhmästä. Kokeneilla arvioijilla määrään suhteutetut tulokset osa-alueittain jakautuvat ryhmän sisällä useammalle arvioijalle. Eli vaikka kokeneiden arvioijien keskinäiset erot arvioinnin painopisteissä säilyvät, tasaantuvat erot arvioinnin kattavuudessa suhteutettaessa osa-alueisiin jakautuminen havaintojen määrään.

Painopisteistä johtuen kaikilla kokemattomilla arvioijilla vaikutti olevan suuri ero heikoimman ja vahvimman osa-alueensa välillä. Useimmiten selvän piikin muodostivat yksi tai kaksi osa-aluetta, kun taas kahdesta kolmesta osa-alueesta ongelmia löytyi vain vähäinen määrä. Vaikka kokeneilla arvioijilla ongelmat jakautuvat tasaisemmin eri osa-alueiden välille, lähes kaikilla heilläkin oli tuloksissa nähtävissä jokin arvioinnin osa-alue selkeästi ylitse muiden. Vähälle huomiolle jääviä osa-alueita ei kuitenkaan ole yhtä selvästi kuin kokemattomilla.

Eri arvioijat ovat tulosten pohjalta luonnollisesti herkempiä havaitsemaan tiettyihin osa-alueisiin liittyviä havaintoja, kun taas jotkut osa-alueet jäivät väistämättä vähemmälle huomiolle. Myös Slavkovic ja Cross (1999) havaitsivat, että kaikki arvioijat ohittivat arvioinnissaan useita kategorioita, mutta olivat tehokkaita joissain kategorioissa. Koska kokeneilla asiantuntijoilla henkilökohtainen vaihtelu eri osa-alueiden välillä on kuitenkin tasaisempaa kuin kokemattomilla, on oletettavaa, että kokeneet arvioijat ovat kuitenkin kyenneet oppimaan myös luonnostaan heikompien osa-alueiden arviointia. Esimerkiksi Nielsen (1992) totesi tutkimuksessaan, että poistumisteiden tarjoamisen heuristiikkaan liittyvät ongelmat olivat hankalimpia löytää. Hänen mukaansa arvioijien pitäisi yksinkertaisesti kiinnittää enemmän huomiota kyseiseen heuristiikkaan. Slavkovic ja Cross puolestaan suosittelevat tärkeimpien osa-alueiden tarkistuslistoja tai eri osa-alueiden suuntaamista eri arvioijille.

#### 5.1.4 Havaintojen abstraktiotaso

Arviointien tuloksia analysoitiin myös arvioijien kirjaamien yksittäisten ongelmien laajuuden perusteella. Kirjoittaja luokitteli ongelmat analysointivaiheessa neljään eri luokkaan, jotka kuvaavat ongelmien abstraktiotasoa<sup>10</sup>. Analysoinnissa käytetyt luokat esiteltiin kappaleessa 4.4.2.

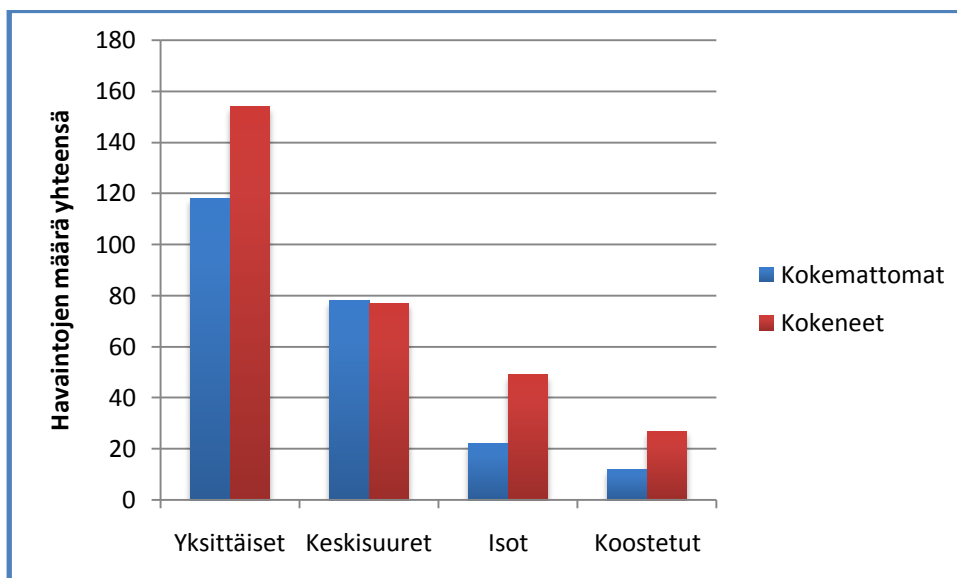
Sekä kokeneet että kokemattomat arvioijat kirjasivat eniten yksittäisiä havaintoja; molemmilla ryhmillä yksittäisiä havaintoja oli noin puolet kaikista kirjatusta havainnoista (kokeneilla 50 %, kokemattomilla 51 %). Koska kokeneet arvioijat kirjasivat kokonaisuudessaan enemmän havaintoja, oli yksittäisten havaintojen lukumäärä heillä lähes kolmannes suurempi kuin kokemattomilla. Keskisuuria ongelmia molemmat ryhmät kirjasivat määrällisesti yhtä paljon: suhteellisesti keskisuurten ongelmien osuus kokemattomien kokonaismäärästä oli kolmannes, kokeneilla neljännes. Suurissa ongelmakokonaisuuksissa ja useista havainnoista koostetuista kokonaisuuksissa ilmeni merkittäviä eroja. Näissä molemmissa luokissa kokeneet löysivät yli kaksinkertaisen määrän ongelmia kokemattomiin verrattuna. Kuva 20 esittää abstraktiotasoihin luokitellut ongelmat ryhmäkohtaisesti absoluuttisina määrinä, Kuva 21 puolestaan prosentuaalisesti suhteutettuna kaikkiin ryhmän kirjaamiin ongelmiin.

Kokemattomien arvioijien tekemien havaintojen kokonaismäärästä kymmenisen prosenttia oli abstraktiotasoltaan suuria ongelmia ja viisi prosenttia useammasta erillisestä havainnosta koottuja kokonaisuuksia. Kokeneilla vastaavat määrät olivat kuusitoista ja yhdeksän prosenttia. Molemmissa luokissa ero kokeneiden eduksi on tilastollisesti merkitsevä. Tulosten perusteella siis kokemattomat arvioijat kirjaavat melko tehokkaasti pieniä ja keskisuuria havaintoja, mutta laajempia ongelmakokonaisuuksia he eivät löydä yhtä herkästi.

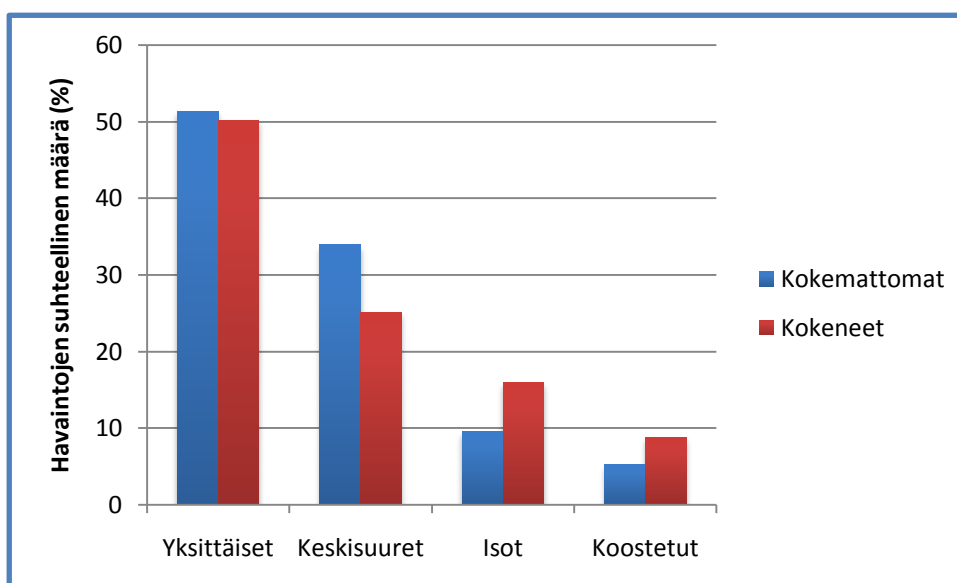
---

<sup>10</sup> On huomattava, ettei havaintojen abstraktiotaso tarkoita samaa asiaa kuin vakavuusaste, joka arvioijia pyydettiin määrittämään kullekin kirjaamalleen havainnolle. Pieni yksityiskohtainen havainto voi jossain tilanteessa olla vaikutukseltaan vakava, toisaalta suurempi kokonaisuus saattaa olla vain lievä ongelma. Toki usein abstraktiotason ja vakavuuden välillä voi olla yhteys.

Vaikka myös kokeneet arvioijat kirjaavat laajempia kokonaisuuksia muita ongelmia vähemmän, he kuitenkin pääsevät arvioinnissaan ongelmien abstraktiotasossa selkeästi syvemmälle.



Kuva 20. Ongelmien määrä abstraktiotasoin tarkasteltuna.



Kuva 21. Ongelmien abstraktiotasojen prosentuaalinen osuus kaikkiin ryhmän kirjaamiin ongelmiin suhteutettuna.

Ero laajemmissa ongelmakokonaisuuksissa korostui tehtävissä 1 ja 2, joissa molemmissa arvioitiin yksittäisiä sivuja. Kuten ongelmien painopisteiden, myös abstraktiotasojen perusteella kokemattomien arvioijien tekemä yksittäisen sivun arviointi jää selvästi useamman sivun kokonaisuutta pinnallisemmaksi. Ensimmäisen tehtävän sivun näkyvät ja selkeät yksityiskohtaiset ongelmat aiheuttivat sen, että kokemattomat tuntuivat jäävän niiden vangiksi pääsemättä arvioimaan sivun luonnetta syvällisemmin. Kokeneet asiantuntijat taas pystyivät pureutumaan yksityiskohtien läpi palveluun syvemmälle ja tarttumaan laajempiin ongelmakokonaisuuksiin.

Tehtävän 2 sivua arvioitaessa kokemattomia arvioijia taas häiritsi ensimmäisen sivun kaltaisten selkeiden ja näkyvien ongelmien puute. He eivät tuntuneet saavan sivusta kovin paljon irti, toisin sanoen löytämään keskeisiä ongelmia.

Aiemmissa tutkimuksissa on todettu, että vakavat ongelmat löytyvät arvioinneissa todennäköisimmin, ja kaikkein vakavimmat kokemuksesta riippumatta varmimmin (Nielsen 1992; Jacobsen ym. 1998). Tämä voi tuntua ristiriitaiselta sen kanssa, että tässä tutkimuksessa kokemattomat arvioijat kirjasivat kokeneisiin verrattuna melko tehokkaasti nimenomaan yksittäisiä ja keskisuuria käytettävyysongelmia, mutta eivät niinkään laajoja ongelmakokonaisuuksia. On kuitenkin muistettava, että havaintojen vakavuusluokitus ei ole suorassa suhteessa abstraktiotasoihin. Tulos on silti samansuuntainen arvioijien tekemän vakavuusluokituksen sekä havaintojen osa-alueiden tarkastelun tulosten kanssa: kokemattomat arvioijat kirjasivat vakavia ongelmia selvästi vähemmän, samoin konseptia ja rakennetta koskevia havaintoja (ks. 5.1.2 ja 5.1.3). Näin ollen voidaan varsin luotettavasti sanoa, että kokemattomat arvioijat löysivät ainakin tässä tutkimuksessa vakavia ongelmia selvästi kokeneita vähemmän. Ristiriita aiempiin tutkimuksiin voitaisiin varmistaa vain objektiivisen vakavuusluokituksen avulla, mutta sellaista ei tässä tutkimuksessa laadittu. Toisaalta myös Desurvire (1994) on havainnut, että kokemattomat eivät löydä vakavia ongelmia yhtä kattavasti.

Kokeneet arvioijat kirjasivat enemmän isoja ongelmakokonaisuuksia, mutta ääneen ajattelun avulla kerätyn laadullisen aineiston perusteella havaintoa ei välttämättä aina tehty suoraan isona kokonaisuutena. Kokenut arvioija saattoi pohtia useamman yksittäisen havainnon suhdetta toisiinsa ja miettiä niiden todellista syytä. Tällöin syvemältä tuotteesta saattoi paljastua keskeisempi ja suurempi ongelma, joka ei sellaisenaan näkynyt suoraan, mutta saattoi aiheuttaa tuotteen käyttöön useita pienempiä yksittäisiä ongelmia. Näiden yksittäisten ongelmien korjaaminen toki helpottaa tuotteen käyttöä, mutta ei välttämättä poista niiden taustalla olevaa varsinaista, huonosta suunnittelusta johtuvaa ongelmaa.

Abstraktiotasoltaan suuremman ongelman määrittäminen vaikutti olevan monelle kokeneelle arvioijalle luontevampaa usean yksityiskohtaisen havainnon sijaan. Yksityiskohtia käytettiin perustelemaan laajempaa kokonaisuutta. Tapa palvelee arvioinnin tarkoitusta ja asiakasta paremmin: tuote voidaan suunnitella uudelleen huolellisemmin tarvittavilta osin. Samasta syystä myös yksittäisten havaintojen kokoaminen mielekkäiksi kokonaisuuksiksi oli monille kokeneille arvioijille luontevaa, vaikka yksittäisten havaintojen taustalla ei ollut havaittavissa laajempaa piilevää ongelmaa.

Kaiken kaikkiaan kokeneiden arvioijien kirjaamisessa ongelmissa ja arviointitavoissa näkyi se, että he muotoilivat arvioinnin tulokset mahdollisimman luonteviksi ja hyödyllisiksi asiakkaalle esittämistä ja tuotteen kehittämistä ajatellen. Kokemattomat usein tyytyivät luettelemaan havaitsemansa yksittäiset ongelmat ajattelematta niiden esittämistä tai korjaamista.

Kuten muissakin käsitellyissä näkökulmissa, myös abstraktiotasoissa esiintyi henkilökohtaisia eroja ryhmien sisällä, mutta erot eivät kuitenkaan olleet niin selkeitä kuin havaintojen absoluuttisessa määrässä tai osa-alueissa. Suurimmillaan henkilökohtaiset erot näkyivät keskisuurten ongelmien kohdalla. Abstraktiotasoltaan suurissa ongelmissa ja koostetuissa ongelmakokonaisuuksissa erot olivat puolestaan varsin pieniä. Kokeneilla arvioijilla tuntuu olevan melko yhtenäinen tapa määritellä abstraktiotasoltaan ja merkitykseltään suurempia kokonaisuuksia, mutta niitä pienempien havaintojen kirjaamisessa näkyvät henkilökohtaiset erot joko ongelmien löytämisessä tai sitten siinä, kuinka tärkeänä niitä pidetään suurempien kokonaisuuksien rinnalla.

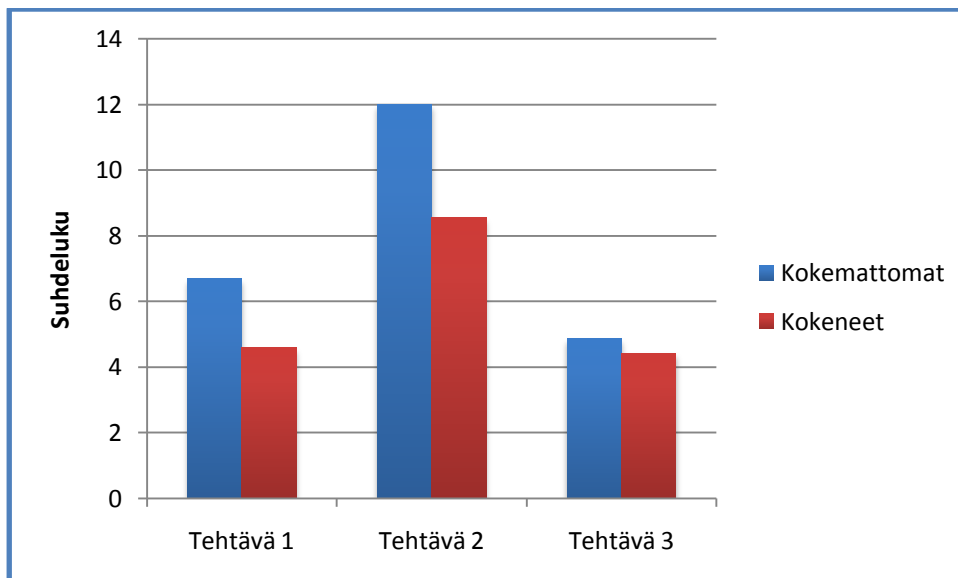
### 5.1.5 Arvioitujen osien paremmuuden vertailu

Arvioijien näkemystä eri tehtävissä käsiteltyjen osien keskinäisestä paremmuudesta voidaan tutkimuksen tulosten avulla lähestyä kahta eri kautta; arvioinnissa tehtyjen luokiteltujen havaintojen sekä arvioijien antamien arvosanojen avulla. Nämä eri lähestymistavat antavat mahdollisuuden tarkastella todellisten tulosten suhdetta subjektiivisiin näkemyksiin. Lisäksi eri ryhmien välillä voidaan vertailla tehtävissä käsiteltyjen osien paremmuusjärjestystä.

Kirjattujen havaintojen avulla voidaan muodostaa eräänlainen arvio kunkin arvioidun osan käytettävyydelle tehtyjen havaintojen määrän ja vakavuusluokituksen avulla. Oheisessa kaaviossa (ks. Kuva 22) eri tehtäviä koskevat vakavuuteen suhteutetut havainnot esitetään kertoimien avulla muodostetuilla suhdeluvuilla. Tehdyt havainnot tehtävittäin suhteutettiin niiden vaikutukseen vakavuusluokitusta vastaavilla kertoimilla siten, että lievä ongelma sai kertoimen 1, merkittävä ongelma kertoimen 2 ja vakava ongelma kertoimen 3. Summat laskettiin kertoimien avulla ryhmittäin ja tehtävittäin. Ryhmä- ja tehtäväkohtainen suhdeluku lasketaan jakamalla kokonaissumma vastaavalla summalla. Suhdeluvut eivät liiku tietyllä skaalalla, koska kirjatulla havainnoilla ei ole maksimiarvoa, mutta niiden avulla on mahdollista vertailla ryhmiä ja tehtäviä.

Ensimmäisen tehtävän sivu muodostui suhdelukujen perusteella molempien ryhmien arvioinnissa toista tehtävää heikommaksi. Ero kokeneiden ja kokemattomien arvioijien välillä näkyy kuitenkin selvästi suhdeluvun suuruudessa molemmissa tehtävissä.

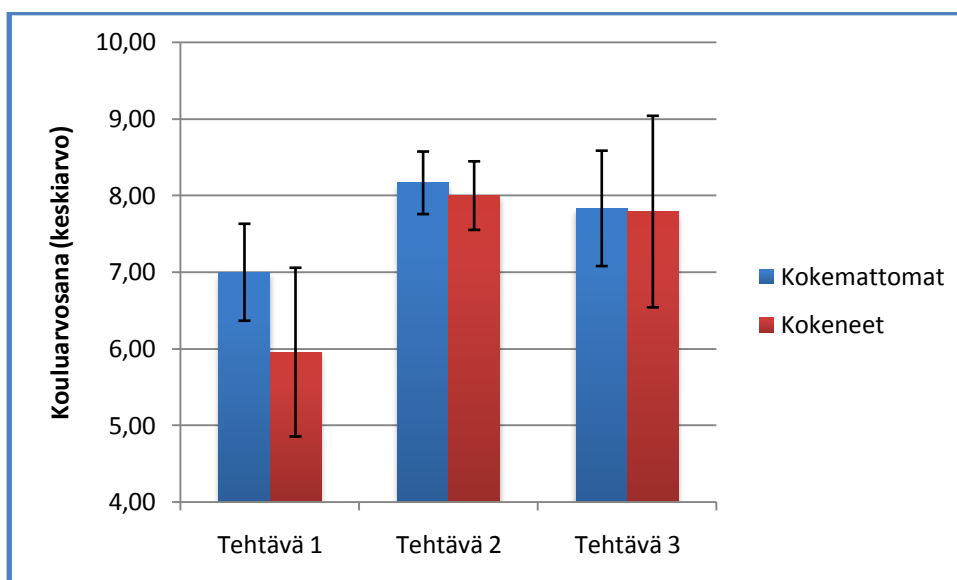
On huomattava, että kolmannessa tehtävässä arvioitiin useita sivuja, joten sitä koskevia havaintoja on ymmärrettävästi muita enemmän. Osa havainnoista koskee kuitenkin kokonaisuutta tai useaa erillistä sivua, joten suoraa vertailua tässä tapauksessa ei voida kolmannen tehtävän osalta tehdä. Tehtävä on kuitenkin mukana tarkastelussa mielenkiinnon vuoksi.



**Kuva 22.** Tehtävissä arvioitujen palveluiden paremmuuden vertailu ongelmien ja niiden vakavuusluokituksen avulla muodostettujen suhdelukujen avulla.

Kuva 23 esittää kaaviossa arvioijien kouluarvosanoina (4-10) antamat subjektiiviset arviot eri tehtävissä arvioitujen osien käytettävyydestä keskiarvoina. Arvosanoilla mitattuna molemmat ryhmät pitivät parhaana toisen tehtävän sivua, ero ryhmien välillä arvosanoissa oli hyvin pieni (kokemattomilla 8,2, kokeneilla 8,0). Ensimmäisen tehtävän sivu sai selvästi

toista huonommat arvosanat (kokemattomilla 7,0; kokeneilla 6,0). Kolmannen tehtävän arvosanoissa ei ryhmien välillä ollut eroa.



**Kuva 23.** Tehtävissä arvioitujen palveluiden paremmuuden vertailu arvioijien antamien kouluarvosanojen avulla. Keskihajonta esitetään kaavion palkeissa pystyjanojen avulla.

Verrattaessa arvosanoja suhdelukujen avulla tehtyyn tarkasteluun havaitaan, että tulokset ovat melko yhdenmukaisia. Ensimmäisen tehtävän sivun käytettävyys oli molempien ryhmien arvioimana heikompi kuin toisen. Molemmat tarkastelutavat lisäksi kertovat, että kokeneiden arviot ovat kokemattomia kriittisempiä, vaikka arvosanoissa ryhmien väliset erot tehtävissä 2 ja 3 ovatkin kosmeettisia. Joka tapauksessa vertailu antaa suuntaa siitä, että subjektiivinen näkemys arvioidusta sivusta korreloi arvioinnissa kirjattujen havaintojen kanssa. Voidaan siis ajatella, että subjektiivinen arvosana perustuisi nimenomaan kirjattuihin ongelmakohtiin, eikä ainakaan pelkästään arvioinnin aikana muodostuneeseen mielikuvaan. Toisin sanoen, kun kokematon arvioija tutkimuksen perusteella kirjaa havaintoja kokeneita vähemmän, näkee hän arvioidun osan myös yleisesti parempana kuin kokenut. Toisaalta arvioijien väliset yksilölliset erot näkyvät myös arvosanojen kohdalla keskihajonnassa.

## 5.1.6 Korjausehdotukset

Tässä kappaleessa esitellään korjausehdotusten tarkastelun tulokset. Arvioijia pyydettiin kunkin tehtävän päätteeksi kertomaan kolme keskeisintä korjausta arvioidulle tuotteelle. Koska havaintoja kirjattiin ylös kolmessa tehtävässä, koskee korjauksien vertailu tehtyihin havaintoihin kolmea arviointitehtävää molemmissa ryhmissä kuudella arvioijalla.

