

Arvioijan vaikutus

19

Raija Perälä

Arvioijan vaikutusta on alettu laajemmin tutkia 1990-luvun loppupuolella. Tutkimus on kohdistunut pääasiassa käytettävyydestä, heuristisen arvioinnin, ääneenajattelun ja kognitiivisen läpikäynnin tuloksiin. Arvioijan vaikutuksella tarkoitetaan sitä, että käytettävyyssiantuntijat arvioidessaan samaa kohdetta samalla menetelmällä havaitsevat olennaisesti erilaisia ongelmia. Tässä työssä etsitään syitä arvioijan vaikutukseen sekä keinoja sen hallintaan. Huomiota kiinnitetään ohjeistuksen, raportoinnin ja ongelmien priorisoinnin merkitykseen. Arvioijan vaikutusta tarkastellaan myös käytettävyysongelmien tunnistamisen näkökulmasta.

19.1. Johdanto

Arvioijan vaikutusta (evaluator effect) käytettävyystudkimuksen tuloksiin on alettu laajemmin tutkia 1990-luvun loppupuolella. Tällöin tutkimus ulotettiin myös muihin käytettävyystudkimuksen arviointimenetelmiin kuin heuristiseen arviointiin; itse asiassa heuristisen arvioinnin tulosten koonnissa yhteen saatiin ensimmäisiä havaintoja arvioijan vaikutuksesta jo 1990-luvun alussa (Nielsen, 1992). Eräs varhaisimmista termiä arvioijan vaikutus käyttäneistä tutkimuksista on Jacobsenin ja muiden (1998) tutkimus arvioijan vaikutuksesta käytettävyydestä tuloksiin. Sitten arvioijan vaikutusta on tarkasteltu lähinnä kolmen eri arviointimenetelmän eli heuristisen arvioinnin, kognitiivisen läpikäynnin sekä käytettävyydestä näkökulmasta (Hertzum & Jacobsen, 2001). Lavery ja muut (1997) käyttävät arvioijan vaikutuksesta termiä *arvioijan harhaisuus (analyst bias)*. Tässä työssä on valittu käyttöön neutraalimpi termi – arvioijan vaikutus.

Arvioijan vaikutuksella tarkoitetaan siis sitä, että käytettävyyssiantuntijat arvioidessaan samaa järjestelmää samaa tutkimusmenetelmää käyttäen havaitsevat olennaisesti erilaisia käytettävyyso ongelmia. Vaikutus ilmenee sekä ongelmien löytämisessä että niiden vakavuusluokituksessa. (Hertzum & Jacobsen, 2001; Jacobsen et al., 1998).

Koska arvioijan vaikutus aiheuttaa sattumanvaraisia arviointituloksia, sen voidaan katsoa vaikuttavan ennen kaikkea tutkimuksen luotettavuuteen eli *reliabiliteettiin (reliability)*. Vaikutus ulottuu kuitenkin myös tutkimuksen *validiteettiin (validity)*. Käytettävyyssiantuntijan tehtävänä on löytää tutkittavasta kohteesta ongelmia, mutta myös päättää, mikä ongelma on käytettävyyso ngelma. Käytettävyyso ngelma voi jäädä kokonaan havaitsematta tai raportoiduksi tulee ongelmia, jotka eivät reaaliaimailmassa ole käytettävyyso ngelmia lainkaan. Siten kahden arvioijan kokoamat raportit havaitsemistaan ongelmista eivät koskaan ole täysin samanlaisia. (Cockton & Woolrych, 2002; Hertzum & Jacobsen, 2001)

Jo Nielsen (1992) havaitsi, että asiantuntijuus ja arviointikokemus vaikuttavat siihen, millaisia ja miten paljon ongelmia arvioija löytää. Kuitenkin myös muut seikat vaikuttavat tuloksiin eikä kokenutkaan arvioija pysty löytämään kaikkia ongelmia. Tässä työssä perustellaan, miksi arvioijan vaikutusta pitäisi tutkia, ja kartoitetaan sen tutkimisessa olevia haasteita. Luvussa käydään aluksi läpi käytettävyyso ngelman tunnistamiseen liittyviä asioita, jonka

jälkeen esitellään tarkemmin aihepiiristä tehtyjä tutkimuksia. Luvun lopussa hahmotellaan keinoja, joilla arvioijan vaikutusta tuloksiin voitaisiin pienentää.

19.2. Käytettävyyden arviointi

Käytettävyyden arviointi koostuu arvioinnin suunnittelusta, tiedon keräämisestä, analysoinnista ja raportoinnista (Cockton et al., 2003). Joka vaiheessa tarvitaan mukaan käytettävyyssiantuntijaa; asiantuntijan tehtävät vaihtelevat siis valittujen arviointimenetelmien mukaan, mutta tavoitteena on etsiä parannuskohteita arvioitavana olevasta sovelluksesta. Kohdan lopuksi kuvataan tarkemmin arvioijan vaikutusta arviointimenetelmien laatuun.

19.2.1. Käytettävyyssiantuntijan rooli ja tehtävät käytettävyyden arvioinnissa

Käytettävyyssarvioinnin suunnittelu sisältää arvioinnin tavoitteiden ja laajuuden sekä käyttäjäryhmien määrittelyn. Arviointimenetelmästä riippuen voi olla tarpeellista laatia tehtäviä tai skenaarioita. Käytettävyyttä koskevaa tietoa kerätään ja käytettävyyssongelmia etsitään usein yhtä tai useampaa menetelmää käyttäen – on havaittu, että eri menetelmillä löydetään erilaisia ongelmia. Analysointivaiheessa ongelmat luokitellaan ja niiden vakavuusaste määritellään. Jos arvioijia on monta, tulokset kootaan yhteen.

Käytettävyyssongelmat kootaan *käytettävyyssarporttiin* (*usability problem report*). Raporttiin kirjoitetaan ongelmien kuvaukset sekä kontekstit, missä ne esiintyivät, mitä heuristiikkaa ne loukkasivat tai mikä on ongelman kriteeri ja vakavuusaste. Raportissa voi olla ratkaisuehdotuksia ongelmien korjaamiseen. Raportin tiedot toimivat uuden suunnittelukierroksen lähtökohtana. Luvussa 13 (Koskinen) käsitellään tarkemmin käytettävyyssarportin sisältöä.

Käytettävyyden arviointia tekevän käytettävyyssiantuntijan rooli riippuu käytetystä arviointimenetelmästä. Empiirisissä menetelmissä mukana on *käyttäjä* (*user*) tai *testikäyttäjä* (*test user*), jolloin käytettävyyssiantuntija tekee arvioinnin tämän suorituksen ja siinä havaittujen käytön ongelmakohtien perusteella. Arvioinnissa tarvittavaa tietoa voidaan kerätä myös haastatteluilla tai kyselylomakkeilla. Analyttisissä menetelmissä kuten heuristisessa arvioinnissa käytettävyyssiantuntija tekee arvioinnin perehtymällä suoraan tutkittavaan kohteeseen.

Käytettävyyssarviointia tekevän henkilön toiminnan tavoitteena on löytää arvioitavana olevan kohteen käytettävyyssongelmat. Hänen on myös tehtävä päätös siitä, mikä tutkittavana olevan sovelluksen ongelma tai virhe on käytettävyyssongelma.

Samana kohteen arvioinnissa voidaan käyttää useampaakin menetelmää. Esimerkiksi heuristisella arvioinnilla löydettyjä käytettävyyssongelmia voidaan verrata käytettävyyssarvestin tuloksiin. Käytettävyyssiantuntijan panos ulottuu arviointiprosessin jokaiseen vaiheeseen aina aineiston keruusta analysointiin ja raportin laatimiseen.

Käytettävyyden arviointi voi olla yksilö- tai ryhmätyötä riippuen käytetystä menetelmästä. Silloinkin kun arviointi tehdään yksilötyönä, eri arvioijien havaitsemat ongelmat ja niiden vakavuusluokitukset kootaan yhteen ja jokainen löydetty ja kriteerit täyttävä käytettävyyssongelma otetaan raporttiin mukaan vain kerran.

Käytettävyyssongelmien vakavuusluokituksella varmistutaan siitä, että suunnittelussa keskitytään tärkeimpien ongelmien korjaamiseen. On tavallista, että kaikkia löydettyjä käytettävyyssongelmia ei ole mahdollista korjata, joten löydettyistä käytettävyyssongelmista on tarpeen nostaa esille kiireellisimmät ja eniten korjaamista vaativat. (Nielsen, 1993)

19.2.2. Käytettävyyssongelman kriteerit

Käytettävyyssarvioinnissa havaittujen ongelmien tulee täyttää tietyt kriteerit, jotta ne voidaan luokitella käytettävyyssongelmiksi. Kriteeristö on erilainen analyttisissä ja empiirisissä menetelmissä, sillä empiiristen menetelmien tuottamia havaintoja käyttäjän toiminnasta ei ole saatavilla analyttisissä menetelmissä.