### 5.1.6.1 Kokeneet arvioijat

Kokeneiden arvioijien korjausehdotukset olivat tyypillisesti muodoltaan valmiita ja suurehkoja kokonaisuuksia, jotka korjaisivat tuotetta sellaisenaan toteutettuina. Osa ehdotuksista liittyi toki yksityiskohtiin, mutta nekin useimmiten olivat selvästi käyttöön vaikuttavia asioita. Kuudesta arvioijasta viiden korjausehdotukset olivat tyypillisesti laajempia, ja keskittyivät korjaamaan suurempia ongelmakokonaisuuksia. Vain yhdellä arvioijalla korjausehdotukset keskittyivät pääosin korjaamaan yksityiskohtia. Arvioijien yhteensä 18 arviointitehtävästä (6 arvioijaa, kullakin 3 tehtävää) 14:ssä korjaukset keskittyivät laajempiin kokonaisuuksiin, neljässä puolestaan yksityiskohtiin. Laajempia kokonaisuuksia olivat esimerkiksi seuraavat korjausehdotukset:

- *Ohjetekstien ja toimintojen yhdenmukaistaminen*
- *Valikon esittäminen selkeämmin*
- *Sivun jäsentely paremmin.*

Kokeneet arvioijat priorisoivat korjausehdotuksia luonnostaan. Osa kokeneista arvioijista mainitsi arvioidun osan jälkeen vain yhden tai kaksi korjausehdotusta, jotka he kokivat niin tärkeiksi, että ne korjaamalla arvioitu palvelu tai sen osa saataisiin merkittävästi parempaan kuntoon. Eli usean seikan mainitsemisen sijaan kokeneet usein suosittelivat keskittymään johonkin selkeään kokonaisuuteen, joka merkittävästi parantaisi tuotteen käytettävyyttä. Esimerkiksi tehtävässä 2 kokeneista arvioijista kaksi nosti valikon uudelleensuunnittelun niin keskeiseksi, etteivät he maininneet muita korjauksia lainkaan: *”Valikkokokonaisuus pitäisi vielä kerran mieltiä uudestaan ... niin sanoisin, että korjatkaa se, ja muut voi jättää väliin”*. Korjausehdotuksien antamisesta näkyi selvästi ajatus järjestelmän parantamisesta, ei niinkään kehitysehdotusten laatiminen suoraviivaisesti tehdyille havainnoille: *”Kun niputan kolme sopivasti, niin saan aika hyvän tuosta”*.

Priorisoinnin lisäksi kokeneet arvioijat saattoivat korjaukset usein myös valmiiksi ymmärrettävän ja helposti sovellettavaan muotoon. He mainitsivat ehdotukset usein myös tärkeysjärjestyksessä; ensin mainittiin koko palvelun osaa koskevia ongelmia, sitten yksittäisiä erillisiä sivuja tai palvelun osia koskevia ongelmia.

Kokeneiden arvioijien antamat korjausehdotukset liittyivät myös paikoin palvelun jatkotoimenpiteisiin ja jatkuvaan hallintaan, ei pelkästään sen nykyiseen sisältöön ja olemukseen: *”Pitäisi päivittää useammin ja tarjota linkkejä eri uutisiin”, ”Pitäisi jaotella sisällön mukaan, täältä ei löydä mitään”*. Lisäksi osa kokeneiden arvioijien korjauksista oli konseptia koskevia ohjeita palvelun tuottamiseen tai ylläpitoon: *”Pitäisi mieltiä, mikä on tärkeää”*.

Kokeneet arvioijat suunnittelivat käyttöliittymää uudelleen usein jo varsinaisen arvioinnin aikana. Ääneen ajattelua tarkasteltaessa havaitaan, että usein korjausehdotus saattoi tulla jo ennen varsinaisen ongelman määrittelyä. Arvioija saattoi esimerkiksi todeta *”Painikkeen terminä pitäisi tässä olla Tilaa”*, minkä jälkeen hän muotoili ongelman uuden ratkaisun pohjalta: *”Hyväksy on väärä termi. Termin pitäisi olla Tilaa”*. Puolet kokeneista arvioijista määrittää useita ongelmia niihin liittyvien ratkaisujen, ehdotuksien tai suunnitelmien kautta.

Kokeneiden arvioijien antamat korjausehdotukset kattoivat melko hyvin heidän kirjaamansa, vakavimmiksi luettelemansa ongelmakohdat. Yhteensä 18 arviointitehtävästä 14:ssä kokeneiden arvioijien keskeisimmät korjaukset vastasivat heidän itsensä määrittelemiä vakavia ongelmia. Vain neljässä tehtävässä yhteyttä vakavimpiin ongelmiin ei ollut havaittavissa.

#### 5.1.6.2 Kokemattomat arvioijat

Kokemattomien arvioijien antamat kolme korjausehdotusta koskivat useimmiten yksityiskohtia ja liittyivät varsin suoraan yksittäisiin havaintoihin. Eli varsin tyypillistä oli, että kokemattomat poimivat kirjaamistaan havainnoista kolme yksittäistä ongelma-kohtaa, joita esittivät korjattaviksi. Ajoittain korjausehdotukset jäivät myös pinnalliseksi viilaukseksi: *”Ensimmäisenä se Home-ikoni ... se voisi olla iso ja graafisesti hienosti viimeistelty”*. Esimerkkinä rajautumisesta yksittäisten ongelmien korjaamiseen on myös se, että yksi kokematon arvioija puuttui tarkkaan linkkien ulkonäköön ja väreihin, muttei kuitenkaan siihen, minkä kyseisellä sivulla pitäisi olla linkki, vaikka juuri se olisi kyseisessä tapauksessa ollut olennaista.



Kuudesta kokemattomasta arvioijasta neljän korjausehdotukset olivat pääosin luonteeltaan yksityiskohtien korjauksia, vain kahdella heistä korjaukset koskivat tyypillisesti laajempia kokonaisuuksia. Tehtävittäin tarkasteltuna kokemattomien arvioijien yhteensä 18 arviointitehtävästä (kuudella arvioijalla kolme tehtävää) viidessä tehtävässä esitetyt korjaukset koskivat laajempia kokonaisuuksia, 13:ssa yksittäisiä ongelmia.

Yksityiskohtien lisäksi kokemattomien arvioijien korjausehdotuksissa oli joitakin hyvin abstrakteja ehdotuksia, joiden pohjalta konkreettiset korjaustoimenpiteet eivät yksin onnistuisi: *”Se pääjutun selkeys, mikä on siinä vasemmassa laidassa”, ”Lisää ohjeita minä laittaisin”* tai ylipäänsä muodossa *”tälle pitäisi tehdä jotain”*.

Korjausehdotusten laatimisessa näkyy myös kokemattomien arvioijien epävarmuus arvioinnista; arvioijalla saattoi olla selkeä näkemys jostakin rakenteellisesta korjauksesta, mutta hän saattoi hylätä sen, koska piti asiaa lopulta mielipidekysymyksenä. Kokemattomien arvioijien subjektiivisuus näkyi myös siten, että korjausehdotus saatettiin laatia itseä miellyttäväksi, vaikka se ei merkittävästi muuttaisi olemassa olevaa ratkaisua.

Kokemattomien arvioijien tekemät korjausehdotukset eivät monesti vastanneet ja kattaneet heidän kirjaamiaan vakavimpia ongelmakohtia. Edellä mainituista 18 tehtävästä joka toisessa eli yhdeksässä arvioijien mainitsemat kolme keskeisintä korjausehdotusta eivät vastanneet heidän itsensä vakavimmiksi luokittelemia ongelmia. Lisäksi joihinkin suurempiin ja vakaviksi luokiteltuihin ongelmiin saatettiin antaa turhan yksittäisiä ja pieniä korjauksia. Joillakin kokemattomilla arvioijilla näkyi silti olevan kyky muodostaa yksityiskohtaistenkin ongelmien pohjalta koko palvelua korjaavia ehdotuksia (tehtävässä 3): *”Sivujen pitäisi olla keskenään samannäköisiä”* ja *”Sivujen välillä pitäisi päästä liikkumaan niin, että kaikissa kohdissa pääsisi taaksepäin Edellinen-painikkeella”*.

## 5.2 Arviointitavat ja arvioinnin eteneminen

Tässä luvussa tarkastellaan arvioijien välisiä eroja arviointitavoissa ja arvioinnin etenemisessä. Edellisessä luvussa todettiin kirjattujen havaintojen määrällisen tarkastelun avulla, että erot näkyvät selvästi arvioinnin tuloksissa (ks. 5.1). Arviointitapojen laadullinen analyysi paljastaa yksilöllisiä eroja arvioinnin etenemisessä erilaisina prosesseina, menetelminä ja painotuksina.

### 5.2.1 Eteneminen tehtävittäin

Arvioinnin etenemiseen ja suorittamiseen vaikuttivat myös koeasetelman keskenään erilaiset arviointitehtävät. Kokeneita asiantuntijoita häiritsi selvästi enemmän arvioitavien yksittäisten sivujen kontekstin puute (tehtävät 1 ja 2). Sivuilta ei nähnyt, mistä niille on tultu, tai mikä sivun suhde on muuhun sivustoon nähden. Kaikki kokeneet arvioijat olisivat halunneet kokeilla sivun linkkejä ja selvittää siten sivun toimivuutta, navigointia ja sivujen keskinäisiä suhteita. Tutkimuksen aikana jotkut kokeneet arvioijat jättivät joitakin havaintoja tietoisesti kirjaamatta, kun eivät voineet varmistaa asiaa siirtymällä sivujen välillä. Suhtautuminen puutteelliseen kontekstiin on mielenkiintoinen, sillä arvioitavaan tuotteeseen ja koko tuoteperheeseen tutustuminen on nähty keskeisiksi tekijöiksi käytännön arviointien laadun kannalta (Sawyer ym. 1996).

Kokemattomia yksittäisen sivun arviointi ei tuntunut yhtä paljon häiritsevän, vain yksi heistä ilmaisi halunsa kokeilla linkkejä ja liikkua sivujen välillä. Tämä voi johtua siitä, että usein asiantuntija-arviointia opetetaan alussa tekemään sivu tai näkymä kerrallaan (sivun ollessa

usein vielä paperitulosteena), mikä on voinut olla hyvin tuoreena tutkimukseen osallistuneiden kokemattomien arvioijien mielessä.

#### 5.2.1.1 Tehtävä 1

Ensimmäisen tehtävän irralliselta sivulta (ks. 4.3.1.1) kokemattomat arvioijat keräsivät yksittäisiä käytettävyysoongelmia varsin kattavasti kokeneisiinkin nähden, mutta sirpaleisesta sivusta johtuvia yksittäisiä ja hajanaisia ongelmia syvemmälle heillä oli arvioinnissa vaikeuksia päästä.

Useimmat kokeneet arvioijat puolestaan pääsivät ensimmäisessäkin tehtävässä arvioinnissaan visuaalisia ja irrallisia havaintoja syvemmälle. Lähes kaikki kokeneet asiantuntijat pohtivat ääneen sivun toteutusta tarkoituksen kannalta, mutta kokemattomista tähän ylsi vain harva. Kokeneesta arvioijasta riippuen pienet, mutta näkyvät ongelmat kirjattiin alta pois heti aluksi tai sitten myöhemmin, mutta arvioinnin aikana näkyi kuitenkin selkeästi tavoite päästä käsiksi varsinaisiin, keskeisempiin ongelmiin. Sirpaleisten ongelmien kirjaamisen aikana kokeneillakin saattoi ajoittain jäädä kirjaamatta jokin keskeisempi ongelma, jonka he ovat havainneet. Syvemmälle päästyään ensimmäisessä tehtävässä kokeneet arvioijat kykenivät löytämään jopa hyviä puolia, vaikka ensivaikutelma sivusta on yleisesti keho.

Kokemattomat arvioijat pitivät selvästi eniten ensimmäisen tehtävän sivun arvioinnista – juuri siksi, että siltä oli melko helppoa poimia yksittäisiä käytettävyysoongelmia. Kokeneet puolestaan vaikuttivat selvästi tuskastuneemmilta. Tuskastuminen johtui joko siitä, että selvien suunnitteluvirheiden määrä turhautti heitä tai siitä, että vakavampiin pinnan alla piileviin käytettävyysongelmiin oli hankalampi päästä käsiksi lukuisista pienistä ongelmista johtuvan sirpaleisuuden ja sekavuuden vuoksi.

#### 5.2.1.2 Tehtävä 2

Arvioinnissa syvemmälle porautumisen vaikeus korostui toisen tehtävän ulkoisesti valmiiksi hiotussa sivussa (ks. 4.3.1.2), josta kokemattomilla oli selvästi vaikeuksia löytää todellisia ongelmia. Ensimmäisen tehtävän sirpaleiseen ja sekavaan sivuun verrattuna toisen tehtävän viimeistellympi ulkoasu johti siihen, että helposti silmäilemällä kerättäviä ongelmia ei juuri ollut. Kun mikään käyttöliittymässä ei tarttunut suoraan silmään, tuntui arviointi monien kokemattomien mielestä hankalalta: *”En jotenkin saa tästä mitään irti”*. Useampi kokematon ei löytänyt toisesta tehtävästä merkittäviä ongelmia, mutta he hieman väkisininkin kirjasivat edes joitakin melko lieviä ongelmia.

Monet kokeneet arvioijat vaikuttivat puolestaan tyytyväisemmiltä toisen tehtävän arvioinnin aikana, kun he sirpaleisten ongelmien keräämisen sijaan pääsivät heti käsiksi sivun luonteeseen ja tarkoitukseen liittyviin ongelmiin.

#### 5.2.1.3 Tehtävä 3

Kolmannen tehtävän vaiheittaisen polun (ks. 4.3.1.3) kohdalla ero näkyi siten, että puolet kokemattomista arvioijista ei arvioinut lainkaan vaiheittaista prosessia kokonaisuutena, vaan yksittäisinä sivuina. Tällöin he eivät myöskään kyseenalaistaneet koko arvioitavan prosessin rakennetta ja toimivuutta kokonaisuudessaan. Kaikki kokeneet asiantuntijat pohtivat puolestaan prosessia kokonaisuutena, joko heti aluksi (he hankkivat käsityksen koko prosessista ja kävivät vaiheet läpi) tai sitten lopuksi vertaamalla eri vaiheiden rooleja ja yhtenäisyyttä.

Valtaosa kokeneista pohti konseptin ja sitä vastaavan toteutuksen toimivuutta jo aikaisessa vaiheessa. Joku heistä pakotti itsensä siihen, kun yksittäiset ongelmat oli kirjattu ylös. Mutta kaikki kokeneetkaan arvioijat eivät tässä onnistuneet. Joka tapauksessa palvelun konseptin, soveltuvuuden ja sisällön kriittinen pohtiminen tai kyseenalaistaminen oli kokemattomille huomattavasti harvinaisempaa. Konsepti hyväksyttiin sellaisenaan ja arvioitiin vain sen esittämistapaa.

## 5.2.2 Arvioinnin etenemistavat

Tyypillisessä verkkosivuston asiantuntija-arviossa arvioitavasta käyttöliittymästä nousee nopeasti esiin useita selvästi havaittavia, mutta sinänsä pieniä käytettävyyso ongelmia. Nämä ongelmat ovat usein tyypillisiä suunnitteluvirheitä, jotka arvioijan on helppo tunnistaa ja kerätä ylös vain silmäilemällä tutkittavaa sivustoa. Tyypillisten virheiden lisäksi palveluita koskee useimmiten myös syvemmällä olevia käytettävyyso ongelmia, jotka liittyvät tarkemmin esimerkiksi juuri kyseisen palvelun rakenteeseen, toimintalogiikkaan, sisältöön tai konseptiin. Näiden ongelmien havaitseminen tai tunnistaminen ei yleensä onnistu välittömästi ja automaattisesti, vaan se vaatii porautumista arvioinnissa pintaa syvemmälle.

Shneiderman (1998) korostaa, että arvioijien tulisi päästä todellisen käyttötilanteen mukaiseen tilanteeseen, esimerkiksi kurssien, manuaalien ja tutoriaalien avulla. Lisäksi Sawyer ym. (1996) toteavat tuotteeseen tutustumisella olevan ison roolin arvioinnin vaikutukseen, samoin kuin koko tuoteperheen tarkastelulla.

Tässä tutkimuksessa tuli vastaan useita erilaisia arviointitapoja ja -tekniikoita. Yksinkertaisin tapa edetä arvioinnissa oli reaktiivinen silmäily. Etenkin kokemattomat arvioijat etenivät pääsääntöisesti silmäilemällä ilman erityistä strategiaa. Strukturoidumpia arviointitapoja olivat palvelun käyminen läpi järjestelmällisesti osa-alue kerrallaan sekä käyttäjän rooliin asettuminen, tai yksinkertaisesti erilaisten näkökulmien tarkastelu järjestelmällisesti yksi kerrallaan. Kokeneet arvioijat vähintään siirtyivät yksittäisten ja selvien ongelmien kirjaamisen jälkeen laajempiin kokonaisuuksiin, jos eivät käyttäneet muita tekniikoita. Käytetyt menetelmät vaihtelivat ryhmien, mutta myös yksittäisten arvioijien välillä.

### 5.2.2.1 Eteneminen silmäilemällä

Monilla kokemattomilla arvioijilla arviointi oli enemmän sattumanvaraista silmäilyä ja hyppimistä eri asioiden välillä kuin järjestelmällistä läpikäyntiä, ja arvioinnit etenivät tyypillisesti melko jäsentymättömästi. Silmäily oli arviointitapana reaktiivista, eli arvioijat keräsivät erilaisia havaintoja lähinnä sattumanvaraisessa järjestyksessä niiden noustessa esiin ja osuessa silmään. Tämä johti siihen, että esimerkiksi käyttöliittymän eri osien suhteet ja sivujen soveltuvuus tarkoitukseensa jäivät helposti yksittäisten elementtien varjoon. Selvästi näkyvät yksityiskohdat hukuttivat usein muut ongelmat alleen, eivätkä kokemattomat arvioijat silloin päässeet arvioinnissaan syvemmälle.

Edellä mainittu ilmiö näkyi erityisesti siirryttäessä ensimmäisen arviointitehtävän sekavan sivun jälkeen toisen tehtävän ulkoisesti selkeämpään ja viimeistellympään sivuun, jolta kokemattomilla arvioijilla oli selvästi hankalampaa löytää käytettävyyso ongelmia ilman johdonmukaista strategiaa. Kokemattomilla arvioijilla ei tuntunut olevan tutkimuksessa valmiuksia käydä läpi arvioinnin kohdetta järjestelmällisesti tai käyttää jotakin tekniikkaa kattavaan läpikäyntiin. Silmäilemällä havaitut yksittäiset ongelmakohdat ohjasivat vahvasti arviointia.

### 5.2.2.2 Tarkastelu osa-alueiden kautta

Toinen menetelmä tarkastella arvioitua palvelua oli käydä sitä läpi osa-alue kerrallaan. Osa-alueita saattoivat olla käytettävyyden osa-alueet (esim. navigointi, rakenne, visuaaliset asiat jne.) tai käyttöliittymän eri osat ja alueet.

Monet kokeneet arvioijat tutkivat arvioidun sivun navigointivalikkoa ja sen suhdetta muuhun sivuun hyvin tarkkaan – etenkin toisessa tehtävässä, jossa arvioitu sivu muodosti yhteenvedon ja koosteen koko palvelun sisällöstä. Valikkoa käytettiin keinona arvioida sivun rakennetta ja sisältöä: jos valikko ei kata sisältöä tai se ei ole looginen, kertoo se mahdollisesti palvelun rakenteellisista ongelmista. Toinen kokeneiden arvioijien käyttämä tapa sivukokonaisuuden hahmottamiseen oli linkkien ja otsikoiden terminologian tutkiminen sisällön ja toiminnan kannalta. Näin he pyrkivät selvittämään, miten otsikko rajaa sivun sisällön, onko otsikon ja sisällön suhde oikea sekä mitä muuta sivukokonaisuus kattaa:

- *"Uutisruudussa ei ole uutisia, vaan juttuja"*
- *"Kuuluuko muutoshistoria liittymään vai Vastaaaja Pro:hon?"*.

Konseptin ja sivun sisällön arvioinnin avulla kokeneet arvioijat pääsivät käsiksi todellisen käyttötilanteen kannalta keskeisimpiin asioihin. He pohtivat selvästi kokemattomia arvioijia useammin sivun tarkoitusta ja tehtävää ja lähtivät arvioimaan sivua tästä näkökulmasta. Kokeneet arvioijat panivat esimerkiksi useammin heti alussa merkille, kuinka suoraan sivu menee varsinaiseen asiaan, kuinka paljon sisällössä on toistoa ja miten sivun sisältö on priorisoitu. He pohtivat sivun sisältöä kattavammin myös suhteessa muuhun sivustoon, vaikka arvioinnin kohteena olisikin ollut vain kyseinen yksittäinen sivu. Lisäksi he arvioivat sisältöä kategorisoimalla sitä (esim. *"tiedotteet eivät tunnu tiedotteilta"*) sekä analysoimalla kieliasua, terminologian ohella myös puhuttelumuotoa ja sen soveltuvuutta palveluun. Tarkastellut osa-alueet ja niiden järjestys vaihteli kokeneilla arvioijilla henkilökohtaisesti, kuten jo kappaleen 5.1.3 tulosten pohjalta voidaan todeta. Joka tapauksessa kokeneet arvioijat keskittyivät ajoittain selvästi johonkin osa-alueeseen.

### 5.2.2.3 Arvioija käyttäjän roolissa

Arvioinnin suorittaminen ei yleensä onnistu pelkästään tarkastelemalla arvioitavaa tuotetta ulkopuolelta, vaan se vaatii tuotteen käyttäjän näkökulman ottamista huomioon erilaisissa tilanteissa. Esimerkiksi Sears (1997) korostaa todellisten tehtävien merkitystä ongelmien löytämisessä: pelkkä heuristinen lista ei auta arvioijaa fokusoitumaan käyttäjän toimintaan, mutta todellisten tehtävien avulla se onnistuu paremmin. Arvioija voi käyttäjän roolissa tarkastella tuotetta todellisen käyttötilanteen mukaisesta näkökulmasta tai esimerkiksi soveltaa introspektiota jollakin tasolla.

#### Tietoinen siirtyminen käyttäjän rooliin

Kokeneilla käyttäjillä näkyi tutkimuksessa kyky asettua tarkoituksenmukaisesti hetkittäin käyttäjän asemaan tilanteen niin vaatiessa. Käyttäjän asemaan asetuttiin tarvittaessa ja tietoisesti jotakin tiettyä toiminnallisuutta, sisältöä tai tiettyä osaa arvioitaessa (esim. *"tämä on yrityksen omaa kieltä, ei minun luultavasti"*). Kokenut arvioija loi ajoittain arviointia varten todellisen käyttötilanteen ja käytettävyydestin kaltaisia luontevia tehtäviä, mikä helpotti arviointia käyttäjän näkökulmasta: *"miten tässä vaiheessa voi tehdä tämän?"*.

Tehtävien tapaan kokeneet arvioijat tekivät arvioinnin tueksi myös järjestelmällisesti joitakin oletuksia ja odotuksia palvelun toiminnasta, sisällöstä ja tarkoituksesta. Näitä luotuja oletuksia vasten he sitten peilasivat toteutusta ja ratkaisuja:

- *"Mikä on ajankohtaista ja tärkeää?"*
- *"Mitä ajankohtaisiin kuuluu?"*
- *"Kuvittelin, että täällä olisi ajankohtainen lista uutistapahtumista, mutta kun täällä on pääkirjoitusta ynnä muuta, olen ollut väärässä"*
- *"Kuinka usein kaikille yrityksen samannimisille haluttaisiin hakea Vastaaja Pro?"*
- *"Mitä ihmettä olen nyt hakemassa?"*

Kokeneet arvioijat käyttivät toisinaan apunaan myös palvelun tuottajan näkökulmaa. Kysymys *"miksi tämä on tehty?"* auttoi myös vertaamaan toteutettuja ratkaisuja palvelun konseptiin ja tarkoitukseen.

Kokemattomilta käyttäjän rooliin asettuminen ja todenkaltaisen tehtävän käyttäminen ei yhtä luontevasti onnistunut. Esimerkiksi yksi kokemattomista arvioijista ei havainnut arviointitehtävässä kovin paljon ongelmia, jolloin hän totesi: *"jos pitäisi tehdä jotain, niin ongelmia ehkä löytyisi"*. Toinen kokematon arvioija puolestaan pohti, miten hänen isoäitinsä selviytyisi palvelussa, mikä oli kuitenkin palvelun tarkoituksen ja todellisen kohderyhmän kannalta oikeastaan merkityksetöntä.

### **Ajautuminen käyttäjän rooliin**

Tietoisesti käyttäjän rooliin siirtymisen sijaan kokemattomat arvioijat tuntuivat ajautuvan arvioinnin aikana toistuvasti käyttäjän rooliin huomaamattaan, eli objektiivisen arvioijan sijaan he olivat itse käyttäjän asemassa. Tällöin heidän omat henkilökohtaiset kokemuksensa ja odotuksensa käyttäjänä ohjasivat arviointia voimakkaasti sen sijaan, että he olisivat arvioineet tilannetta todellisen käyttäjän kannalta. Myös aiemmissa tutkimuksissa (Hertzum & Jacobsen 1999) on havaittu, että kokemattomat arvioijat saattavat ankkuroitua omaan kokemukseensa käyttäjinä ja unohtaa todellisen käyttäjän.

Arvioijat eivät itse välttämättä edustaneet varsinaista kohderyhmää, mikä johti joihinkin virhetulkintoihin sekä subjektiiviseen ja rajoittuneeseen arviointiin. Lisäksi se esti eri kohderyhmien näkökulmien ottamisen huomioon. Kokemattomat arvioijat kokivat toistuvasti ongelmiksi kohtia, joita eivät itse ymmärtäneet, ja henkilökohtainen näkökulma näkyi muutenkin ääneen ajattelussa:

- *"Tämä sivu on epämiellyttävän näköinen"*
- *"En tykkää tästä yhtään"*
- *"En ymmärrä tätä"*
- *"Nyt tässä tapahtui sellaista käyttäjätestiä"*
- *"Odotin listaa, josta näkee, mitä tänään tapahtuu"*
- *"Asiakasnumeroiden esitystapa on epäselvä"* (numerot vieraita arvioijalle).

Luonnollisesti myös kokeneiden arvioijien oma henkilökohtainen mielipide ja painopiste vaikuttivat arviointiin, kuten todettiin jo arvioinnin osa-alueiden tarkastelussa kappaleessa 5.1.3. Useimmiten kokeneet arvioijat pystyivät kuitenkin pitämään arvioinnin objektiivisempaan pohtimalla tietoisesti erilaisia näkökulmia, käyttötilanteita ja käyttäjiä. Todellisen käyttäjän (ja eri käyttäjäryhmien) näkökulma olikin usein vahvasti mukana heidän arvioinnissaan:

- *"Lähtisikö käyttäjä tekemään tätä?"*
- *"Kenelle tämä on tarkoitettu?"*
- *"Tätä ei ole suunnattu kenellekään"*

- *”Tähän kohtaan on sekoitettu useammalle kohderyhmälle tarkoitettua tietoa, jolloin kukaan ei löydä mitään”.*

#### 5.2.2.4 Erialaisten näkökulmien tarkastelu

Käyttäjän aseman lisäksi monet kokeneet arvioijat huomioivat myös muita eri näkökulmia palvelun käytön kannalta ja analysoivat niitä kontekstiin perustuen. Vaikka kokeneilla arvioijilla havaintojen tekeminen ja ongelmanmuodostus tapahtuivat usein melko automaattisesti oman kokemuksen pohjalta, niin analysoimalla tutkittavaa sivua tarkemmin heidän oli kuitenkin mahdollista pureutua tarkemmin juuri kyseisen ratkaisun toimivuuteen.

Tutkimuksessa yksi kokenut arvioija esimerkiksi tarkasteli palvelun sopivuutta ensikäyttäjälle ja kokeneemmalle käyttäjälle, kun se oli palvelun luonteen kannalta olennaista. Toinen arvioija taas pohti hakutoiminnon yhteydessä monimutkaisten hakujen onnistumista. Kokeneet arvioijat myös siirtyivät luontevasti tietoisesti yhdestä näkökulmasta toiseen, kun siihen oli tarve. Seuraavassa on luoteltu joitakin esimerkkejä kokeneiden arvioijien soveltamista näkökulmista:

- Arvioija pitää tuotteessa käytettyä ratkaisua selkeänä, mutta ei kovin joustavana. Hän ei kuitenkaan alkaisi tekemään web-käyttöliittymää joustavammaksi selkeyden kustannuksella (eri näkökulmien priorisointi).
- Arvioija näkee arvioitavan palvelun kiistatta tarpeelliseksi. Sille on tilausta, mutta sen toteutus kyseisessä tapauksessa on heikko (konseptinäkökulma erillään toteutuksesta).
- Käytetty käyttöliittymäratkaisu on arvioijan mielestä yleisesti huono, mutta juuri kyseisessä tapauksessa kuitenkin opittava, eikä siis mahdoton (eri näkökulmien analyysi johtaa poikkeukseen).
- Arvioija näkee, että ratkaisu ei ole helppo ensikäyttäjälle, mutta kaiken tarpeellisen pystyy sen avulla kuitenkin tekemään (eri käyttötilanteiden huomioiminen).

#### 5.2.3 Havaintojen tekeminen

Havaintojen tekemisellä tarkoitetaan sitä, millä tavalla arvioinnin aikana muodostuvat havainnot käytännössä nousevat esiin. Käytettävyysongelmien sijaan tässä luvussa puhutaan enimmäkseen havainnoista, sillä etenkin kokeneiden arvioijien havainnoissa oli ongelmien lisäksi myös joitakin hyviä ominaisuuksia.

##### 5.2.3.1 Havaintojen kirjaaminen, tahti ja etenemisjärjestys

Tutkimuksessa osa kokeneista arvioijista alkoi kirjata silmiin osuvia ”helppoja” yksityiskohtaisia havaintoja heti arvioinnin alussa. Näiden kirjaamisen jälkeen he siirtyivät tutkimaan monimutkaisempia, palveluun tarkemmin liittyviä syvempiä ongelmia. Toiset kokeneet taas jättivät yksityiskohdat tietoisesti huomioimatta arvioinnin alussa ja pohtivat ensin tarkemmin konseptin, rakenteen ja sisällön ongelmia. Kun arviointiin varattua aikaa oli jäljellä vähän, he ryhtyivät rutiininomaisesti kirjaamaan pienempiä ja selkeitä havaintoja. Silloin tällöin joku kokenut arvioija jätti selkeitä yksityiskohtia kokonaan kirjaamatta keskittyessään tarkempaa analysointia vaativiin kokonaisuuksiin, vaikka luettelikin niitä ääneen. Kaikkien kokeneiden arvioijien etenemisessä näkyi havaintojen priorisointi, mutta he kuitenkin toimivat hieman eri tavoin sen suhteen.

Tavallisin henkilökohtainen ero erityisesti kokeneiden arvioijien välillä oli se, että osa heistä kirjasi havaintoja ylös yksitellen niitä tehdessään, osa taas pohti kerralla tiettyä laajempaa kokonaisuutta ja alkoi kirjata kokonaisuutta koskevia havaintoja ryppäänä sen jälkeen. Tavat

näkyivät vastaavasti myös ääneen ajattelussa; ensimmäisessä tapauksessa havainnot mainittiin ääneen samaan tahtiin kirjaamisen kanssa, jälkimmäisessä taas arvioija saattoi pohtia kokonaisuutta hetken hiljaa, minkä jälkeen alkoi purkaa kokonaisuutta auki sanallisesti. Kokemattomille arvioijille tavallista oli se, että he alkoivat pääsääntöisesti kirjata havaintoja yksitellen sitä mukaa kun niitä tekivät.

Kokeneen arvioijan määrittämä ongelmakokonaisuus, sen perustelu ja vastaava korjausehdotus hahmottuivat usein havaintoa kirjoitettaessa. Kirjoittaminen oli monilla selvästi osa ongelmanmäärittämisprosessia, ja havainnon kirjoittaminen saattoi myös tuoda esiin uusia näkökulmia ja havaintoja.

Kun arvioitavana oli yksittäistä sivua laajempi kokonaisuus, selasivat monet kokeneet arvioijat aluksi eri sivut läpi hahmottaakseen kokonaisuuden ja sen osien suhteet. Lisäksi he liikkuvat muutenkin arvioinnin aikana usein yksittäisten sivujen välillä edestakaisin. Kokemattomat arvioijat puolestaan kävivät tyyppillisesti kokonaisuuden läpi sivu kerrallaan ja arvioivat vaiheittaisen sivusarjan osia yksittäisinä sivuina. Kokonaisuuden hahmottaminen auttoi kokeneita arvioijia huomioimaan palvelun toimintalogiikkaa suunnitelmallisesti ja järjestelmällisesti:

- *"Täytyy katsoa tuleeko tuo vastaan myöhemmin"* (myöhemmissä vaiheissa).
- *"Missä vaiheessa annoin hakuehdon Vastaaja Pro:n tilaus?"*.

### 5.2.3.2 Havaintojen esittäminen ja perustelu

Kokeneet arvioijat pääsääntöisesti käsittelivät ja muotoilivat havaintojaan erilaisilla tavoilla, joita heiltä ei kuitenkaan pyydetty: he antoivat oma-aloitteisesti korjausehdotuksia tekemilleen havainnoille, tekivät usein suullisen yhteenvedon arvioinnin päätteeksi (kun aika oli loppumassa) sekä jäsentelivät ja luokittelivat tekemiään havaintoja. Lisäksi kokeneet arvioijat puntaroivat ja pohtivat perusteluja tekemilleen havainnoille yleensä hyvin kattavasti ja huolellisesti – luontevana osana arviointiprosessia. Kokemattomien arvioijien kertomat perustelut eivät useimmissa tapauksissa kertoneet merkittävästi kirjattua havaintoa enempää, tai sitten perustelu saattoi olla yksinkertaisesti viittaus tiettyyn periaatteeseen tai sääntöön. Havaintojen esittämisessä ja perustelussa näkyi se, että kokeneet arvioijat ovat oppineet ja tottuneet perustelevaan havaintonsa perusteellisesti suoraan asiakkaalle tai arviointiraporttiin, jolloin pelkkä sääntöihin tai konventioihin viittaaminen ei riitä. Myös yhteenvedon tekeminen ja havaintojen jäsentely helpottaa tulosten esittämistä ja välittämistä eteenpäin.

Siitä huolimatta, että korjausehdotuksia ei pyydetty arvioinnin aikana, kokeneet arvioijat useimmiten mainitsivat havaitsemansa ongelman yhteydessä myös ehdotuksen sen korjaamiseksi. Jotkut kokeneet arvioijat suunnittelivat jo arvioinnin aikana kokonaan uutta ratkaisua. He pohtivat erilaisia havaintoja kokonaisuutena ja kehittivät ratkaisun, joka korjaisi kokonaisuuden mahdollisimman laajasti – ja samalla useita ongelmia. Eli sen sijaan, että kullekin yksittäiselle ongelmakohdalle olisi laadittu korjausehdotus, saatettiin ehdotuksena antaa laajempi ja kokonaan uusi, paremmin tilanteeseen soveltuva ratkaisu. Yksi kokenut arvioija kykeni arvioinnin aikana jopa luomaan korjausehdotuksen kautta täysin uuden suunnittelusäännön, joka olisi sovellettavissa ja yleistettävissä myös jatkossa.

Kokeneet arvioijat johtivat ongelmakuvauksen ajoittain korjausehdotuksen tai uudelleen suunnittelun kautta. He muodostivat itselleen kuvan hyvästä ratkaisusta ja kirjasivat ongelmiksi sellaiset asiat, jotka eivät tukeneet tätä mallia. Tämä lähestymistapa vaikutti usein myös havaintojen kielelliseen muotoiluun; ne esitettiin monesti valmiiksi mielessä olevan ehdotuksen pohjalta, esimerkiksi: *"Tietojen syöttäminen ei sovi vahvistamisen kanssa*

*samalla sivulle*” (arvioija oli muodostanut ääneen ajatellessaan ratkaisun, joka olisi optimaalinen vaiheittaiselle tilausprosessille tehtävässä 3).

Kokemattomat arvioijat puolestaan eivät muodostaneet korjausehdotuksia oma-aloitteisesti kovin usein. He tyytyivät useimmiten mainitsemaan ja kirjaamaan vain havaitsemansa ongelmakohdan sellaisenaan, eivätkä välttämättä järjestelmällisesti pohtineet parempaa vaihtoehtoa tai ratkaisua.

### 5.2.3.3 Mitä pidetään käytettävyysohjelmana?

Arviointia voidaan tarkastella käytettävyyden eri osa-alueilla sekä näkökulmilla, joita arvioijat huomioivat arvioinnin kuluessa. Kirjattujen havaintojen analyysissä todettiin, että henkilökohtaiset erot arvioinnin painopisteissä vaikuttavat siihen, etteivät kokeneetkaan arvioijat osaa ottaa huomioon kaikkia käytettävyyden osa-alueita ja näkökulmia yhtä tehokkaasti (ks. 5.1.3). Joitakin vakaviakin käytettävyysohgelmia jäi arvioijilta arvioinnin aikana havaitsematta, koska kaikkiin osa-alueisiin ei välttämättä kiinnitetty huomiota.

Kokemattomien arvioijien arviointituloksiin vaikutti lisäksi se, mitä he mielsivät käytettävyysohjelmaiksi. Kokemattomilla arvioijilla oli hyvin vaihtelevia näkemyksiä siitä, mitä asioita tulisi mainita havaintoina tai käytettävyysohgelmina. Lukuisia havaintoja jäi kirjaamatta vain siksi, että arvioijat eivät pitäneet niitä käytettävyysohjelmina. Esimerkiksi yksi kokemattomista arvioijista ei pitänyt asioiden ryhmittelyä, sisällön järjestelyä tai kuvien valintaa käytettävyysohjelmina. Muutenkin kokemattomat arvioijat mielsivät monesti käytettävyysohgelmeiksi vain käyttöliittymässä selvästi näkyvät virheet, muut jäivät kirjattujen havaintojen ulkopuolelle:

- *”tää on kyllä ainoa käytettävyysohgelma, joka liittyy tähän”*
- *”mä en taas löydä käytettävyysohgelmia tästä”*
- *”mä en nyt tiedä, kun tää on tämmöstä sisällön järjestelyä taas”.*

Kokeneiden arvioijien näkemys käytettävyysohgelmaista oli selvästi laajempi ja yhtenäisempi. He hyväksyivät ja kirjasivat käytettävyysohgelmeiksi kaikki sellaiset havaitsemansa asiat, jotka jollakin tavalla saattoivat ylipäänsä käyttäjää haitata. Tästä huolimatta yksittäisissä havainnoissa näkyi mielipide- ja näkemyseroja kokeneiden arvioijien välillä. Esimerkiksi yksi kokenut arvioija sanoi arvioinnin aikana, että kollega puuttuisi sivulla olevaan visuaaliseen hälyyn, mutta itse hän ei pitänyt sitä varsinaisena ohgelmana. Kokeneet asiantuntijat ovat myös oppineet toisiltaan, ja saattavat tietää, mihin toinen asiantuntija kiinnittäisi huomiota. Arvioijat tuntuivat käyttävän tätä oman arviointinsa tukena, mutta uskalsivat silti olla eri mieltä. Useammankin kerran tutkimuksen aikana kokeneet arvioijat olivat jopa vahvasti eri mieltä jostain palvelun ominaisuudesta.

## 5.2.4 Arviointia ohjanneita asioita ja piirteitä

Tässä luvussa esitellään seikkoja, jotka vaikuttivat ohjaavasti arvioijien työskentelyyn. Asioilla saattoi olla joko positiivinen tai negatiivinen vaikutus.

### 5.2.4.1 Kokemattomat arvioijat: epävarmuus ja suoriutumisen tarve

Kokemattomien arvioijien työssä näkyi tutkimuksen aikana selvää epävarmuutta, erityisesti ongelmien havaitsemisessa ja kirjaamisessa. Epävarmuus tuli esiin ääneen ajattelun avulla; arvioijat saattoivat toisaalta rönsyillä asiassa, toisaalta puhua ja tehdä havaintoja hyvin epäröiden tai selvästi epäillä omia näkemyksiään. Kokemattomilta arvioijilta jäi joitakin selkeitä ongelmia kirjaamatta siksi, etteivät he olleet varmoja, onko kyseessä ylipäänsä



ongelma, tai kuinka vakavana sitä voidaan pitää. He pitivät joitakin ratkaisuja mahdollisina ongelmina, muttei kuitenkaan isoina puutteina. Useampi arvioija jätti jonkin havainnon kirjaamatta, kun ei ollut varma siitä, onko kyseessä ongelma. Asiaan vaikutti kyllä myös se, että käytettävyysongelman käsitettä tulkittiin hyvin eri tavoin. Pelkästään huonona asiana epävarmuutta ongelmien määrittelyssä ei kuitenkaan voi pitää, sillä se johtaa myös arvioinnin kannata hyödylliseen kyseenalaistamiseen:

- *"Täytyy alkaa miettimään, että onko siinä ongelmaa"*
- *"Katsotaan vielä, etten ihan sekoile".*

Ajoittain jotkut kokemattomat arvioijat pystyivät jopa kyseenalaistamaan tiettyä ratkaisua tai konseptia, esimerkiksi: *"Onko tässä pakko olla tarkennettu haku erikseen?"*. Useimmiten kuitenkin arviointi jäi heillä olemassa olevan toteutuksen pohtimiseen, jolloin he lähinnä totesivat, mitä arvioitavasta ratkaisusta puuttuu – eivät niinkään pohtineet, mikä on palvelun kannalta keskeistä, mikä taas kenties toissijaista.

Kokemattomilla arvioijilla oli selvästi tarve suoriutua arvioinnista hyvin ja vaikuttaa uskottavalta. Koeasetelma luonnollisesti jollakin tavalla korosti oman osaamisen näyttämistä sekä lisäsi jännitystä ja mahdollisesti epävarmuutta. Kuitenkin kokeneiden arvioijien varmuus arvioinnissa oli selvästi vahvempaa, vaikka osa heistäkin kertoi jännittäneensä koeasetelmaa. Kokemattomilla arvioijilla oli ainakin koetilanteessa tarve löytää ylipäänsä riittävästi ongelmia arvioinnin uskottavuuden kannalta, eli päästä siten todistamaan osaamistaan. Välillä tämä ajoi selvästi arvioinnin hyödyllisyyden edelle: *"En oikein osaa kirjoittaa tuosta mitään, niin kirjoitan jotain muuta"* (arvioija pohti hankalasti käsiteltävää ja purettavaa ongelmakokonaisuutta – vastaava tilanne esiintyi ainakin kahdella kokemattomalla arvioijalla).

Suorittamisen vaikutelmaa korosti entisestään arviointiin käytettävissä olevan ajan rajoittaminen. Ajan vähetessä näkyi, että arvioijien keskittyminen oli usein oman arviointityön etenemisessä, ei niinkään arvioinnin ja sen tulosten hyödyllisyydessä:

- Ohjaaja: *"Aikaa on jäljellä viisi minuuttia"*. Arvioija: *"En tiedä haluanko viidessä minuutissa enää mitään ruveta sen tarkemmin, ehkä ne olivat siinä"*.
- Arvioija: *"Nyt alkaa tulla se vaihe, että yrittää etsiä vielä yhden, jos on vielä minuutti aikaa"*.

Kun aika ei enää riittänyt kirjoittamiseen, joku kokematon arvioija luetteli vielä suullisesti useita havaintoja. Suorittamisesta ja uskottavuuden luomisesta johtuen kokemattomat arvioijat vaikuttivat tyytyväisiltä saadessaan kerättyä paljon pieniä ongelmia.

Epävarmuudesta kertoo jotain sekin, että jotkut kokemattomat arvioijat eivät pohtineet tai perustelleet havaintojaan johdonmukaisesti ääneen, vaan muodostivat usein näkemyksen mielessään, minkä jälkeen lausuivat ja kirjasivat havainnon valmiiksi muotoiltuna. Arvioijat siis ikään kuin pelkäsivät tuoda pohdiskeluaan ja perustelujaan esiin. Argumentointi ja perustelu oli siis joko puutteellista, tai sitten arvioijat vain eivät itse luottaneet sen pitävyyteen. Myös kirjattuihin havaintoihin päätyi huomioita, joita arvioija ei joko osannut tai halunnut muotoilla, perustella ja esittää selkeästi sekä ymmärrettävästi, esimerkiksi:

- *"Outo kovalinkki"*
- *"Keskeytä-nappi ei ole hyvä"*
- *"Uutisten linkit ja nuolet"*.

Kokemattomat arvioijat olivat vastikään oppineet käytettävyyden ja sen arvioinnin perusteet käytettävyyden peruskursseilla. Arvioinnissa he pohtivat mielessään ja mainitsivat ääneen joitakin yksittäisiä sääntöjä ja periaatteita (esim. Nielsenin heuristiikat) ja yrittävät soveltaa niitä arviointiin. Sääntöjen ja periaatteiden tukena ei ollut havaittavissa vielä kovin paljon omaa näkemystä tai kokemusta. Ongelmana oli lisäksi se, että kokemattomat arvioijat eivät käyttäneet sääntöjä kattavasti, eivät muistaneet niitä kaikkia, eivätkä aina soveltaneet niitä järkevästi, jolloin arviointi jäi helposti hajanaiseksi tai yksipuoliseksi. Joillakin kokemattomilla arvioijilla oli pieniä puutteita jo käytettävyyden perustiedoissakin, mikä johti joihinkin kyseenalaisiin tai jopa virheellisiin havaintoihin ja johtopäätöksiin:

- Arvioija vaati päätteellistä fonttia luettavuuden helpottamiseksi. Päätteellinen kirjasintyyli on nimenomaan hankalampaa näytöllä esitettynä, ja soveltuu paremmin tulostettuun tai painettuun tekstiin.
- Arvioija ehdotti painikkeen poistamista pudotusvalikon perästä. Vaikka nykyisin navigointiin tarkoitetut pudotusvalikot monesti toimivat itsenäisesti ilman painiketta, niin lähtökohtaisesti on kuitenkin suositeltu myös valinnan hyväksyvän painikkeen käyttämistä valikon kanssa.
- Arvioija kehui hyvää otsikkoa (Ajankohtaista). Otsikko ei vastaa kyseisen sivun sisältöä kovin hyvin – kyseessä voi tosin olla enemmän myös sisällön ongelma.
- Arvioija kehui sivun keskellä olevaa valikkoa, jonka ”saa helposti haltuun, kun on vain muutama linkki”. Valikon sisältö ja sijainti ovat itse asiassa varsin epäonnistuneita kokonaisuuteen nähden.

#### 5.2.4.2 Kokeneet arvioijat: elementtien tunnistaminen, opitut mallit ja skeemat

Kokeneet arvioijat havaitsivat ja tunnistivat arvioitavasta sovelluksesta melko nopeasti tuttuja käyttöliittymäratkaisuja ja elementtejä, joihin he ovat törmänneet aiemmin useinkin. Ääneen ajattelu paljasti useita tilanteita, joissa arvioija tunnisti nopeasti ongelmallisen ja heikoksi todetun ratkaisun. Osa tällaisista tilanteista johtui tyyppillisistä ja toistuvista suunnitteluvirheistä, joita kokeneen arvioijan oli helppo tuotteessa havaita. Kokeneille arvioijille on muodostunut vahvoja skeemoja, joiden avulla he kykenivät tunnistamaan tuttuja ratkaisuja ja rakenteita. Sama ilmiö on havaittu lukuisilla muilla tutkimusaloilla kokemusta tutkittaessa (ks. luku 3).

Toisaalta kokeneiden arvioijien huomio kiinnittyi nopeasti tyyppillisiin web-käyttöliittymän osiin, kuten valikot, otsikot, visuaaliset palkit ja navigointikomponentit. Pääosin huomio kiinnittyi siihen, jos elementti sisälsi yleisen käytettävyysongelman tai jos elementtien käyttö tai niiden väliset suhteet olivat epäkonventionaalisia. Esimerkiksi ensimmäisen arvioitavan sivun oikean reunan linkkilistaa jotkut kokeneet arvioijat pitivät sivuston valikkona, koska he nopeasti tunnistivat sen sellaiseksi muotonsa perusteella. Kuitenkin kyseessä oli luettelo linkeistä muihin sivun sisältöön liittyviin asioihin.

Arviointia ohjasi selvästi kokeneiden arvioijien oma kokemus ja näkemys. Vaikka etenemistavat arvioinnissa olivat yksilöllisiä, yhteistä kokeneille arvioijille oli se, että missään vaiheessa arviointia he eivät viitanneet tiettyihin tarkkoihin käytettävyyssääntöihin. Sen sijaan he perustelivat havaintojaan enemmän omin sanoin ja omalla kokemuksellaan, vaikka kyseessä olisikin ollut yleisesti tunnettu asia tai periaate, esimerkiksi:

- *”Olen tämän aika monta kertaa testissä nähnyt”*
- *”Montakohan kertaa joudun vielä urani aikana sanomaan, että...”*
- *”Tätä ei voi vääntää liikaa rautalangasta, se on lopputulema näistä kokemuksista”.*

Kokemukseen nojautuminen näkyi arvioinnissa ulospäin selvänä itsevarmuutena. Toisaalta itsevarmuus ja myös koeasetelman aiheuttama suoriutumisen tarve johtivat välillä siihen, että arvioijilla oli tarve todistella asiantuntemustaan tai näpäyttää joillakin tiedoilla ja arvioinneissa usein toistuvilla ongelmakohtilla. Arviointi oli tällöin luonteeltaan lähinnä käyttöliittymäoikolukua, eli näkyvät ja selkeät yksityiskohdat mainittiin, mutta ei liikuttu kovin syvällä palvelun luonteessa ja tarkoituksessa. Ajoittain jotkut kokeneet arvioijat purkautuivat pitkään joistakin pienistä ja usein toistuvista havainnoista, mikä vei huomiota ja aikaa keskeisemmiltä havainnoilta. Toistuvat virheellisesti suunnitellut yksityiskohdat tuntuivat turhauttavan joitakin kokeneita arvioijia niin, että he jäivät näiden kohtien vangiksi pitkäksi ajaksi. Ajoittaisesta kärkkäästä suullisesta kritiikistä ja jopa haukkumisesta huolimatta kokeneet arvioijat muotoilivat ja kirjasivat tehdyt havainnot korrektiin ja rakentavaan muotoon. Tutkimuksessa ei pyydetty erikseen muotoilemaan ongelmaotsikoita. Useimmat arvioijat (sekä kokeneet että kokemattomat) kirjasivatkin ne lähinnä omina muistiinpanoina ylös. Molemmista ryhmistä vain pari arvioijaa muotoili otsikoista muille ymmärrettäviä.

Arvioinnin aikana kokeneiden arvioijien työssä näkyi joitakin selviä valmiita malleja, jotka auttoivat arvioijia käymään tarpeelliset asiat läpi arvioitavassa palvelussa. Toiminta oli usein automaattista, mutta välillä myös tietoista ja harkittua. Esimerkiksi yksi kokenut arvioija kävi läpi useita hakutoimintoon liittyviä yksityiskohtia järjestelmällisesti selvästi vanhasta muistista. Eli arvioijan mielessä oli selkeä lista, jonka hän ilmeisesti käy läpi jokaisen arvioitavan hakutoiminnon kohdalla.

Myös havainnoille annettavat korjausehdotukset tulivat usein valmiiden mallien pohjalta, ikään kuin kokeneen arvioijan omasta ”ratkaisupankista”. Välillä arvioija saattoi myös kirjata havaintoja valmiin ratkaisun pohjalta. Eli arvioija kirjasi ylös havaintoina kohtia, jotka eivät noudattaneet hyväksi todettua ratkaisua ja mallia (esim. *”Tietojen syöttäminen ei sovi vahvistamisen kanssa samalle sivulle”*).

Kokeneiden arvioinnissa näkyi konkreettisesti se, että arvioidun osan ja sen ratkaisujen vertaaminen muihin vastaaviin on hyvin luonnollista. He vertasivat arvioitua sivua tai osaa lähinnä aiemmista arvioinneista ja käytettävyytsteistä kertyneeseen tietoon tai omaan käyttökokemukseensa palveluista. Vertailukohtana saattoivat olla muiden organisaatioiden ja yritysten samantapaiset palvelut tai saman organisaation muut palvelut. Erityisesti visuaalista ilmettä ja elementtejä verrattiin saman organisaation muihin sivuihin. Lisäksi vertailua tehtiin arvioinnin sisällä. Koeasetelmaan sisältyi kolme erillistä osaa saman yrityksen eri verkkopalveluista. Useampikin kokenut asiantuntija vertasi arvioinnin aikana myöhemmin arvioidun osan ratkaisuja aiempaan, eli tarkasteli näiden yhdenmukaisuutta.

#### 5.2.4.3 Rajatun ajan vaikutus arviointiin

Tutkimuksessa arviointiin käytettävän ajan rajoittaminen vaikutti arvioijiin eri tavoilla. Useimmat kokeneet arvioijat etenivät arvioinnissaan systemaattisesti käyden läpi järjestyksessä käytettävyyden eri osa-alueita tai arvioitavan sivun eri osia (ks. 5.2.2.2). Joillakin kokeneilla arvioijilla jäi rajatun ajan vuoksi kirjaamatta useita havaintoja, sillä heillä oli kiire käydä läpi kaikki eri näkökulmat. Tämä näkyi siinä, että arvioijat saattoivat ajan vähetessä luetella kiireessä ääneen useita havaintoja, joita he eivät kuitenkaan kirjanneet ylös tuloksiin.

Lisäksi jotkut kokeneet arvioijat jäivät ajoittain puhumaan ja rönsyilemään joistakin yksittäisistä asioista pitkäksi ajaksi, jolloin ajan vähetessä osa asioista jäi kokonaan käymättä läpi. Myös joidenkin arvioijien into havaintojen kirjoittamiseen hyttyi selvästi ajan

loppumista kohden. Toisaalta osa kokeneista arvioijista ei käyttänyt aikaa loppuun asti, vaan kertoi olevansa valmis, kun oli mielestään löytänyt keskeisimmät ongelmakohdat palvelusta.

Kokemattomat arvioijat käyttivät ajan useimmiten melko tarkkaan loppuun asti. Jos selkeimmät löydetyt havainnot oli kirjattu ylös, käytettiin jäljellä oleva aika yleensä samojen teemojen pohtimiseen tai palvelun silmäilyyn. Arvioinnin intensiteetti kuitenkin laski selvästi ajan loppua kohden. Monilta kokemattomilta arvioijilta aika näytti loppuvan kesken, mutta tutkimuksen päätteeksi tehdyssä haastattelussa he sanoivat silti ajan riittäneen. Kokemattomat arvioijat vastasivat pääsääntöisesti käyttävänsä luonnollisessa arviointitilanteessa arviointiin enemmän aikaa, mutta ei kuitenkaan yhtäjaksoisesti, vaan pienissä pätkissä tai vaihteittain.

Kokeneet arvioijat kokivat ajan niin ikään riittäväksi – ja kuten todettiin, kaikki eivät käyttäneet aikaa loppuun asti. Toisaalta pari kokenutta arvioijaa yllättyi kuluttaessaan kaiken käytössä olleen ajan. Lähtökohtaisesti aika tuntui pitkältä, mutta arviointi veikin yllättäen aikaa. Monet kokeneet olisivat halunneet palata palveluun pienen tauon jälkeen, jolloin arviointiin olisi voinut ottaa toisen näkökulman tai olisi voinut keskittyä tiettyihin asioihin. Uusi arviointikierron nähtiin hyödylliseksi myös siksi, että havaintoja saa haudutella mielessään jonkin aikaa:

- *”Palaisin myöhemmin ja katsoisin, ovatko havainnot fiksuja, kun on niin paljon sälää”*
- *”Tässä on todennäköisesti joku ongelma, jota en ole huomannut, eli tämä jäisi nyt muhimaan.”*

Pari kokenutta arvioijaa totesi myös haluavansa yleensä arvioinnin välissä keskustella toisen asiantuntijan kanssa. Tauon ja yhteisen keskustelun jälkeen arvioinnin voisi tehdä loppuun tukevammalta pohjalta.

## 5.2.5 Arvioijien oma kuvaus arvioinnista

Tässä kappaleessa vertaillaan kokeneiden ja kokemattomien arvioijien omin sanoin antamia kuvauksia asiantuntija-arvioinnista heidän todelliseen arviointiinsa sekä loppuhaastattelun vastauksiin. Kuvaukset pyydettiin ennen arviointeja ja loppuhaastattelu tehtiin koko arvioinnin päätteeksi.

### **Kokeneet arvioijat**

Kokeneiden arvioijien kuvauksissa arvioinnin keskeiseksi näkökulmaksi nähtiin asiantuntijalle kertynyt kokemus aiemmista arvioinneista, testeistä ja tuotteista. Kaikki kokeneet asiantuntijat kuvasivat arviointiprosessin hieman eri tavoin, mutta yhteistä oli se, että prosessi on kehittynyt kokemuksen kautta, eikä mitään yksittäistä tunnettua menetelmää sovelleta tarkasti. Kuvauksissa tunnettuja opetettuja menetelmiä ei juuri edes mainittu nimiltä.

Kuvausten perusteella arvioinnit perustuvat kontekstin ja konseptin hahmottamiseen, käyttäjäryhmän tuntemiseen sekä yleisten, toimiviksi havaittujen suunnittelutapojen soveltamiseen. Kokeneiden arvioijien mukaan on kuitenkin tärkeää tunnistaa juuri kyseenomaisen tuotteen tarpeet ja muokata arviointia sen mukaisesti. Osa kertoi tarvittaessa soveltavansa jotakin tarkempaa menetelmää, jos tuotteen arviointi sitä tuntuu vaativan. Yleisesti mainittiin myös introspektio: asetutaan tietoisesti käyttäjän asemaan, luodaan todellisia käyttötilanteita ja tehtäviä sekä tutkitaan tuotteen soveltuvuutta niiden avulla.

Siinä missä kokemattomille arvioijille asiantuntija-arviointi tarkoitti lähinnä heuristista arviointia muistissa olleiden sääntöjen avulla, kokeneille se oli tuotteen tarkoituksen ja toiminnan sekä käyttäjäryhmän ymmärtämistä, oman kertyneen kokemuksen ja teoreettisen tiedon yhdistämistä sekä tuotteen parantamista niiden avulla. Omalla kokemuksella – tiedosta toimivista käyttöliittymäratkaisuksista, vastaavista tuotteista, ihmisen toiminnasta sekä seuratuista käytettävyydesteistä – oli kuvauksissa suuri rooli.

Viisi kokenutta arvioijaa kuudesta noudatti arvioinnissaan melko tarkasti etukäteen antamaansa kuvausta. Kuudennella arvioijalla kuvaus perustui näkemykseen siitä, miten tekisi asiantuntija arvioinnin yleensä, mutta tutkimuksessa arvioitujen osien luonne muutti hieman tätä menetelmää; asiantuntija jäi kiinni joihinkin osa-alueisiin enemmän kuin oli ajatellut. Tämän hän itse tunnisti ja kertoi arvioinnin jälkeisessä haastattelussa.

Kokeneiden arvioijien arviointi perustui pitkälti heille kertyneeseen kokemukseen aiemmista arvioinneista ja testeistä, joten koetilanteessakin (ilman sääntöjä tai muita arvioinnin apuvälineitä) he pystyivät tekemään arvioinnin totutulla ja heidän kuvaamallaan tavalla. He perustelivat tekemiään havaintoja useaan otteeseen aiemmilla kokemuksillaan, esimerkiksi: *”Olen tämän aika monta kertaa testissä nähnyt”*.

Lisäksi kokeneet arvioijat pohtivat havaintojaan useista eri näkökulmista ja suhteuttivat ne kyseisen palvelun luonteeseen. Useimmat heistä mainitsivat näiden seikkojen tärkeyden ennen arviointia antamassaan kuvauksessa. Tarkempia teoreettisia menetelmiä (kuten heuristinen arviointi ja kognitiivinen läpikäynti) he eivät näkyvästi käyttäneet, vaan sovelsivat itselleen muodostunutta omaa menetelmää, joka koostui kaikilla jossain määrin palvelun läpikäynnistä ja hyväksi havaituista toteutustavoista, suunnittelutavoista ja säännöistä.

Kokeen päätteeksi kaikilla kokeneilla arvioijilla oli selkeä ja todenmukainen kuva siitä, millä tavalla arviointi oli edennyt. Yhtä aiemmin mainittua arvioijaa lukuun ottamatta kokeen päätteeksi annetut arviointitapaa koskevat haastatteluvastaukset olivat myös yhteneviä ennen arviointia annettujen kuvausten kanssa. Lisäksi kokeneet arvioijat mainitsivat yleensä vastauksissaan myös ne seikat, jotka poikkesivat kuvauksesta ja suunnitellusta arviointitavasta. Kokeneilla arvioijilla tuntui olevan hyvä käsitys heidän käyttämistään arviointitavoista, samoin kuin vahvuuksista ja heikkouksista. He mainitsivat vastauksissaan johdonmukaisesti osa-alueet, jotka saivat eniten huomiota sekä tavat, joilla he kävivät läpi arvioitavaa tuotetta. Erityisesti kaikki arvioijat korostivat oman muodostetun tekniikkansa ja kokemuksensa merkitystä arvioinnin tekemisessä, eivätkä nähneet tarvetta esimerkiksi sääntölistoille. Kaikki arvioijat perustivat arviointinsa omaan pohjaansa, johon he myös tuntuivat luottavan. Kokeneet arvioijat olivat kriittisiä oman arviointinsa kattavuutta kohtaan: vaikka he uskoivat löytäneensä keskeisimmät ongelmakohdat, uskoivat he myös sivuuttaneensa lukuisia ongelmia.

Kokeneet asiantuntijat kokivat tutkimuksessa tehdyn arvioinnin koetilanteesta huolimatta melko tavanomaisena rajoitettuihin aikoihin, sillä todellisuudessa arviointiin on käytännössä varattu tietty aika. Ainoastaan yksittäisten sivujen arviointi ei ollut kokeneille arvioijille kovin luonnollista, sillä kontekstin ymmärtäminen oli tällä tavalla hankalampaa.

### **Kokemattomat arvioijat**

Kokemattomien arvioijien kuvaukset asiantuntija-arvioinnista olivat keskenään selvästi yhtenäisempiä kuin kokeneiden. Tämä johtui lähinnä siitä, että ne painoutuivat selvästi opeteltuihin yksittäisiin menetelmiin (lähinnä heuristiseen arviointiin) ja vastikään opittuun teoriaan. Vaikka kokemattomat arvioijat tunsivat muitakin menetelmiä jossain määrin,

useimmissa kuvauksissa kerrottiin arvioinnin olevan tutkittavan tuotteen vertaamista listoihin, sääntöihin tai muuhun teoriaan. Asiantuntijuuden merkitys nähtiin yksittäisten menetelmien hallitsemisena: asiantuntija kykenee soveltamaan menetelmiä ilman erillistä materiaalia. Erilaisista arvioituista tuotteista, aiemmista arvioinneista sekä testeistä saadun kokemuksen merkitys puuttui kokemattomien arvioijien kuvauksista lähes kokonaan.

Kuvauksille yhtenäistä oli se, että arviointia pidettiin pitkälti mekaanisena toimintana, vaikka tätä ei suoraan ilmaistukaan. Arvioinnin keskeisenä tuloksena kokemattomat arvioijat pitivät käytettävyysongelmia. Muutoksien suunnittelua tai ylipäänsä tuotteen parantamista havaintojen pohjalta ei kuvauksissa juuri mainittu. Käyttäjryhmän merkityksestä kokemattomat arvioijat eivät puhuneet lainkaan, eivätkä myöskään erityyppisten tuotteiden vaikutuksesta.

Kokemattomien arvioijien kokeessa soveltamat arviointitavat eivät useimmiten vastanneet ennen arviointia annettuja kuvauksia. Kun kuvauksissa yleisesti kerrottiin teoreettisten menetelmien soveltamisesta, todellinen arviointi jäi usein vain pinnalliseksi silmäilyksi ja yksittäisten ongelmien poimimiseksi. Tämä saattoi johtua siitä, että koeasetelmassa heillä ei ollut käytössä sääntöjä tai muita apuvälineitä arvioinnin tukena, eikä heidän kokemuksestaan ollut vielä riittävästi tukea arviointiin.

Vain yhdellä arvioijalla kuudesta ennen tutkimusta annettu kuvaus arvioinnista vastasi tutkimuksessa sovellettua arviointitapaa. Muilla arvioijilla arviointi jäi todellisuudessa annettua kuvausta hajanaisemmaksi, pinnallisemmaksi ja yksipuolisemmaksi.

Arvioinnin päätteeksi kokemattomatkin arvioijat osasivat kuitenkin kertoa melko tarkasti, kuinka arvion todellisuudessa tekivät. Useimmiten haastattelussa todettiin, että arviointi oli pitkälti haja-ammuntaa ja se eteni epäjärjestelmällisesti. Osa kokeneista myös tunnisti itse, ettei todellinen arviointi vastannut etukäteen ajateltua ja kuvattua tapaa. Monet eivät päässeet todellisiin käyttötilanteisiin ja tehtäviin käsiksi, minkä he myös kertoivat haastattelussa. Heitä häiritsi se, ettei joillakin sivuilla tarvinnut tehdä mitään, mikä olisi ohjannut arviointia luontevasti. Tämä kertoo, etteivät he osanneet oma-aloitteisesti pohtia käyttäjien tehtäviä ja tavoitteita kyseisillä sivuilla.

Jotkut kokemattomat arvioijat kertoivat käyneensä läpi mielessään niitä Nielsenin heuristiikkoja, jotka muistivat ulkoa. Tämä aiheutti sen, että arviointi jäi yksipuoliseksi ja vaillinaiseksi, kun muihin näkökulmiin ei silloin juuri osattu puuttua. Kokemattomat arvioijat pitivät tyypillisesti arviointiaan melko kattavana, he eivät uskoneet sivuuttaneensa kovin paljon ongelmia, ainakaan vakavia.

Mielenkiintoista on se, että vaikka varmuutta omasta arviointitavasta ei vielä ollut, eikä ennalta annettu kuvaus vastannut kovin hyvin arviointia käytännössä, kykenivät myös kokemattomat arvioijat kertomaan arviointiprosessistaan todenmukaisesti arvioinnin jälkeen. He pitivätkin tutkimuksessa tehtyä arviointia hieman poikkeavana ja erikoisena tilanteena tyypilliseen arviointiin nähden (ei materiaalia arvioinnin tukena).

### 5.3 Silmänliiketutkimuksen tulokset

Tässä luvussa esitellään arviointien aikana tehdyn silmänliiketutkimuksen tulokset. Silmänliikedatan analysointitavat esiteltiin kappaleessa 4.4.3. Tulokset esitetään seuraavissa kappaleissa tehtävittäin tietyillä ajanhetkillä arvioinnin alusta lähtien.

### 5.3.1 Ensimmäinen tehtävä

#### 2–4 sekuntia

Aivan välittömästi sivun avauduttua (ensimmäiset 2 sekuntia) kokeneiden arvioijien katse selvästi ohjautui visuaalisesti voimakkaan yläpalkin alapuolelle. Palkin alapuolella huomio osui sisältöalueen otsikoihin ja linkkeihin. Kokemattomien arvioijien katse ohjautui puolestaan enemmän itse palkkiin ja siinä olevaan otsikkoon. Kyseinen ero näkyi erityisen selvästi neljän sekunnin kohdalla (ks. Kuva 24). Molemmilla ryhmillä näkyi neljän sekunnin kohdalla katseen kohdistuminen sivun oikean reunan palkin yläosan kuvakkeeseen sekä sisältöalueen yläreunan otsikoihin ja linkkeihin. Nämä olivat visuaalisesti ja luonteeltaan sellaisia elementtejä, jotka tyypillisesti keräävät huomiota sivun yläosassa.



**Kuva 24.** Fiksaatioiden kohdistuminen sivun yläosaan neljän ensimmäisen sekunnin aikana (fiksaatioiden määrä). Kokemattomien arvioijien lämpökartta vasemmalla, kokeneiden vastaava oikealla.

#### 8–15 sekuntia

Erittäin pitkän sivun alareunaan kummallakaan ryhmällä ei osunut lainkaan fiksaatiopisteitä ensimmäisen kahdeksan sekunnin aikana. Tämän voisi ajatella johtuvan siitä, että sivua olisi lähdetty käymään läpi järjestelmällisesti sen alusta alkaen, koska alussa olivat myös sivun arvioinnin kannalta keskeisimmät asiat. Todellisuudessa ilmiö johtunee kuitenkin siitä, että ensimmäisen kahdeksan sekunnin aikana sivun välittömästi näkyvän osan sisältö ja toiminnot veivätkin aluksi huomion niin, ettei sivun pituutta (ja selaimen vierityspalkkia) itse asiassa vielä edes havaittu tuossa ajassa. Valtaosa arvioijista kuitenkin rullasi sivun alas ja takaisin ylös heti pituuden havaitsemisen jälkeen. Ensimmäiset fiksaatiopisteet sivun alareunassa esiintyivät 8–15 sekunnin välillä. Nämä pisteet ovat yksittäisiä, ja ovat tulleet juuri käytäessä sivun alareunassa ennen paluuta takaisin ylös. Viidentoista sekunnin kohdalla kokemattomat ovat lämpökarttojen mukaan silmäilleet kokonaan läpi oikean reunan linkkilistan, kun taas kokeneilla huomio on jakautunut sivun yläosassa tasaisemmin.

### 30 sekuntia – 2 minuuttia

Puolen minuutin kohdalla molemmat ryhmät olivat pistäneet merkkeille sivun alareunassa sivua koskevat vastuu- ja päivitystiedot. Muu alareunan sisältö muodostui aiemmista tiedotteista, jotka eivät keränneet huomiota tarkemmin. Suurin ero datassa näkyy kuitenkin minuutin ja kahden minuutin (ks. Kuva 25) lämpökarttojen kohdalla. Tässä vaiheessa kokeneiden arvioijien huomio on ollut aivan sivun alareunassa hyvin selvästi kokemattomia enemmän. Merkittävää on se, että kokemattomien arvioijien lämpökartassa näkyy lähes olemattomia muutoksia puolen minuutin ja kahden minuutin välillä, kun taas kokeneilta sivun alareuna saa juuri tällä välillä hyvin reilusti huomiota. Kokemattomat tuntuvatkin sivuuttavan sivun alareunan arvioinnissaan ensimmäisten minuuttien aikana sen jälkeen, kun ovat sen ensin vilkaisseet läpi nopeasti. Kokeneiden arvioijien kiinnostus sivun alareunaan saattaa kertoa sivun pituuden pohtimisesta, mihin viittaa myös ääneen ajattelun analyysi kyseisen tehtävässä.



**Kuva 25.** Fiksaatioiden kohdistuminen sivun alaosaan kahden ensimmäisen minuutin aikana (fiksaatioiden määrä). Kokemattomien arvioijien lämpökartta vasemmalla, kokeneiden vastaava oikealla.

Ensimmäisessä tehtävässä arvioidun sivun oikean reunan linkkilista ei ollut varsinainen valikko, minkä vuoksi toisen tehtävän kohdalla havaittua ilmiötä valikon havaitsemisen etenemistä ei ole nähtävissä tällä sivulla (vrt. 5.3.2). Molemmat ryhmät havaitsivat linkkilistan melko yhtenäisesti. Ensimmäisten minuuttien aikana kokemattomat arvioijat kiinnittivät siihen huomiota jopa kokeneita kattavammin.

Kuten myös arvioinnin toisessa tehtävässä (ks. 5.3.2), ensimmäisessä kokemattomien arvioijien katse jakautui sivulla alaspäin aiemmin, kun kokeneilla se keskittyi aluksi selvemmin sivun yläosan toimintoihin ja sisältöön. Minuutin jälkeen sisältöalueen alemman osan aloittava kuvake oli saanut selvästi kokemattomien huomion (ks. Kuva 26). Toisaalta sitten jo kahden minuutin kohdalla kokeneet arvioijat olivat ohittaneet sisältöalueen keskikohdan fiksaatioiden määrässä kokemattomat.





**Kuva 26.** Fiksaatioiden kohdistuminen sivun yläosaan ensimmäisen minuutin aikana (fiksaatioiden määrä). Kokemattomien arvioijien lämpökartta vasemmalla, kokeneiden vastaava oikealla.

Kahden minuutin arvioinnin kohdalla tehtävä fiksaatioiden keston suhteellinen tarkastelu<sup>11</sup> kertoo, että suhteellisesti eniten huomiota ovat keränneet molemmilla ryhmillä yksittäiset visuaaliset elementit, kuten:

- sivun otsikko
- sisältöalueen alussa käytetty hieman orpo väliotsikko
- punaisella esitetty linkki tapahtumien lisäämiseksi
- linkkilistan alussa oleva kuvake
- linkkilistan linkit
- pari sivulla käytettyä visuaalista otsikkoa.

<sup>11</sup> Suhteellinen tarkastelu huomioi fiksaatioiden pituudet suhteessa toisiinsa. Toisin sanoen kohde, jota on tarkasteltu pitkään, esitetään lämpökartassa punaisempana kuin kohde, jota on vilkaistu nopeasti.



**Kuva 27.** Fiksaatioiden keston jakautuminen sivun yläosassa kahden ensimmäisen minuutin aikana (fiksaatioiden suhteellinen kesto). Kokemattomien arvioijien lämpökartta vasemmalla, kokeneiden vastaava oikealla.

#### 4–10 minuuttia

Vielä neljän minuutin kohdallakaan kokeneet arvioijat eivät olleet käyneet tarkkaan ja kattavasti läpi oikean reunan linkkilistaa, vaan huomio oli alkanut jakaantua enemmän sisältöalueelle. Kokemattomien arvioijien katseet olivat kattaneet linkkilistan alas asti selvästi tarkemmin. Kokemattomat arvioijat eivät puolestaan olleet tarkastelleet sivun alareunaa vielä kolmen minuutin kohdallakaan, mutta neljän minuutin jälkeen alareuna oli alkanut saada enemmän kokemattomienkin arvioijien huomiota.

Lopulta kymmenen minuutin arvioinnin jälkeen molempien ryhmien lämpökartat näyttävät melko yhtenäisiltä. Suurin osa fiksaatioista oli osunut sivun yläreunaan, linkkilista oli tutkittu läpi kokonaisuudessaan ja sisältöalueen keskiosakin on saanut huomiota varsin tasaisesti. Suhteellisesti tarkasteltuna kymmenen minuutin kohdalla katsottaessa suurimman huomion molemmilta ryhmiltä ovat saaneet yleisesti väliotsikot ja linkit (ks. Kuva 28).



**Kuva 28.** Fiksaatioiden keston jakautuminen sivun yläosassa kymmenen ensimmäisen minuutin aikana (fiksaatioiden suhteellinen kesto). Kokemattomien arvioijien lämpökartta vasemmalla, kokeneiden vastaava oikealla.

### 5.3.2 Toinen tehtävä

#### 2–4 sekuntia

Välittömästi arvioitavan sivun avauduttua (ensimmäiset 2 sekuntia) molemmat ryhmät katsoivat erityisesti sivun otsikkoa, valikon alkua, palvelun logoa ja sen vieressä olevia yläreunan elementtejä. Kokeneiden arvioijien katse kiinnittyi ensimmäisten kahden sekunnin aikana selvästi tarkemmin sivun pääotsikkoon ja ensimmäiseen väliotsikkoon. Kokemattomien arvioijien fiksaatiot jakautuivat sivulla hieman hajanaisemmin, vaikkakin samat elementit kyllä havaittiin.

Neljän sekunnin jälkeen lämpökartasta on havaittavissa, että kokeneet arvioijat olivat tuohon mennessä jo silmäilleet valikon läpi kokonaisuudessaan (ks. Kuva 29). Ääneen ajattelun perusteella voidaan ajatella, että kokeneet arvioijat saattoivat hakea valikon avulla arvioitavalle sivulle rakennetta ja tarkoitusta sekä taustaa arvioinnin tueksi. Ero kokemattomiin vaikuttaa selvältä, sillä kokemattomat olivat tässä vaiheessa edelleen kiinnittäneet huomiota vain valikon alkuun. Sitä vastoin pääosa kokemattomien arvioijien huomiosta on kiinnittynyt yläpalkin elementteihin. Toisaalta kokemattomien arvioijien huomio oli ylipäänsä hajaantunut sivulla selvästi enemmän, sillä yksikään kokeneista arvioijista ei ollut vilkaissut sivun sisältöön vielä neljän sekunnin kohdalla. Valikon ohella kokeneiden arvioijien huomio oli muutenkin sivun yläreunassa; lähinnä yläpalkin alapuolella, kuten myös ensimmäisessä tehtävässä.



**Kuva 29.** Fiksaatioiden kohdistuminen sivun yläosaan neljän ensimmäisen sekunnin aikana (fiksaatioiden määrä). Kokemattomien arvioijien lämpökartta vasemmalla, kokeneiden vastaava oikealla.

### 8–15 sekuntia

Erot valikon ja sisällön katsomisessa ovat näkyvissä lämpökartoissa vielä kahdeksan sekunnin jälkeenkin. Yksittäisiä fiksaatioita lukuun ottamatta kokemattomat arvioijat olivat edelleen katsoleet vain valikon yläosaa, toisaalta kokeneilla arvioijilla vasta muutama fiksaatiopiste osui sivun varsinaiseen sisältöön. Kokemattomat olivat kahdeksan sekunnin jälkeen kiinnittäneet jo selvästi enemmän huomiota sisällön yläosaan.

Sivun yläreunan hieman poikkeuksellinen navigaatiovalikko (Valitse tarpeesi -pudotusvalikko) keräsi kahdeksan sekunnin aikana kokeneiden arvioijien katsetta kokemattomia enemmän. Varovasti tulkiten tämän voisi ajatella kertovan siitä, että kokeneet arvioijat pitivät arvioinnin alussa huomionsa arvioinnin kannalta tärkeissä asioissa: navigaatiossa sekä sivun tarkoituksen ja toiminnan ymmärtämisessä. Kokemattomilla arvioijilla puolestaan katseen eteneminen saattoi olla lähempänä palvelun käyttäjän katsetta sivulla.

Viidentoista sekunnin jälkeen kokemattomat arvioijat eivät vielä olleet käyneet läpi valikkoa kokonaisuudessaan. Kokeneet arvioijat olivat vilkaisseet sisällön yläosaa ja sivun alareunan elementtejä sivun rakenteen hahmottamisen vuoksi. Sisältöä sellaisenaan ei vielä tässä vaiheessa ollut tutkittu. (ks. Kuva 30.)



**Kuva 30.** Fiksaatioiden kohdistuminen sivun yläosaan viidentoista ensimmäisen sekunnin aikana (fiksaatioiden määrä). Kokemattomien arvioijien lämpökartta vasemmalla, kokeneiden vastaava oikealla.

### 30 sekuntia

Puolen minuutin jälkeen kokemattomat arvioijat olivat lämpökarttojen perusteella tarkastelleet sivua kokonaisuutena kattavammin. Esimerkiksi sivun väliotsikoita he olivat katsoneet kokeneita tarkemmin, samoin sivun sisältöä. Puoleen minuuttiin mennessä kokemattomat arvioijat olivat käyneet katseellaan läpi myös navigaatiovalikon. Kokeneiden arvioijien fiksaatiot ensimmäisen puolen minuutin aikana painottuivat selvästi enemmän sivun yläreunaan.

Tarkasteltaessa fiksaatioita suhteellisesti havaitaan, että molemmilla ryhmillä suhteellisesti suurin osuus fiksaatioista kiinnittyi samoihin yksittäisiin elementteihin (ks. Kuva 31):

- logoon ja etenkin sen vieressä olevaan Matkaviestintä-tekstiin
- valikon ensimmäiseen linkkiin (joka on nostettu tyhjän tilan avulla muista selvemmin esiin)
- Kirjautu ulos -toimintoon
- otsikon alussa oleva kuvakkeeseen
- pudotusvalikko ja siihen liittyvä painikkeeseen
- Ohjeet-linkkiin sivun yläkulmassa.

Syy yksittäisten elementtien yhtenäiseen havaitsemiseen lienee luonnollinen: kyseisten kaltaiset erottuvat elementit keräävät tyypillisesti huomiota puoleensa, eikä katseen osumista niihin voi juuri välttää. Todennäköisesti kyseiset elementit keräisivät katsetta myös käytettävyydestin aikana todellisilta käyttäjiltä, jos samaa sivua siinä testattaisiin.



**Kuva 31.** Fiksaatioiden keston jakautuminen sivun yläosassa ensimmäisen puolen minuutin (30 sekuntia) aikana (fiksaatioiden suhteellinen kesto). Kokemattomien arvioijien lämpökartta vasemmalla, kokeneiden vastaava oikealla.

### 1–2 minuuttia

Puolen minuutin ja yhden minuutin välillä fiksaatiot keskittyivät pääsääntöisesti jo aiemmin havaittuihin alueisiin ja kohtiin molemmilla ryhmillä. Uusia alueita ei tällä välillä juurikaan havaittu, vaan huomio kiinnittyi tarkastelemaan samoja alueita tarkemmin ja syvemmin. Suhteellisesti tarkasteltuna kokemattomilla arvioijilla fiksaatioiden suurin osuus siirtyi tässä vaiheessa yläpalkista valikon ensimmäisiin kohtiin.

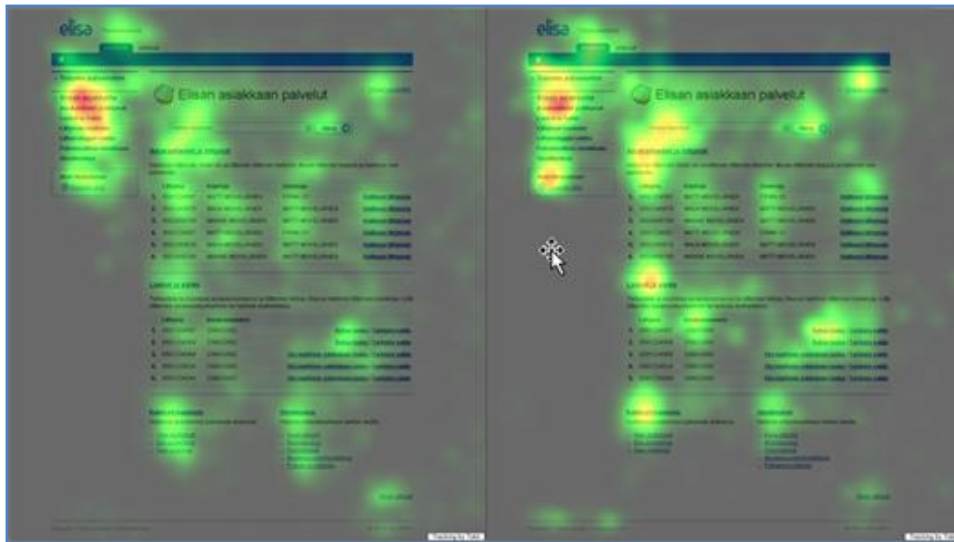
Kahden minuutin arvioinnin jälkeen (2 min) puolestaan kokeneet arvioijat olivat menneet kokemattomien ohi tarkasteltujen alueiden kattavuudessa (ks. Kuva 32). Tässä vaiheessa kokeneet arvioijat olivat tarkastelleet myös sivun varsinaista sisältöä kokemattomia johdonmukaisemmin. Sivun puoliväliä alempana olevaa sisältöä kokemattomat arvioijat olivat tarkastelleet tässä vaiheessa selvästi vähemmän. Molemmilla ryhmillä suhteellisesti selvästi suurin osa huomiosta osui tässä vaiheessa valikkoon.



**Kuva 32.** Fiksaatioiden kohdistuminen sivun yläosaan kahden ensimmäisen minuutin aikana (fiksaatioiden määrä). Kokemattomien arvioijien lämpökartta vasemmalla, kokeneiden vastaava oikealla.

#### 4–10 minuuttia

Kahden minuutin jälkeen ero sivun tarkastelemisen kattavuudessa tasoittui (etenkin alareunan osalta) vaiheittain niin, että neljän minuutin kohdalla ero oli hyvin pieni ja lopulta kymmenen minuutin kohdalla käytännössä olematon. Suhteellisesti tarkasteltuna havaitaan, että kymmenen minuutin jälkeen kokeneiden arvioijien katse oli painottunut sivun rakenteen ja navigoinnin kannalta selkeisiin elementteihin (ks. Kuva 33). Suhteellisesti suurimmat osuudet kohdistuivat valikkoon, yksittäisiin navigointielementteihin sekä sivun väliotsikoihin. Vaikka kokemattomat arvioijat kiinnittivät arvioinnin aikana huomiota pääsääntöisesti samoihin asioihin, valikkoa lukuun ottamatta kyseiset elementit eivät kuitenkaan olleet kymmenen minuutin jälkeen keränneet katsetta suhteellisesti yhtä selkeästi.



**Kuva 33.** Fiksaatioiden keston jakautuminen sivun yläosassa kymmenen ensimmäisen minuutin aikana (fiksaatioiden suhteellinen kesto). Kokemattomien arvioijien lämpökartta vasemmalla, kokeneiden vastaava oikealla.

## 6 JOHTOPÄÄTÖKSET JA POHDINTA

---

Tässä luvussa esitetään johtopäätökset tutkimuksen tulosten pohjalta. Johtopäätökset esitetään perustuen tutkimuskysymyksiin, joihin luvussa pyritään vastaamaan. Tutkimuskysymyksiin vastaamisen lisäksi johtopäätöksissä käsitellään myös joitakin muita näkökulmia, joita tulosten analyysi nosti esiin tutkimuskysymysten ulkopuolelta.

### 6.1 Arvioinnin tulokset

Luvussa painotutaan tarkastelemaan asiantuntija-arvioinnin tuloksia, erityisesti kirjattuja havaintoja ja niille annettuja korjausehdotuksia. Luku vastaa ensimmäiseen tutkimuskysymykseen (ks. 1.4.1).

#### 6.1.1 Arvioinnin syvyys

Kokeneet arvioijat tekivät arvioinnissa absoluuttisena määränä kokemattomia enemmän havaintoja lähes kaikilla eri tavoilla mitattuna. Vaikka erot eivät pääsääntöisesti ole suuria tai tilastollisesti merkitseviä, voidaan kuitenkin melko luotettavasti sanoa kokeneiden löytävän määrällisesti enemmän havaintoja, sillä erot näkyvät juuri useista eri näkökulmista tarkasteltuna. Tulos on yhdenmukainen monien aiempien tutkimuksien kanssa (mm. Jeffries ym. 1991; Nielsen 1992; Desurvire 1994).

Absoluuttista määrää keskeisempi havainto tässä tutkimuksessa on kuitenkin se, että kokeneet arvioijat pääsevät arvioinnin tuloksissa selvästi kokemattomia syvemmälle, jos arvioinnin tuloksia tarkastellaan havaintojen osa-alueiden, abstraktiotasojen tai vakavuusluokituksen mukaan. Toisin sanoen kokeneet arvioijat pureutuvat näkyvien yksittäisten ongelmien alle palvelun kannalta keskeisempiin ja laajempiin ongelmiin. Yksittäisten ja yksityiskohtaisten, helposti havaittavien ongelmien nostaminen esiin voidaan nähdä eräänlaisena käyttöliittymäoikolukuna, joka toki sekin parantaa tutkittavaa tuotetta. Monesti kuitenkin todelliset tuotteen ongelmat ovat pintaa syvemmillä, eikä niitä oikolukemalla saada kerättyä esiin. Vaikka tuttujien elementtien ja ongelmakohtien tunnistaminen (ks. 6.2.2) johti kokeneet arvioijat joskus käyttöliittymäoikolukuun, pääsivät he selvästi useammin arvioinnissaan myös syvemmälle.

Abstraktiotasolla mitattuna pienten ja keskisuurten yksittäisten käytettävyysongelmien löytämisessä ryhmien väliset erot ovat melko pieniä, mutta laajojen ongelmakohtien ja useasta yksittäisestä havainnosta muodostettujen kokonaisuuksien kohdalla kokeneet erottuivat kokemattomista selvästi. Kokeneille arvioijille tuntuu usein olevan luontevaa esittää useiden yksittäisten havaintojen sijaan jokin laajempi ongelmakokonaisuus, joka voi olla joko koostettu yksityiskohtaisista pienemmistä havainnoista tai olla syvemmillä piilevä – usein rakenteen tai konseptin ongelma, joka aiheuttaa käyttöön erilaisia pienempiä ongelmia. Lisäksi useammat erilliset ongelmat saattavat vaikuttaa toistensa vakavuuteen, vaikka muuten olisivatkin toisistaan riippumattomia.

Kokemattomat arvioijat eivät kirjanneet laajempia ongelmakokonaisuuksia samassa laajuudessa. Tähän voisi nähdä kaksi vaihtoehtoista selitystä: joko kokemattomat eivät yksinkertaisesti pysty havaitsemaan arvioinnissa laajempia ongelmakokonaisuuksia tai sitten he erittelevät havaintonsa herkemmin yksityiskohtaisten ja pienten ongelmien avulla. Tähän tämä tutkimus ei kuitenkaan vastaa, sillä se vaatisi yksittäisten ongelmien tarkastelua ja vertailua eri arvioijien välillä.



Arvioinnin osa-alueittain tarkasteltuna kokeneet kykenevät tunnistamaan kokemattomia enemmän tuotteen konseptiin ja toimintoihin liittyviä ongelmia. Selvästi näkyviin tekstisisällön, navigoinnin ja rakenteen ongelmiin myös kokemattomat pääsevät tarttumaan melko tehokkaasti. Toisin sanoen ongelmakohdat, jotka pistävät silmään, havaitaan melko tasaisesti, mutta syvemmällä tuotteessa olevat ongelmat saattavat jäädä kokemattomilta helpommin löytämättä.

Aiemmissa tutkimuksissa on todettu, että vakavimmat ongelmat löydetään arvioinnissa varmimmin arvioijasta riippumatta (Nielsen 1992; Jacobsen ym. 1998). Tämä tutkimus kuitenkin näyttää, että kokemattomilla arvioijilla on vaikeuksia löytää juuri vakavampia ongelmia tutkituista tuotteista. Arvioijien tekemän kolmiportaisen vakavuusluokittelun tarkastelu paljasti, että kaikissa luokissa kokeneet arvioijat kirjasivat kokemattomia enemmän havaintoja, mutta suhteellisesti ero on suurin vakavien ongelmien kohdalla. Toisaalta myös Desurvire (1994) on havainnut, että kokemattomat eivät löydä vakavia ongelmia yhtä tehokkaasti.

On huomattava, että ilmiö voi johtua myös siitä, että kokemattomat arvioijat eivät välttämättä luokittele tekemiään havaintoja yhtä herkästi vakaviksi ongelmiksi. Ylipäänsä arvioijien itsensä määrittelemä vakavuusluokitus ei ole tutkimuksen näkökulmasta objektiivinen ja yhtenäinen. Tämä on kuitenkin asia, jonka vaikutusta ei tämän tutkimuksen pohjalta voida sanoa varmasti. Yksi erillisen jatkotutkimuksen aihe voisikin olla vakavuusluokituksen tutkiminen; mihin arvioijat perustavat vakavuusluokituksensa, jos heille ei anneta siihen täsmällisiä ohjeita ja kuinka herkästi skaalan ääriarvoja tutkimuksessa käytetään.

Toisaalta ongelmien vakavuutta voidaan tässä tutkimuksessa pohtia myös kirjoittajan määrittelemien osa-alue- ja abstraktiotasoluokitusten kautta. Konseptia koskeva havainto tai laaja ongelmakokonaisuus voidaan ajatella tyypillisesti vakavammaksi kuin esimerkiksi yksittäinen visuaalista suunnittelua koskeva havainto tai abstraktiotasoltaan pieni havainto. Vaikka yksiselitteisesti näin ei olekaan, eikä abstraktiotasoilla ja osa-alueilla ole suoraa yhteyttä vakavuusluokituksiin, tukee niiden tarkastelu kuitenkin sitä, että laajemmat ja vaikuttavammat ongelmat jäisivät kokemattomilta arvioijilta selvästi kokeneita useammin kirjaamatta. Kun eri tarkastelutavoilla saadut tulokset havaintojen vakavuutta koskien ovat samansuuntaisia ja tukevat toisiaan, voidaan tulosta pitää melko luotettavana huolimatta eri tarkastelutapoihin liittyvistä puutteista.

Arvioijien subjektiivista arviota eri arviointikohteiden paremmuudesta selvitetessä (annettiin kouluarvosanoilla) pitivät kokemattomat arvioijat arviointikohteita parempina kuin kokeneet. Tämä on linjassa sekä havaittujen ongelmien määrän että myös vakavuusluokittelun tulosten kanssa; kokemattomat löysivät arviointikohteista vähemmän vakavia ongelmia sekä ongelmia ylipäänsä. On hankala sanoa, kuinka paljon subjektiivinen arvosana on perustunut nimenomaan kirjattuihin ongelma-kohtiin, kuinka paljon taas arvioinnin aikana muodostuneeseen mielikuvaan; on mahdollista, että kokemattomien arvioijien mielikuva arvioiduista palveluista ei ole ollut kokeneita parempi. Joka tapauksessa voidaan todeta, että kun kokematon arvioija tutkimuksen perusteella kirjaa havaintoja kokeneita vähemmän, näkee hän arvioidun osan myös yleisesti parempana kuin kokenut.

Kokemuksen merkitys arvioinnin syvyydelle voidaan tämän tutkimuksen perusteella nähdä hyvin selvänä. Myös Karoulis ja Pombortsis (2000) korostavat asiantuntijuuden merkitystä arvioinnissa. Koska esimerkiksi opiskelijoita on kuitenkin helpompi saada suorittamaan arviointia, he ovat pohtineet keinoja tukea kokemattomien arvioijien työtä esimerkiksi

ohjevihkosten ja seminaarien avulla. Lisäksi he mainitsivat, että arvioinnin laadun varmistamiseksi eritasoisille arvioijille voisi olla käytössä erilaiset sääntölistat.

### 6.1.2 Yksilölliset erot

Tutkimuksessa keskityttiin tarkastelemaan kokeneiden ja kokemattomien arvioinneissa ilmeneviä eroja. Vaikka useita selviä eroja löytyi juuri näiden ryhmien välillä, erityisen huomattavaa on se, miten suuria henkilökohtaisia eroja ryhmien sisällä oli havaittavissa. Henkilökohtaiset erot näkyivät sekä arviointitavoissa että tehtyjen havaintojen määrässä ja laadussa. Eri arvioijat kirjaavat havaintoja hyvin vaihtelevia määriä, toisaalta myös kirjattujen havaintojen abstraktiotasot vaihtelevat, joten yksiselitteisesti arvioinnin tuloksia on vaikea laittaa paremmuusjärjestykseen. Nielsen ja Molich (1990) totesivat tutkimuksessaan, että osa arvioijista löytää paljon yleisesti löydettyjä ongelmia, osa taas vähemmän, mutta myös joitakin vaikeammin löydettäviä. Aiemmin on myös havaittu, että eri arvioijat kiinnittävät huomiota erilaisiin asioihin (mm. Virzi 1997; Jacobsen ym. 1998). Tämä tutkimus vahvisti saman tuloksen. Tässä tutkimuksessa henkilökohtaiset erot näkyvät konkreettisimmin havaintojen osa-alueiden tarkastelun kautta. Tutkimuksen perusteella kaikilla arvioijilla on kokemuksesta riippumatta jokin tai joitakin osa-alueita, joihin he tavallaan ovat erikoistuneita, eli kiinnittävät niihin huomiota muita enemmän.

Osalla arvioijista tehtyjen havaintojen painopiste oli selvästi tutkittujen sivujen sisällössä tai rakenteessa, toisilla taas selvästi visuaalisessa suunnittelussa. Mielenkiintoista on erityisesti se, että henkilökohtaiset erot eri osa-alueiden painotuksissa olivat suurempia kokeneiden arvioijien kesken kuin kokemattomien kesken. Varsinkin kun muistetaan, että kokeneet arvioijat olivat saman asiantuntijayrityksen asiantuntijoita: tämä ei kuitenkaan tehnyt heidän arvioinneistaan yhtenäisiä. Slavkovic ja Cross (1999) havaitsivat tutkimuksessaan, että kaikki arvioijat ohittivat arvioinnissaan useita kategorioita, mutta olivat tehokkaita joissain niistä. He suosittelivat arviointiin tärkeimpien asioiden tarkistuslistoja ja eri osa-alueiden kohdentamista eri arvioijille. Nielsen (1992) totesi, että arvioijien pitäisi kiinnittää tietoisesti huomiota asioihin, jotka luontaisesti jäävät heiltä vähemmälle huomiolle.

Tutkimuksen perusteella kokeneet arvioijat kuitenkin useimmiten tiedostavat omat painopisteensä eri osa-alueiden välillä, kuten myös erot painotuksissa muihin kokeneisiin arvioijiin nähden. Tiedosta huolimatta tämän tutkimuksen perusteella painopisteisiin vaikuttaminen tietoisesti tai puutteellisiin osa-alueisiin keskittyminen on erittäin hankalaa. Arvioijat vain yksinkertaisesti tuntuvat olevan herkempiä havaitsemaan joitakin asioita kuin toisia. Tämä saattaa johtua siitä, että erot ovat itse asiassa syvällä arvioijan persoonallisuudessa, motivaatiossa, kiinnostuksen kohteissa, koulutustaustassa tai muussa henkilökohtaisessa kokemuksessa. Jonkin verran muista painopisteistä varmasti opitaan jo ihan siitä syystä, että kokeneet arvioijat ovat tietoisia eroista. Kokeneet arvioijat myös jossain määrin hyväksyvät eroavaisuudet ja kaipaavat tästä syystä arvioinnin jälkeen tulosten purkoa, keskustelua ja yhteenvetoa muiden arvioijien kanssa. Myös Hertzum ym. (2002) totesivat, että arvioijien väliset erot eivät välttämättä tarkoita erimielisyyttä, vaan eri lähestymistapoja ja näkökulmia samoihin asioihin. He esittivät, että erot syntyvät siitä, että arviointi vaatii tulkintaa; vaikka osa ongelmista on itsestään selviä, niin useimmat ongelmat vaativat tarkempaa vuorovaikutuksen analyysia (Hertzum & Jacobsen 2003).

Käytettävyytutkimuksessa on pitkään korostettu useamman arvioijan tarvetta kattavien tulosten saavuttamiseksi (mm. Nielsen & Molich 1990; Virzi 1997). Tässä tutkimuksessa havaitut henkilökohtaiset erot perustelevat edelleen tarvetta useamman arvioijan näkemyksen hyödyntämiselle. Nielsen (1992) päätyi siihen, että kolmesta viiteen kokeneen arvioijan käyttäminen on tyyppillisessä arvioinnissa tehokkainta suhteutettaessa arvioinnin

tulokset käytettäviin resursseihin. Kun kolmen arvioijan ryhmää ajatellaan tämän tutkimuksen kannalta, saadaan tutkimuksen tuloksiin mielenkiintoinen näkökulma yhteistuloksista. Valitsemalla kokeneiden arvioijien ryhmän sisältä mitä tahansa kolmen arvioijan joukkoja saadaan heidän arvioinneistaan selvästi kattavampi ja tasaisempi yhteistulos yksittäisen arvioijan tuloksiin verrattuna. Toisaalta mielenkiintoista on sekin, että tarkkaan valittu kolmen kokemattoman arvioijan ryhmä voi yltää arvioinnissa joitakin kolmen kokeneen arvioijan ryhmiä kattavampaan kokonaistulokseen. Joten kolmekaan arvioijaa ei välttämättä takaa kattavaa arvioinnin tulosta, mikä on todettu jo aiemmissakin tutkimuksissa.

Arviointiprosessissa henkilökohtaiset erot näkyivät myös selvästi, mutta selvää yhteyttä tuloksiin prosesseista ei ole havaittavissa. Tavallisin ero arviointiprosessissa liittyi havaintojen kirjaamistahtiin: osa arvioijista kirjaa havaintoja ylös yksitellen, osa taas pohtii kokonaisuuksia ensin pidempään ja kirjaa havaintoja ryppäissä. Arvioinnin alussa osa kokeneista arvioijista alkaa kirjata selvästi havaittavia yksityiskohtia ylös välittömästi päästäkseen sen jälkeen pureutumaan syvemmällä piileviin ongelmiin. Osa taas kykenee näkemään sirpaleisten yksityiskohtaisten ongelmien alle keskeisiin ongelmakokonaisuuksiin ja kirjaamaan niitä heti ylös. Kokemattomat arvioijat eivät pääsääntöisesti päässeet kovin usein lainkaan yksityiskohtia syvemmälle.

Hertzum ja Jacobsen (2003) pohtivat, johtavatko yksilölliset erot siihen, että arvioinnit eivät tuota valideja tuloksia. Tämän tutkimuksen perusteella kysymykseen voisi vastata, että käytännön kannalta, toisin sanoen tuotteiden parantamiseksi tulokset ovat varmasti käyttökelpoisia. Kokonaan eri asia kuitenkin on, kuinka vertailukelpoisia eri arvioijien tulokset ovat keskenään.

### 6.1.3 Käytettävyysoingelman määrittely

Luvussa 6.1.1 todettiin, että kokeneet arvioijat löytävät ongelmia määrällisesti kokemattomia enemmän. Samaan tulokseen on päädytty myös aiemmissa tutkimuksissa, joissa arvioijien välisiä eroja on tutkittu juuri tehtyjen havaintojen määrien avulla (mm. Nielsen 1992; Desurvire 1994). Monissa tutkimuksissa arvioidut tuotteet ovat kuitenkin olleet hyvin rajattuja tai muuten sellaisia, että ongelmien määrittelemine on ollut hyvin suoraviivaista (mm. Nielsen & Molich 1990; Nielsen 1992). Kun ongelmajoukko ei ole yksiselitteinen, on määrällinen vertailu vaikeaa, ja kaikkien ongelmien määrän hahmottaminen on yksinkertaisesti vaikeaa (Sears 1997).

Kuten kirjattujen havaintojen analyysi paljastaa (ks. 5.1.4), havaintojen määrälliseen vertailuun liittyy kuitenkin se merkittävä ongelma, että arvioijat kirjaavat havaintoina ylös hyvin eritasoisia kokonaisuuksia. Näin ollen havaintojen abstraktiotaso hankaloittaa suoraan absoluuttisista määristä tehtäviä johtopäätöksiä. Kun havaintojen absoluuttinen määrä vaihtelee sekä ryhmien välillä että sisällä, vaihtelee myös yksittäisten havaintojen abstraktiotaso. Siinä missä yksi arvioija saattaa kirjata joukon yksittäisiä havaintoja, niin toinen saattaa muodostaa samoista havainnoista yhden laajemman kokonaisuuden, jonka kirjaa yhtenä havaintona. Syyt eritasoisiin kokonaisuuksiin ovat varmasti yksilökohtaisia, mutta ainakin kokeneilla arvioijilla oli usein lähtökohtana jotenkin luonnollinen kokonaisuus, jolle on määritettävissä myös luonteva ratkaisuehdotus.

Tämän tutkimuksen pohjalta voidaan sanoa, että tehtyjen havaintojen määrällinen vertailu luotettavasti vaatisi ensin käytettävyysoingelman käsitteen määrittelyä. Myös Hertzum ja Jacobsen (2003) totesivat, että ilman ongelmakriteeristöä käytettävyysoingelmien

erottaminen muista on hankalaa, ja ongelman käsite tulisi määritellä, jotta arvioijilla olisi yhtenäinen kuva käytettävyysoongelmasta.

Käytettävyytutkimuksessa puhutaan yleisesti käytettävyysongelmista, ja kyseessä onkin tutkimusalueen kannalta erittäin keskeinen ja tunnettu käsite. Siitä huolimatta käsitettä ei juuri ole ainakaan kattavasti tai tunnetusti määritelty. Vakavuusluokituksen näkökulmasta käytettävyysongelmaa on toki määritelty (mm. Nielsen & Molich 1990; Karat 1994), lisäksi ongelmille on muodostettu erilaisia osa-alueisiin perustuvia luokittelutapoja (mm. Hvannberg & Law 2003; Chattrichart & Lindgaard 2003). Kuitenkin kokonaan määrittelemättä on se, että millä abstraktiotasolla liikutaan; puhutaanko yhden ongelman kohdalla yksityiskohdasta vai isosta kokonaisuudesta? Käytettävyysongelmien taksonomia (Keenan ym. 1999) on hyvä lähtökohta ongelmien globaalien tason analyysiin ja vertailuun, mutta siihen pitäisi pystyä liittämään myös mainittu abstraktiotasojen näkökulma. Lisäksi huomion arvoista on se, että käytettävyyden käsitteen määrittely vaikuttaa varmasti myös käytettävyysongelman määrittelyyn; onko arvioinnissa taustalla esimerkiksi ISO:n (ISO 9241-11) määritelmä, joka ottaa huomioon myös esimerkiksi tuotteen konseptin ja sisällön soveltavuuden, vai Nielsenin (1993) määritelmä, joka jättää ne määritelmän ulkopuolelle.

Käytännön arviointiprojekteissa käytettävyysongelman käsitteellisellä määrittelyllä ei useimmiten ole suurta merkitystä, koska niissä keskitytään yksinkertaisesti parantamaan tutkittavaa tuotetta. Tällöin kaikki tuotteen käyttöä ja käyttäjää millä tahansa tavalla mahdollisesti häiritsevät seikat voidaan mieltää käytettävyysongelmiksi. Tätä ajatusta noudattivat myös Jeffries ym. (1991), jotka määrittelivät ongelman miksi tahansa käyttäjän toimintaa vaikeuttavaksi tekijäksi. Myöskään ongelman abstraktiotasolla ei käytännön arvioinnissa ole juuri merkitystä – keskeistä on vain se, että ongelma kuvataan sellaisella tasolla, että siihen on luontevaa antaa sopiva korjausehdotus.

Käsitteen määrittelylle onkin tarvetta oikeastaan vain tämän tutkimuksen kaltaisissa tilanteissa, joissa eri arviointien tuloksia halutaan vertailla keskenään ja suhteuttaa toisiinsa luotettavasti. Ericsson ja Smith (1991) toteavat, että asiantuntijuuden tutkimisessa keskeistä on juuri muodostaa kriteerit, joilla paremmuus voidaan yksiselitteisesti selvittää. Ilman käytettävyysongelman käsitteen määrittelyä arvioijien paremmuuden vertailu löydettyjen ongelmien perusteella on kyseenalaista. Jos yleistä määrittelyä käytettävyysongelman käsitteelle ei onnistuta luomaan, jää vertailtavuuden varmistaminen tutkijan vastuulle. Tutkimuksen tekijällä on tällöin käytännössä kolme vaihtoehtoa:

- määrittää etukäteen abstraktiotaso, jolla arvioijia pyydetään kirjaamaan havaintojaan
- muokata analyysivaiheessa havainnoista keskenään samantasoisia kokonaisuuksia
- tarkastella analyysivaiheessa tuloksia useammasta näkökulmasta, kuten tässä tutkimuksessa on tehty.

Käytettävyysongelman määrittelemättömyys näkyi tutkimuksessa myös siinä, että kokemattomilla arvioijilla oli keskenään vaihtelevia käsityksiä siitä, millaisia asioita tulisi kirjata ylös käytettävyysongelmina ja mitkä taas jäivät arvioinnin ulkopuolelle. Kokemattomilta jäi tästä syystä kirjaamatta useita selkeitä havaintoja, jotka he kuitenkin mainitsivat ääneen arvioinnin aikana. Kokeneilla arvioijilla näkyi arviointia observoitaessa selvästi lähtökohta, jonka mukaan käytettävyysongelmina ylös kirjataan kaikki sellaiset havainnot, jotka saattaisivat haitata tuotteen käyttöä millä tahansa tavalla. Tämä vastaa tyyppillistä käytännön arviointiprojektia. Yksittäisten ongelmien vaikutuksesta käyttäjän toimintaan toki kokeneetkin arvioijat saattoivat olla keskenään eri mieltä.

Vaikka ero käytettävyysongelman käsitteen tulkinnassa ryhmien välillä on tämän tutkimuksen kannalta mielenkiintoinen, korostaa se entisestään tarvetta yleiselle

käytettävyysohjelman määrittelylle. Tämä tutkimus ei tuo vastausta siihen, miten käsite tulisi määritellä, mutta asiaa olisi mahdollista selvittää eri tavalla rajatun jatkotutkimuksen avulla.

#### 6.1.4 Käytännön tarpeet tulosten esittämisessä

Kokeneille arvioijille käytännön arviointiprojekteista kertynyt kokemus tuli näkyvimmin esiin tulosten esittämisessä, muotoilussa ja priorisoinnissa. Kokeneiden arvioijien esittämät havainnot olivat selvästi kokemattomia useammin ratkaisulähtöisiä, eli ne perustuivat nimenomaan arvioidun ratkaisun parantamiseen, ei niinkään vain ongelman löytämiseen kuten usein kokemattomilla arvioijilla.

Ratkaisulähtöisyys näkyy siten, että kokeneet arvioijat muotoilevat usein havainnon ja ongelmakuvauksen korjausehdotuksen näkökulmasta (esim. ”tämän tulisi olla siten, että...”). Kokeneet arvioijat myös muodostavat usein havainnoista korjauksen kannalta mielekkäitä laajempia kokonaisuuksia, joilla parannetaan kerralla luontevasti useita pienempiä ongelmakohtia. Vertailun vuoksi voidaan todeta, että kokemattomat laativat ehdotukset melko suoraan yksittäisistä havainnoista, tai sitten korjaukset ovat turhan abstrakteja.

Usein käytännössä ratkaisuehdotusten kannalta joudutaan pohtimaan laajempia kokonaisuuksia, eikä pelkkien yksityiskohtien korjaaminen irrallisesti välttämättä johda hyvään lopputulokseen. Lisäksi arvioinnin tilanteelle asiakkaalle laajojen kokonaisuuksien havaitseminen on usein arvokkaampaa tietoa, vaikka pienten ongelmien korjaaminen olisikin ehkä helpompaa ja halvempaa. Käytännön arvioinnissa kun olennaista ei ole tehtyjen havaintojen määrä, vaan se, kuinka tuotteeseen tarvittavat muutokset saadaan parhaiten vietyä eteenpäin. Tavoite toki määritellään erikseen jokaisessa arviointityössä.

Kokeneiden arvioijien esittämissä havainnoissa näkyy muutenkin se, että he ovat käytännön työssä oppineet esittämään ja perustelemaan tekemiään havaintoja tutkimuksen tilaajalle, kun pelkkä viittaus tiettyyn sääntöön ei riitä perusteluksi. Tämä lisää arvioinnin luotettavuutta ja toisaalta arvioijan uskottavuutta. Myös Capra (2006) havaitsi kokeneiden tuottavan kokemattomia laadukkaampia ongelmakuvauksia. Tässä tutkimuksessa arvioijia ei pyydetty erikseen muotoilemaan ongelmaotsikoita, joten useimmat arvioijat (kokeneet ja kokemattomat) kirjasivatkin ne lähinnä omina muistiinpanoina ylös muotoilematta niitä tarkemmin.

Vaikka perusteluja ja taustoja kirjatulle havainnoille ei tutkimuksessa pyydetty, kokeneet arvioijat perustelivat usein havaintoja suullisesti varsin kattavasti, ja toisaalta pohtivat ääneen eri näkökulmia. Kokeneet arvioijat kiinnittivät myös kirjattujen havaintojen esittämisessä kokemattomia enemmän huomiota havaintojen välittämiseen eteenpäin, niiden esitystapaan ja ymmärrettävyyteen – vaikkakin eri tavoilla. He luokittelivat, jäsentelivät ja kirjoittivat kuvauksia auki kokemattomia useammin. Lisäksi he useammin muodostivat havainnoista korjausten kannalta sopivampia kokonaisuuksia. Lähtökohtana havaintojen muotoilussa oli, että ne olisivat selkeästi hyödynnettävissä palvelua parannettaessa.

Kokeneiden arvioijien arvioinnin keskeisin tavoite – tuotteen parantaminen – näkyi tutkimuksessa myös korjausehdotuksien kohdalla. Pyydettyä tuotteen arvioinnin päätteeksi kolmea keskeisintä korjausta kokeneet arvioijat eivät useimmiten pelkästään maininneet kolmea kirjattua ongelmaa korjattavaksi (kuten kokemattomat useimmiten tekivät), vaan priorisoivat korjattavia asioita ja toisaalta muodostivat suurempia kokonaisuuksia, joiden korjaaminen olisi mielekkästä. He saattoivat välillä myös mainita

ainoastaan yhden heidän mielestään tärkeän asian, jota muuttamalla tuote kokonaisuudessaan kuitenkin paranee merkittävästi (esim. *"Jos keskitytte tähän, niin tästä tulee jo huomattavasti parempi"*). Kokeneiden arvioijien korjaukset kuitenkin kattavat lähes aina vakavimmiksi luokitellut ongelmat, kun taas kokemattomilla arvioijilla ei aina ole selkeää yhteyttä korjausehdotusten ja vakavimpien ongelmien välillä. Korjausehdotusten rooli onkin mielenkiintoinen, sillä esimerkiksi Sawyer ym. (1996) korostavat tarkkojen parannusehdotusten merkitystä yhdessä vakavuusluokituksen ja neljän muun tekijän ohella käytännön arvioinnin laadun varmistamiseksi. Toisaalta arviointimenetelmät eivät kuitenkaan suoraan tarjoa keinoja ongelmien korjaamiseen (Nielsen & Molich 1990). Shneiderman (1998) jättäisikin korjausehdotukset suunnittelijoille, ei arvioijille. Tämä tutkimus kuitenkin antaa viitteitä, että kokeneille arvioijille korjausehdotusten laatimisesta on tullut keskeinen osa arviointiprosessia, ja ne syntyvät luontevasti osana ongelmien kirjaamista ylös.

Kokeneiden arvioijien priorisointi viestii käytännön kokemuksesta; usein käytännön projekteissa kaikkiin tarvittaviin muutoksiin ei ole aikaa tai muita resursseja, joten on keskityttävä vain olennaisimpiin korjattaviin kohtiin. Tutkimuksen mukaan kokeneet arvioijat osaavat jo oma-aloitteisesti arvioinneissaan huomioida tämän ja priorisoida jonkin keskeisen seikan muiden edelle. Käytännön arvioinnissa olisi myös erityisen tärkeää määritellä arvioinnin tavoite ja tuotteen valmiusaste; pyritäänkö arvioinnissa löytämään suunnittelun tueksi keskeisiä suuntaviivoja ja ratkaisuja, vai onko kyseessä vain käyttöliittymän siistiminen, johon esimerkiksi heuristisen arvioinnin on todettu sopivan (Nielsen & Molich 1990).

## 6.2 Arviointiprosessi

Tässä luvussa esitetään johtopäätökset asiantuntija-arvioinnin prosessin ja etenemisen osalta. Luku keskittyy vastaamaan toiseen tutkimuskysymykseen (ks. 1.4.2).

### 6.2.1 Silmäily vai järjestelmällinen läpikäynti?

Kokeneille arvioijille on muodostunut arviointia varten erilaisia tekniikoita, menetelmiä ja strategioita, joilla he pyrkivät käymään läpi arvioitavaa tuotetta mahdollisimman kattavasti tai ainakin tuotteelle olennaisilta osilta. Kokeneet arvioijat kykenevät esimerkiksi käymään arvioitavaa palvelua läpi järjestelmällisesti eri näkökulmista tai käytettävyyden osa-alue kerrallaan, jolloin arvioitava tuote tulee katettua kattavammin. Lisäksi siinä missä kokemattomat arvioijat saattoivat ajautua huomaamattaan käyttäjän rooliin, oli kokeneilla arvioijilla tutkimuksessa havaittavissa kyky siirtyä tarkoituksenmukaisesti käyttäjän asemaan arviointia varten. Kokeneet arvioijat siirtyivät käyttäjän asemaan aina tarvittaessa, he loivat todenmukaisia tehtäviä ja hypoteeseja sekä pyrkivät pääsemään todelliseen käyttökontekstiin. Arvioijien on todellisuudessa erittäin tärkeää päästä pureutumaan todellisen käytön mukaiseen tilanteeseen ja todellisiin tehtäviin (Sears 1997; Shneiderman 1998). Ja vaikka arviointi on käytännössä nopea ja helppo tehdä, tuotteeseen perehtymisen merkitystä korostetaan usein (mm. Sawyer ym. 1996; Shneiderman 1998).

Tämä tutkimus ei kuitenkaan pysty vastaamaan siihen, miten edellä mainitut menetelmät ovat muodostuneet; ovatko ne tulleet tiedostamatta pelkästään suoritettujen arviointien tai käytettävyydestien määrän seurauksena vai ovatko arvioijat itse vaikuttaneet arviointitapansa kehittymiseen tarkkailtuaan omaa arviointiaan. Joka tapauksessa kokeneet arvioijat tuntuvat tietävän oman arviointiprosessinsa sekä arviointinsa heikkoudet ja vahvuudet melko hyvin.

Kokemattomilla arvioijilla ei tutkimuksen perusteella ole valmiuksia soveltaa vastaavanlaisia menetelmiä, tai ylipäänsä käydä läpi tutkittavaa tuotetta täysin johdonmukaisesti. Heidän arviointitapansa perustuu useimmiten yksinkertaisesti tutkittavan tuotteen silmäilyyn ja siihen, että sivuja silmäillessä esiin nousevat asiat ohjaavat arviointia. Kokeneisiin verrattuna kokemattomien arvioijien läpikäyntitapa on satunnainen ja impulsiivinen, minkä vuoksi arvioinnin tulokset saattavat jäädä helposti pinnallisiksi. Tyypillisten ja helposti tunnistettavien virheiden lisäksi palveluita koskee useimmiten myös syvemmällä olevia käytettävyysongelmia, jotka liittyvät tarkemmin esimerkiksi juuri kyseisen palvelun rakenteeseen, toimintalogiikkaan, sisältöön tai konseptiin. Näiden ongelmien havaitseminen tai tunnistaminen ei yleensä onnistu välittömästi ja automaattisesti, vaan se vaatii ratkaisun analysoimista tarkemmin todellisen käyttötilanteen kannalta. Toisin sanoen porautuminen arvioinnissa pintaa syvemmälle vaatii jonkin tekniikan tai menetelmän soveltamista.

Ajoittain itse asiassa kokemattomien arvioijien tapa hahmottaa ja tarkastella arvioitavia sivuja muistuttaa tapaa, jolla sivustojen todelliset käyttäjät niitä tarkastelevat. Usein käytettävyysteesteissäkin havaitaan, että käyttäjät silmäilevät sivua aluksi hieman jäsentymättömästi, minkä jälkeen sivu ja sillä käytetyt elementit alkavat itsessään ohjata havaitsemista ja sivun käyttöä. Kokemattomat arvioijat ajautuivat arvioinnin aikana usein käyttäjän rooliin tiedostamattaan, eli pohtivat ratkaisuja omalta kannaltaan, vaikka eivät todelliseen käyttäjäryhmään itse kuuluneetkaan. Myös ääneen ajattelu oli ajoittain samankaltaista käytettävyytestien käyttäjien kanssa. Toki on huomattava, että palvelun käyttäjän ja sitä arvioivan henkilön tavoitteet ovat erilaisia sivuja tarkasteltaessa.

Tutkimuksen kuuluneen silmänliiketutkimuksen tulokset antoivat lisäksi joitakin viitteitä kokemattomien arvioijien havaitsemisen yhtäläisyyksistä käytettävyysteesteissä tehtyihin havaintoihin käyttäjien hahmottamisesta. Kun tietyllä hetkellä melko pian sivun avaamisen jälkeen kokeneiden arvioijien huomio oli arvioinnin kannalta keskeisissä elementeissä, kokemattomilla se muistutti tyypillistä käytettävyydestä, eli katse kiinnittyi tietyllä tavalla hallitseviin, katsetta kerääviin ja ohjaaviin elementteihin. Johtopäätökseen on kuitenkin suhtauduttava hieman varauksellisesti, sillä tutkimuksessa arvioituja sivuja ei kuitenkaan testattu käyttäjien kanssa, joten luotettavaa vertailua ei voida tehdä. Mahdollisissa jatkotutkimuksissa yksi kiinnostava näkökulma voisi kuitenkin olla käyttäjän näkökulman vertaaminen tarkemmin kokeneiden ja kokemattomien arvioijien tapaan havainnoida arvioitavia tuotteita.

## **6.2.2 Kokemuksen tuoma pohja**

Arvioinnin suorittaminen ja havaintojen tekeminen vaatii jonkinlaista pohjaa, jota vasten tuotetta analysoidaan. Tässä luvussa kerrotaan kokemuksen tuoman pohjan vaikutuksesta arviointiin.

### **6.2.2.1 Epävarmuus, varmuus ja kyseenalaistaminen**

Kokemattomien arvioijien työssä arvioinnin aikana näkyi selvästi epävarmuutta havaintojen tunnistamisessa ja kirjaamisessa. Epävarmuus näkyi ääneen ajattelun avulla siten, että useita arvioijien ääneen pohtimia teemoja jäi kirjaamatta havainnoiksi, koska arvioijat eivät niistä olleet varmoja. Kokemattomat arvioijat eivät esimerkiksi olleet varmoja, ovatko jotkut havainnot todellisia käytettävyysongelmia, ja jättivät ne ehkä kirjaamatta. Vertailun vuoksi voidaan todeta, että kokeneet yleensä kirjasivat kaikki havainnot, mutta jättivät useammin havainnon vakavuusasteen luokittelematta, jos eivät olleet varmoja sen todellisesta vaikutuksesta.

Kokemattomat arvioijat myös välillä selvästi arkailivat tuoda pohdintojaan esiin. Epävarmuus johtui todennäköisesti siitä, että ilman sääntölistoja tai muita apuvälineitä he eivät saaneet riittävästi tukea tai perustelua arvioinnilleen. Kokemattomat arvioijat epäroivät monien selvien ongelmakohtien kohdalla, koska eivät olleet varmoja havaintojen relevanssista. Tätä kautta epävarmuus heijastui myös arvioinnin tuloksiin. Vaikka arvioijat sitä eivät ääneen maininneetkaan (tai osanneet mainita), heidän arviointinsa tuntui kärsivän siitä, että heillä ei ollut materiaalia tukena. Osa kokemattomista arvioijista yritti esimerkiksi selvästi soveltaa heuristista arviointia ja Nielsenin kymmentä sääntöä (Nielsen 1993), mutta koska he eivät muistaneet niitä kaikkia, jäivät arvioinnin tulokset vaillinaisiksi.

Kokeneilla arvioijilla varmuus oli arvioinnissa selvästi suurempaa. Kokeneilla arvioijilla arviointi pohjautuu pääosin heidän omaan kokemukseensa, johon he myös luottavat vahvasti. Kokeneet arvioijat eivät soveltaneet arvioinnissaan ainakaan näkyvästi heuristista arviointia tai kognitiivista läpikäyntiä, eivätkä he juuri koskaan viitanneet esimerkiksi Nielsenin sääntöihin, sitä vastoin he perustelivat havaintojaan lähinnä omin sanoin. Silti heidän arviointiprosessinsa yhdisteli mainittuja menetelmiä jollakin tasolla. He saattoivat pohtia toimintojen tai navigoinnin etenemistä selvästi läpikäynnin tapaisesti, mutta tarttua taas tiettyihin vaiheisiin, sivuihin tai elementteihin enemmän heuristisen arvioinnin keinoin. Tutkimuksen perusteella vaikuttaakin siltä, että kokeneiden arvioijien prosessissa olisi samoja elementtejä kuin heuristisessa läpikäynnissä (Sears 1997).

Oma kokemus tuntuu toimivan sellaisena pohjana, johon arvioinnissa voidaan luottavaisesti tukeutua. Arvioijat voivat olla varmempia yksittäisistä havainnoista, koska he ovat törmänneet niihin aiemmissa arvioinneissa ja käytettävyytsteissä. Lisäksi koko arviointiprosessi etenee sujuvammin, kun kaikkia havaintoja ei tarvitse pohtia yksityiskohtaisesti. Monet puutteellisesti suunnitellut tai toteutetut elementit tunnistetaan ensisilmäyksellä ja osataan mainita havaintoina välittömästi. Siinä missä kokemattomat arvioijat etenivät arvioinnissa usein vertaamalla sääntöjä ja periaatteita käyttöliittymäratkaisuihin, niin kokeneet etenivät monesti toisinpäin: he näkivät ja tunnistivat ongelman suoraan, mutta hakivat sille sen jälkeen tarvittaessa perustelun.

Kokeneiden arvioijien kokemus on tuottanut heille eräänlaisia tietokokonaisuuksia, malleja ja skeemoja tyypillisten elementtien tai ongelmakohtien havaitsemiseen. Näin ollen kokeneet arvioijat kykenevät nopeasti näkemään tutkittavassa tuotteessa tuttuja ja toistuvia – usein puutteellisia ratkaisuja ja elementtejä. Ilmiö on tuttu kognitiivisen psykologian tutkimuksesta, jossa esimerkiksi shakki-ekspertheille (mm. Simon & Chase 1973) ja lääketieteen ekspertheille (mm. Patel & Groen 1991) on todettu muodostuneen kyky tunnistaa omalla ongelmallaan tyypillisiä rakenteita. Esimerkiksi lääketieteen asiantuntijat kykenevät muodostamaan diagnoosin nopeasti yksittäisten oireiden perusteella (Patel & Groen 1991). Arvioinnissa ilmiö näkyy siten, että kokeneet arvioijat pystyvät tunnistamaan yksityiskohtaisten pienten ongelmien takana olevan laajemman kokonaisuuden, joka todellisuudessa vaatii korjausta. Toisaalta valmiit mallit näkyvät ajoittain myös astetta kehittyneemmin: kokeneet arvioijat saattavat ongelmakohtien sijaan tunnistaa välittömästi optimaalisen ja toimivan ratkaisun kyseiseen tilanteeseen. Tällöin kokeneilla arvioijilla on mielessään kuva hyvästä ja toimivasta ratkaisusta, johon arvioitavaa sovellusta verrataan. Eteneminen ratkaisu edellä on luonnollista myös arvioinnin tulosten esittämisen kannalta (ks. 6.1.4).

Kokeneiden arvioijien itsevarmuuden ja kokemattomiin nähden automaattisemman arvioinnin vaikutus arvioinnin tuloksiin ei kuitenkaan ole aivan yksiselitteinen. Epävarmuus voi johtaa arvioinnin kannalta terveeseen ja hyödylliseen kyseenalaistamiseen, kun taas itsevarmuus pahimmillaan virhetulkintoihin kontekstin riittämättömän huomioimisen vuoksi.



Esimerkiksi kokenut arvioija voi omien malliensa ja skeemojensa pohjalta tunnistaa tietyn ongelmakohtan, vaikka se ei juuri kyseisessä palvelussa tai kontekstissa olisikaan ongelma. Toisaalta myös yleisesti hyvä ratkaisu voi olla huono jossakin kontekstissa. Tutun tilanteen tunnistaminen voi johtaa siihen, että se mielletään ongelmaksi vanhoilla perusteilla, kyseenalaistamatta arviota riittävästi (esim. *”Olen tämän aika monta kertaa nähnyt”*). Kokemattomat arvioijat joutuvat vastaavien skeemojen puuttuessa analysoimaan ratkaisua aina tarkemmin juuri kyseisessä tilanteessa ja kontekstissa. Toisaalta kokeneet arvioijat pystyvät kyseenalaistamaan ratkaisuja kokemattomia paremmin tarkastelemalla niitä useasta eri näkökulmasta. Tämän vuoksi tutkimuksen perusteella onkin mahdotonta todeta, mikä on epävarmuuden ja toisaalta kyseenalaistamisen vaikutus arvioinnin tuloksiin. Joka tapauksessa itsevarmuudesta ja tunnistamisesta johtuva vaara on olemassa. Asiaa tulisi kuitenkin tutkia tarkemmin luotettavien johtopäätösten saamiseksi.

#### 6.2.2.2 Arvioinnin pohja

Tutkimuksessa arvioinnit tehtiin asiantuntija-arvioineina ilman tukena käytettäviä sääntölistoja tai muita apuvälineitä. Arvioinnit nojautuivat siis arvioijien omaksumiin menetelmiin sekä omaan näkemykseen ja kokemukseen. Tarkempia menetelmiä kuten heuristinen arviointi tai kognitiivinen läpikäynti ei kuitenkaan kielletty käyttämästä, mutta tehtäväksi annettiin vain etsiä tuotteesta sen käyttöä hankaloittavat käytettävyysongelmat. Asetelma ei tuntunut vaikuttavan kokeneiden arvioijien työhön, mutta kokemattomat tuntuivat tulosten valossa siitä hieman kärsivän. Heillä tuntuu arvioinnissa olevan vielä tarve sitoa havainnot vahvasti oman näkemyksen turvana toimiviin sääntöihin. Kokeneilla arvioijilla taas säännöt ovat useimmiten kasvaneet osaksi omaa kokemusta ja ehkä osittain sellaisenaan kadonneetkin heille kehittyneiden mallien ja skeemojen tieltä.

Tutkimuksen perusteella johtopäätöksenä voidaankin pitää sitä, että arviointiin tarvitaan aina joku tuki, jonka pohjalta tuotteen ongelmakohtia ja muita havaintoja voidaan luotettavasti tunnistaa. Arvioijan kokemus toimii tehokkaasti tällaisena tukena, jos sitä on kertynyt riittävästi aiempien arviointien ja käytettävyydestien kautta, kuten tutkimuksessa kokeneilla arvioijilla. Jos oma kokemus ei yksin riitä arvioinnin tueksi, voidaan parempiin tuloksiin päästä joidenkin apuvälineiden kanssa, kuten esimerkiksi sääntölistat, ohjeet, läpikäyntimallit ja muut vastaavat. Tutkimuksen pohjalta vaikuttaakin siltä, että kokemattomat olisivat voineet päästä parempiin tuloksiin, jos he olisivat saaneet käyttäen tukeaan joitakin arvioinnin apuvälineitä, kuten esimerkiksi sääntölistoja. Varmaksi tätä ei kuitenkaan tämän tutkimuksen pohjalta voida sanoa.

Mielenkiintoinen koeasetelma olisikin sellainen, jossa kokemattomille arvioijille annettaisiin tueksi esimerkiksi heuristiseen arviointiin tai kognitiiviseen läpikäyntiin tarvittavaa aineistoa, ja kannustettaisiin niiden käyttöön. Kokeneet arvioijat taas tekisivät edelleen arviointinsa vain oman näkemyksensä pohjalta. Toisaalta kiinnostavaa olisi myös tutkia, mikä vaikutus olisi sillä, jos myös kokeneet rajautuisivat arvioinnissaan puhtaasti esimerkiksi heuristiseen arviointiin ja sen sääntölistoihin – olisiko sillä positiivinen vai negatiivinen vaikutus heidän tuloksiinsa?

Toinen mielenkiintoinen kysymys on, missä vaiheessa arvioijan kokemus ylittää sen rajan, että oma kokemus riittää yksin kattavan arvioinnin tueksi? Käytettävyyden opiskelijoille opetetaan arviointia sääntölistojen ja ohjeiden avulla, toisaalta tämän tutkimuksen perusteella kokeneet arvioijat nojaavat arviointinsa yksinomaan kokemukseensa. Kokemukseen ovat toki sulautuneet jollakin tavalla myös kyseiset säännöt ja teoria, mutta valtaosa kokemuksesta on silti tullut arvioinneissa tutkittujen tuotteiden ja varsinkin käytettävyydestien kautta. Jokainen käytettävyydestiä seurannut tai ohjannut tietää, miten

konkreettisesti ja tehokkaasti käytettävyydestä seuraamalla voi omaksua tietoa. Tutkimuksen pohjalta vaikuttaakin siltä, että eron tekevään kokemukseen vaikuttaa suoritettujen asiantuntija-arviointien määrän sijaan enemmän juuri käytettävyydestien vetämiseen liittyvä historia. Tämä tutkimus ei kuitenkaan kerro, missä vaiheessa ja miten siirtyminen arvioinnin apuvälineistä oman kokemuksen hyödyntämiseen tapahtuu – eli missä vaiheessa arvioijan oma kokemus muodostaa yksin riittävän pohjan arvioinnin tueksi niin, että apuvälineitä ei enää tarvita.

Edellä mainittua kysymystä voisi pohtia myös käytettävyyden arvioinnin opetuksen näkökulmasta. Useimmiten opetuksessa keskitytään yleisimpiin menetelmiin, kuten heuristiseen arviointiin ja kognitiiviseen läpikäyntiin, ja niiden jollakin tasolla formaaliin soveltamiseen. Voitaisiinko opetuksessa jollakin tavalla ottaa jo huomioon kokemuksen kertyessä tapahtuva muutos arvioinnin tekemisessä ja pohjustaa kokemukseen nojaavaa arviointia jo aikaisessa vaiheessa? Vai onko niin, että oikotietä ei ole, vaan kehittyminen vaatii teorian ja tarkkojen menetelmien opettelemista alussa, ja sen jälkeen yksinkertaisesti työtä arviointien ja käytettävyydestien parissa näkemyksen kerryttämiseksi? Joka tapauksessa askel ensimmäiseen suuntaan voisi olla heuristisen arvioinnin ja läpikäyntien yhdistelmien (esim. Heuristinen läpikäynti, Sears 1997) käyttöön kannustaminen, koska kokeneiden arvioijien prosessi tuntuu olevan enemmän sen suuntainen. Searsin tutkimuksen mukaan heuristinen läpikäynti auttaa kokemattomia arvioijia, ja hän olettaa myös kokeneiden hyötyvän menetelmästä. Tämän tutkimuksen pohjalta oletusta on tavallaan vaikea vahvistaa tai kumota: kokeneiden arviointitavoissa on jo olemassa samoja elementtejä heuristisen läpikäynnin kanssa, mutta epäselvää on, auttaisiko menetelmän käyttäminen heitä vielä parempaan tulokseen.

Myös muita keinoja voidaan pohtia kokemattomien arvioijien arviointien parantamiseen. Karoulis ja Pombortsis (2000) korostavat asiantuntijuuden merkitystä, ja pohtivat, miten esimerkiksi opiskelijoiden arviointia voitaisiin tukea niin, että tulokset olisivat lähempänä kokeneita. He loivat kokemattomien arvioijien tueksi ohjevihkosien sekä seminaarin, joilla arvioijia opastettiin. Lisäksi he pohtivat mahdollisuutta tarjota erilaisia heuristisia sääntöjä vastaamaan eritasoista kokemusta. Skov ja Stage (2005) puolestaan kehittivät työkalun ongelmien tunnistamiseen (*Conceptual tool for problem identification*), joka auttaa kokemattomia pääsemään arvioinneissaan parempiin tuloksiin niin määrässä kuin vakavuudessaakin. Näitä lähestymistapoja voi puolustaa myös tämän tutkimuksen avulla.

### 6.2.3 Arviointiin käytettävä aika

Tutkimuksessa arviointiin käytettävä aika oli rajoitettu – lähinnä käytännön syistä. Useimmat arvioijat totesivat arvioinnin jälkeen ajan olleen heidän suorittamalleen arvioinnille riittävä, ainakin kerralla tehtävää arviointia varten. Jotkut kokemattomat arvioijat olisivat tosin omien sanojensa mukaan käyttäneet kerralla aikaa enemmän. Kokemattomat arvioijat useimmiten käyttivät ajan loppuun saakka, mutta kokeneet aina eivät. Suurin ero oli kuitenkin siinä, että useimmat kokemattomat arvioijat eivät uskoneet mahdollisen lisäajan tuovan merkittäviä muutoksia tuloksiin.

Kokeneet arvioijat pitivät käytössä ollutta aikaa kerralla arviointiin riittävänä, mutta sitä vastoin monet heistä olisivat halunneet palata palvelun arviointiin uudelleen hieman myöhemmin. He olisivat halunneet käydä arvioitavaa tuotetta läpi uudelleen annettuaan asian hautua mielessään tai etenkin keskusteltuaan muiden asiantuntijoiden kanssa havainnoista, mitä pidettiin erittäin tärkeänä. Palatessaan palveluun toisella arviointikerralla he voisivat myös ottaa arviointiin jonkin toisen näkökulman.

Tämä tutkimus ei siis vastaa siihen, mikä olisi rajoittamattoman ajan vaikutus arvioinnin tuloksiin ryhmien välillä. On mahdollista, että kokemattomat pääsisivät tuloksissa lähemmäs kokeneita, mutta on myös mahdollista, että kokeneet arvioijat pääsisivät vapaalla ajalla noudattamaan paremmin omaa totuttua arviointiprosessiaan, mikä taas johtaisi parempiin ja huoleellisempiin tuloksiin. Rajoittamattoman ajan avulla saataisiin tutkittua arvioijien välisiä eroja, mutta on huomattava, ettei tilanne olisi kuitenkaan kovin luonnollinen, sillä käytännön työssä aika on aina rajattu (Sears 1997). Ajan vaikutusta voitaisiin tutkia esimerkiksi siten, että arvioinnin kohde annettaisiin arvioinnin tekijöille vapaasti esimerkiksi pariksi päiväksi, minkä jälkeen tulokset toimitettaisiin tutkimuksen tekijälle raportoituna.

Toinen mahdollinen variaatio olisi jakaa käytettävissä oleva aika kahteen osaan, joiden välillä on tauko. Monet arvioijat sanoivat haastattelussa, että olisivat halunneet jättää asian hautumaan joksikin ajaksi. Arviointi voitaisiin tällä tavalla suorittaa esimerkiksi kahdessa eri istunnossa peräkkäisinä päivinä, jolloin mielenkiintoista olisi verrata tuloksia yhdellä istumalla tehtyyn arviointiin.

- Anderson, J. R. (2000) *Cognitive psychology and its implications*. W.H. Freeman and Company, New York, 5th edition.
- Anzai, Y. (1991) Learning and use of representations for physics expertise. Ericsson, K.A. & Smith, J. (Eds.) *Toward a general theory of expertise: prospects and limits*. Cambridge University Press.
- Bastien, J.M.C. & Scapin, D.L. (1995) Evaluating a user interface with ergonomic criteria. *International Journal of Human-Computer Interaction*, Vol. 7, no. 2, pp. 105-121.
- Capra, M. (2006) Usability problem description and the evaluator effect in usability testing. Dissertation thesis, Virginia Polytechnic Institute and State University.
- Card, S.K., Moran, T.P. & Newell, A. (1983) *The psychology of human-computer interaction*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Chattratchart, J. & Brodie, J. (2003) HE-Plus – Towards usage-centered expert review for website design. *Proc. forUse 2003*, Ampersand Press, pp. 155-169.
- Chattratchart, J. & Lindgaard, G. (2008) A comparative evaluation of heuristic-based usability inspection methods. *Conference on Human Factors in Computing Systems*, pp. 2213-2220.
- Chi, M.T.H., Feltovich, P.J. & Glaser, R. (1981) Categorization and representation of physics problems by experts and novices. *Cognitive Science*, Vol. 5, Iss. 2, pp. 121-152.
- Cuomo, D.L. & Bowen, C.D. (1994) Understanding usability issues addressed by three user-system interface evaluation techniques. *Interacting with computers*, Vol. 6, no. 1, pp. 86-108.
- de Groot, A.D. (1965) *Thought and choice in chess*. The Hague: Mouton.
- Desurvire, H., Kondziela, J. & Atwood, M.E. (1992) What is gained and lost when using evaluation methods other than empirical testing. *People and Computers VII: Proceedings of HCI 92*, York., pp. 89-102.
- Desurvire, H., Lawrence, D. & Atwood M. (1991) Empiricism versus judgement: Comparing user interface evaluation methods on a new telephone-based interface. *ACM SIGCHI Bulletin*, Vol. 23, no. 4, pp. 58-59.
- Desurvire, H.W. (1994) Faster, cheaper!! Are usability inspection methods as effective as empirical testing? Nielsen, J. & Mack R.L. (Eds.) *Usability inspection methods*. John Wiley & Sons, Inc.
- Doubleday, A., Ryan, M., Springett, M. & Sutcliffe, A. (1997) A comparison of usability techniques for evaluating design. *Proceedings of the 2nd conference on Designing interactive systems: processes, practices, methods, and techniques*. pp. 101-110.
- Dumas, J.S. & Redish, J.C. (1993) *A Practical Guide to Usability Testing*. Norwood. Ablex Publishing Company, New Jersey.
- Ericsson, K.A. & Smith, J. (1991) Prospects and limits of the empirical study of expertise: an introduction. Ericsson, K.A. & Smith, J. (Eds.) *Toward a general theory of expertise: prospects and limits*. Cambridge University Press.

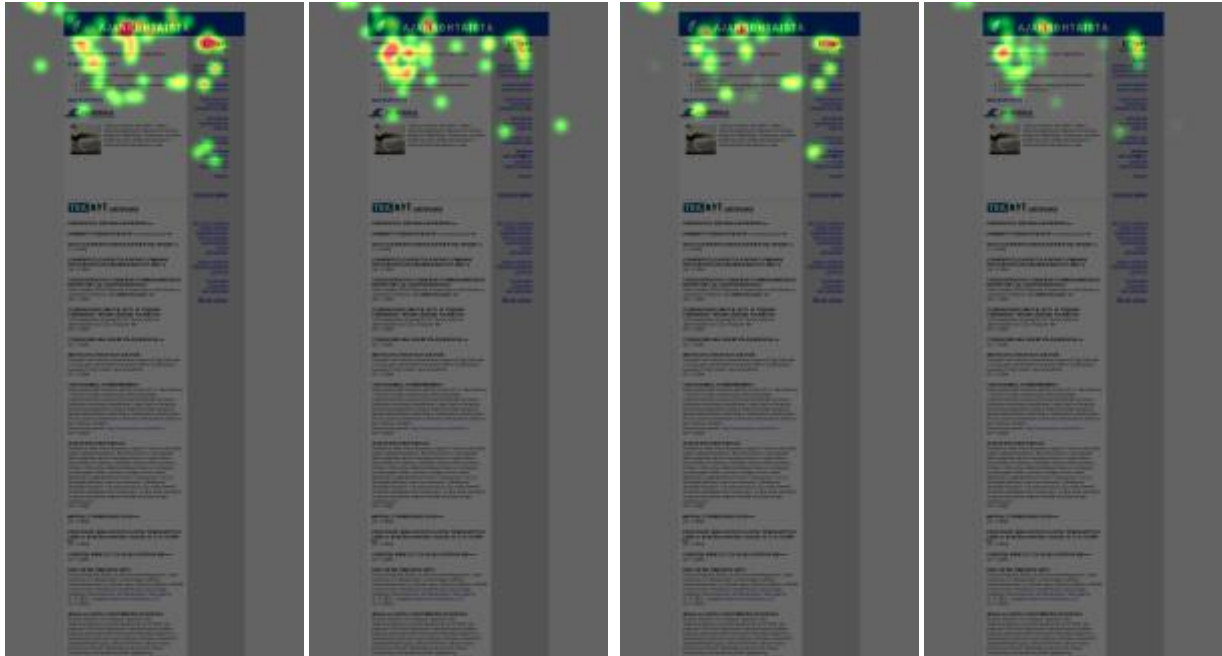
- Fitts, P. M. & Posner, M. I. (1967) *Human Performance*. Brooks Cole, Belmont.
- Goldberg, J.H. & Kotval, X.P. (1998) Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, Vol. 24, No. 6, pp. 631-645.
- Goldberg, J.H., Stimson, M.J., Lewenstein, M., Scott, N. & Wichansky, A.M. (2002) Eye Tracking in Web Search Tasks: Design Implications. *Proceedings of the 2002 symposium on Eye tracking research & applications*. pp. 51-58.
- Gould, J.D. & Lewis, C. (1983) Designing for usability -- Key principles and what designers think. *Proceedings of the ACM CHI'83 Conference on Human Factors in Computing Systems*. pp. 50-53.
- Hertzum, M. & Jacobsen, N. (1999) The Evaluator Effect during First-Time Use of the Cognitive Walkthrough Technique. *Human-Computer Interaction: Ergonomics and User Interfaces. Proceedings of HCI International '99*, Vol. I, pp. 1063-1067.
- Hertzum, M. & Jacobsen, N. (2003) The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-Computer Interaction*, Vol. 15, Iss. 1, pp. 183-204.
- Hertzum, M., Jacobsen, N. E. & Molich, R. (2002) Usability inspections by groups of specialists: perceived agreement in spite of disparate observations. *CHI '02 extended abstracts on Human factors in computing systems*, pp. 662-663.
- Hollinsed, T. & Novick, D.G. (2007) Usability inspection methods after 15 years of research and practice. *Proceedings of the 25th annual ACM international conference on Design of communication*, pp. 249-255
- Hvannberg, E. & Law, E. (2003) Classification of Usability Problems (CUP) Scheme. *Interact*, pp. 655-662.
- Instone, K. (1997) Site Usability Heuristics For The Web. <http://instone.org/heuristics> (viitattu 22.6.2008).
- ISO/DIS 9241-11 (1996) Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability. *International Organization for Standardization*.
- ISO/FDIS 13407 (1999) Human-centred design processes for interactive systems. *International Organization for Standardization*.
- Jacobsen, N., Hertzum, M. & John, B. (1998) The Evaluator Effect in Usability Tests. *CHI 98 conference summary on Human factors in computing systems*, pp. 255-256.
- Jeffries, R. & Desurvire, H. (1992) Usability testing vs. heuristic evaluation: was there a contest? *ACM SIGCHI Bulletin*, Vol. 24, Iss. 4, pp. 39-41.
- Jeffries, R., Miller, J., Wharton, C. & Uyeda, K. (1991) User interface evaluation in the real world: a comparison of four techniques. *Proceedings of the ACM CHI'91 Conference on Human Factors in Computing Systems*. pp. 119-124.
- Karat, C. (1994) A comparison of user interface evaluation methods. Nielsen, J. & Mack R.L. (Eds.) *Usability inspection methods*. John Wiley & Sons, Inc.
- Karat, C., Campbell, R. & Fiegel, T. (1992) Comparison of empirical testing and walkthrough methods in user interface evaluation. *Proceedings of the ACM CHI'92 Conference on Human Factors in Computing Systems*, pp. 397-404.

- Karat, J. (1997) User-centered software evaluation methodologies. Helander, M., Landauer, T.K. & Prabhu, P. (Eds.) *Handbook of human-computer interaction*, 2nd ed. Elsevier Science. pp. 689-704.
- Karoulis, A. & Pombortsis, A. (2000) Heuristic Evaluation of Web-sites: The Evaluators' Expertise and the Heuristic List. Aristotle University of Thessaloniki, 7 pages.
- Keenan, S., Hartson, H., Kafura, D. & Schulman, R. (1999) The Usability Problem Taxonomy: A Framework for Classification and Analysis. *Empirical Software Engineering*, Vol. 4, Issue 1, pp.71-104.
- Laarni, J. (2004) Silmänliikkeiden rekisteröinti käyttöliittymien tutkimuksessa. *Psykologia* 39 (2004), Iss. 2, pp. 134-142.
- Law, E. & Hvannberg, E. (2004) Analysis of strategies for improving and estimating the effectiveness of heuristic evaluation. *Proceedings of the third Nordic conference on Human-computer interaction*, pp. 241-250.
- Lewis, C., Polson, P.G., Wharton, C. & Rieman, J. (1990) Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*, pp. 235-242.
- Loftus, G.R. & Mackworth, N.H. (1978) Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 4, No. 4, pp. 565-572.
- Molich, R. & Nielsen, J. (1990) Improving a human-computer dialogue. *Communications of the ACM*, Vol. 33, no. 3, pp. 338-348.
- Nielsen, J. (1992) Finding usability problems through heuristic evaluation. *Proceedings of the ACM CHI'92 Conference on Human Factors in Computing Systems*, pp. 373-380.
- Nielsen, J. (1993) *Usability Engineering*. Morgan Kaufmann Publishers, San Diego.
- Nielsen, J. (1994) Enhancing the explanatory power of usability heuristics. *Proceedings of the ACM CHI'94 Conference on Human Factors in Computing Systems*. pp. 152-158.
- Nielsen, J. (1995) Scenarios in discount usability engineering. Carroll, J.M. (Ed.) *Scenario-based design: Envisioning work and technology in system development*. John Wiley & Sons, Inc. pp. 59-83.
- Nielsen, J. & Landauer, T.K. (1993) A mathematical model of the finding of usability problems. *Proceedings of the Association for Computing Machinery INTERCHI'93*. pp.206-213.
- Nielsen, J. & Mack R.L. (1994) Usability inspection methods. Nielsen, J. & Mack R.L. (Eds.) *Usability inspection methods*. John Wiley & Sons, Inc.
- Nielsen, J. & Molich, R. (1990) Heuristic evaluation of user interfaces. *Proceedings of the ACM CHI'90 Conference on Human Factors in Computing Systems*. pp. 249-256.
- Patel, V.L. & Groen, G.J. (1991) The nature of medical expertise: a critical look. Ericsson, K.A. & Smith, J. (Eds.) *Toward a general theory of expertise: prospects and limits*. Cambridge University Press.
- Polson, P., Lewis, C., Rieman, J. & Wharton, C. (1992) Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies*, Vol. 36, Iss. 5, pp. 741-773.

- Riihiaho, S. (2000) Experiences with usability evaluation methods. Licentiate's thesis. Helsinki University of Technology.
- Saariluoma, P. (1990) *Taitavan ajattelun psykologia*. Kustannusosakeyhtiö Otava, Helsinki.
- Saariluoma, P. (1988a) Ihmisen muisti. Hautamäki, A. (Ed.) *Kognitiotiede*. Gaudeamus, Helsinki.
- Saariluoma, P. (1988b) Ajattelu kognitiivisena prosessina. Hautamäki, A. (Ed.) *Kognitiotiede*. Gaudeamus, Helsinki.
- Salvucci, D.D. & Anderson, J.R. (2001) Automated Eye-Movement Protocol Analysis. *Human-Computer Interaction*, Vol. 16, Iss. 1, pp. 39-86
- Sawyer, P., Flanders, A. & Wixon, D. (1996) Making a difference - The impact of inspections. *Proceedings of the ACM CHI'96 Conference on Human Factors in Computing Systems*. pp. 376-382.
- Sears, A. (1995) Heuristic walkthroughs: Combining the advantages of existing evaluation techniques. *Proceedings of the Third Annual Mid-Atlantic Human Factors Conference*, pp. 29-35.
- Sears, A. (1997) Heuristic walkthroughs: Finding the problems without the noise. *International Journal of Human-Computer Interaction*, Vol. 9, no. 3, pp. 213-234.
- Shneiderman, B. (1998) *Designing the User Interface*. Addison Wesley Longman. Third Edition.
- Simon, H.A. & Chase, W.G. (1973) Skill in chess. *American Scientist*, 61, pp. 394-403
- Skov, M., & Stage, J. (2005) Supporting problem identification in usability evaluations. *Proceedings of the 19th conference of the computer-human interaction special interest group*, pp. 1-9.
- Slavkovic, A. & Cross, K. (1999) Novice heuristic evaluations of a complex interface. *CHI '99 extended abstracts on Human factors in computing systems*, pp. 304-305.
- Sloboda, J. (1991) Musical expertise. Ericsson, K.A. & Smith, J. (Eds.) *Toward a general theory of expertise: prospects and limits*. Cambridge University Press.
- Smith, S.L. & Mosier, J.N. (1986) Guidelines for designing user interface software. [http://www.userlab.com/Downloads/Smith\\_Mosier\\_guideline\\_.pdf](http://www.userlab.com/Downloads/Smith_Mosier_guideline_.pdf) (viitattu 22.6.2008).
- Tversky, A. & Kahneman, D. (1974) Judgment under uncertainty: Heuristics and Biases. *Science* 27, Vol. 185, No, 4157, pp. 1124-1131.
- Wharton, C., Rieman, J., Lewis, C. & Polson, P. (1994) The cognitive walkthrough method: A practitioner's guide. Nielsen, J. & Mack R.L. (Eds.) *Usability inspection methods*. John Wiley & Sons, Inc.
- Vilbergsdóttir, S., Hvannberg, E. & Law, E. (2006) Classification of usability problems (CUP) scheme: augmentation and exploitation. *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*, pp. 281-290
- Virzi, R.A. (1997) Usability inspection methods. Helander, M., Landauer, T.K. & Prabhu, P. (Eds.) *Handbook of human-computer interaction*, 2nd ed. Elsevier Science. pp. 705-715.

# LIITE A: SILMÄNLIIKETUTKIMUKSEN TULOKSET LÄMPÖKARTTOINA

## Tehtävä 1



**Sarja 1.** Fiksaatioita kuvaavat lämpökartat kahden sekunnin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 2.** Fiksaatioita kuvaavat lämpökartat neljän sekunnin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).





**Sarja 3.** Fiksaatioita kuvaavat lämpökartat kahdeksan sekunnin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



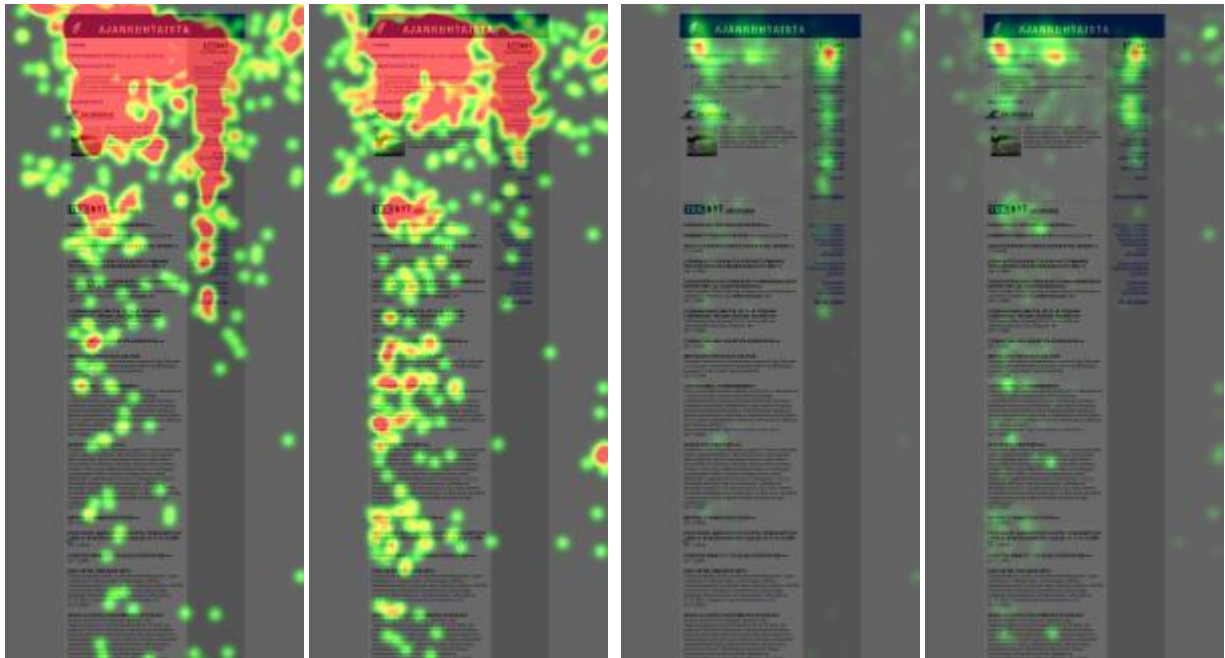
**Sarja 4.** Fiksaatioita kuvaavat lämpökartat viidentoista sekunnin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 5.** Fiksaatioita kuvaavat lämpökartat kolmenkymmenen sekunnin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



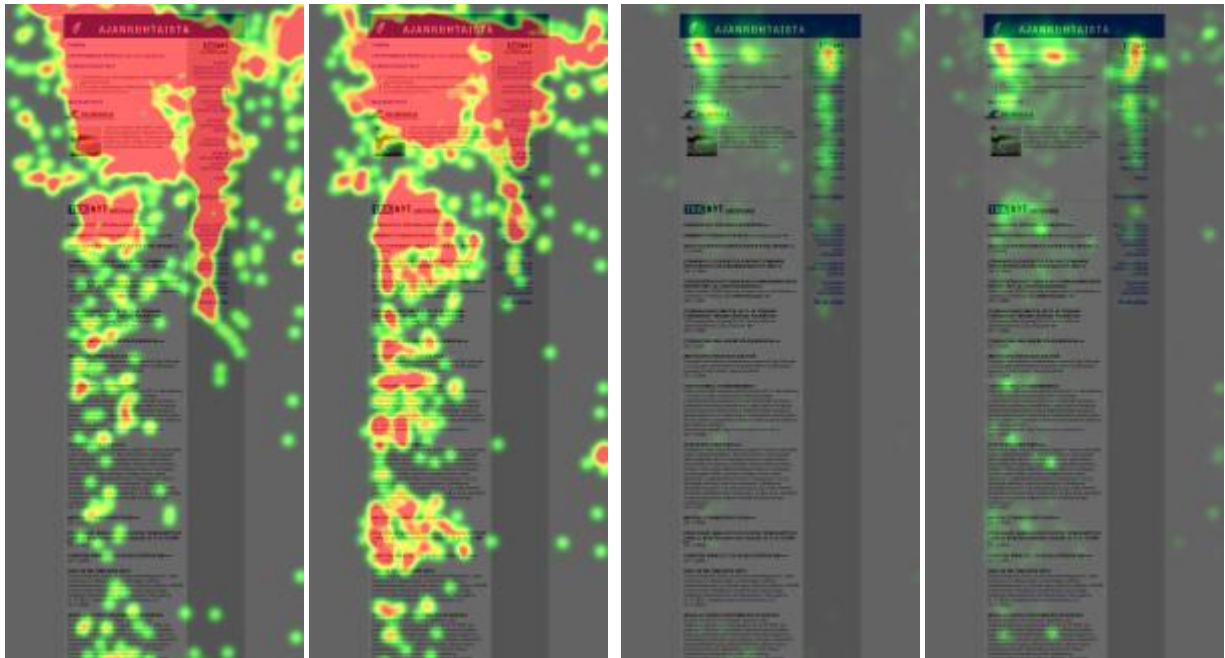
**Sarja 6.** Fiksaatioita kuvaavat lämpökartat yhden minuutin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 7.** Fiksaatioita kuvaavat lämpökartat kahden minuutin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 8.** Fiksaatioita kuvaavat lämpökartat kolmen minuutin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 9.** Fiksaatioita kuvaavat lämpökartat neljän minuutin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 10.** Fiksaatioita kuvaavat lämpökartat kymmenen minuutin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).

## Tehtävä 2



**Sarja 11.** Fiksaatioita kuvaavat lämpökartat kahden sekunnin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 12.** Fiksaatioita kuvaavat lämpökartat neljän sekunnin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 13.** Fiksaatioita kuvaavat lämpökartat kahdeksan sekunnin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 14.** Fiksaatioita kuvaavat lämpökartat viidentoista sekunnin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



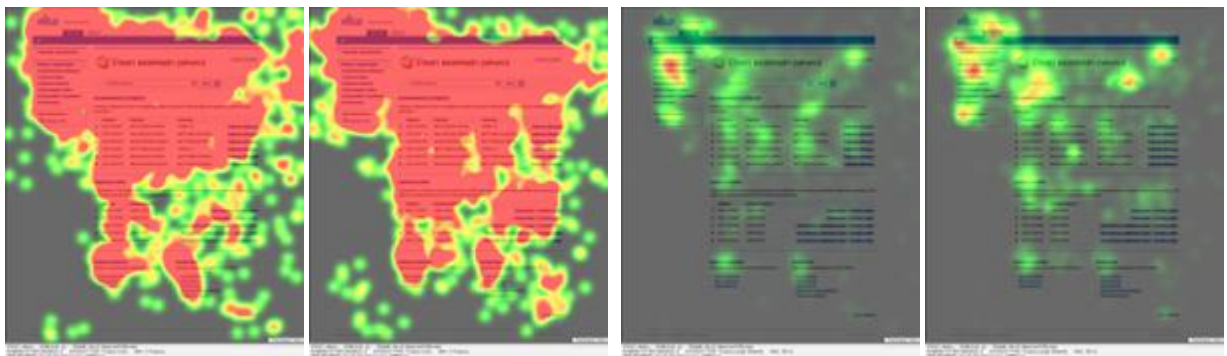
**Sarja 15.** Fiksaatioita kuvaavat lämpökartat kolmenkymmenen sekunnin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



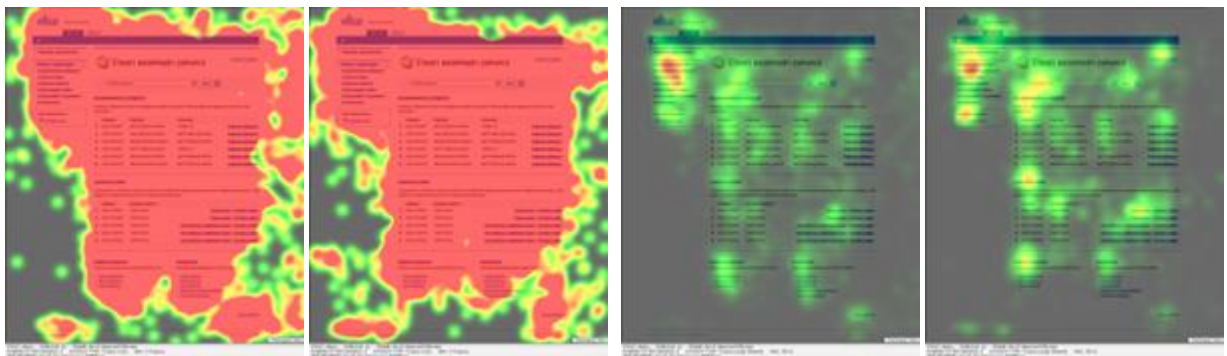
**Sarja 16.** Fiksaatioita kuvaavat lämpökartat kahden minuutin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 17.** Fiksaatioita kuvaavat lämpökartat kolmen minuutin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 18.** Fiksaatioita kuvaavat lämpökartat neljän minuutin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).



**Sarja 19.** Fiksaatioita kuvaavat lämpökartat kymmenen minuutin arvioinnin jälkeen. Kuvat vasemmalta oikealle: fiksaatioiden määrä (kokemattomat), fiksaatioiden määrä (kokeneet), fiksaatioiden suhteellinen kesto (kokemattomat) ja fiksaatioiden suhteellinen kesto (kokeneet).