Analyttisissä menetelmissä käytettävyyssarviointi on ennakoivaa, toisin sanoen tutkitaan, mikä sovelluksen ominaisuus tai toiminto tulee olemaan mahdollinen käytettävyyssongelma. Kognitiivisessa läpikäynnissä arviointi tehdään käyttäjän toimintojen näkökulmasta ja siinä käytettävyyssongelmien kriteerit liittyvät opittavuuteen. Kriteereinä käytetään neljää kysymystä, jotka on esitelty luvussa 9 (Ranne). Heuristisessa arvioinnissa ongelmien kriteereinä käytetään Nielsenin kymmenen heuristiikan listaa (luku 8, Korvenranta): tarkistetaan, loukkaako jokin sovelluksen ominaisuus jotakin heuristiikkaa. Muita analyttisten menetelmien käytettävyyssongelmien kriteerejä ovat mm. erilaiset suunnitteluohjeet ja standardit.

Koska empiirisissä menetelmissä arviointitapahtumassa on aina mukana todellinen tai testi-käyttäjä, arvioinnin painopiste on siinä, mitä seurauksia käytettävyyssongelmasta käyttäjälle aiheutuu. Empiirisissä menetelmissä kuten käytettävyyssongelmassa käytettävyyssongelmien kriteerit voivat olla myös mitattavia eli kuinka paljon aikaa tehtävän suorittamiseen kuluu.

Jacobsenin ja muiden (1998) tutkimuksessa, jossa menetelmänä oli käytettävyydestä, käytettiin etukäteen määriteltyjä käytettävyyssongelmien kriteerejä. Osa kriteereistä perustui suoriin, osa käyttäjän subjektiivisiin kokemuksiin sovelluksen käytöstä. Havaittu ongelma voitiin luokitella käytettävyyssongelmaksi, jos

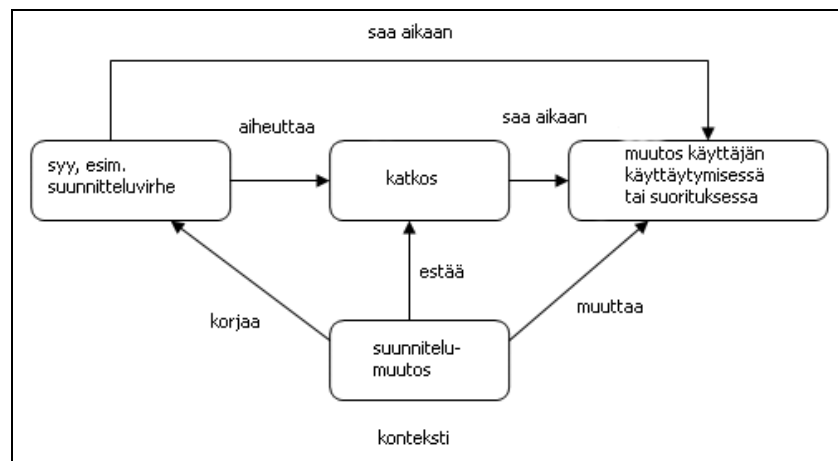
- osallistuja ei onnistunut saavuttamaan tavoitettaan kolmen minuutin sisällä,
- osallistuja luovutti ja tehtävä jäi tekemättä,
- osallistuja koetti etsiä ratkaisua vähintään kolme kertaa päästäkseen tavoitteeseensa,
- osallistuja oli hämmentynyt sovellusta käyttäessään,
- osallistujalla voitiin havaita negatiivisia tunteita tai hän sanoi, että jokin asia on ongelmallista,
- osallistuja teki suunnitteluohjeiden sovelluksen parantamiseksi tai
- järjestelmä kaatui.

Lavery ja muut (1997) määrittelevät käytettävyyssongelman sen osatekijöiden kautta (kuva 1). Käytettävyyssongelma muodostuu ongelman syystä, käyttäjän ja järjestelmän väliseen vuorovaikutukseen tulevasta *katkoksesta (breakdown)* sekä seurauksesta, joka ongelmasta aiheutuu. Katkos voi olla epätarkoituksenmukainen tavoite tai toiminto, jonka käyttäjä ongelman takia tekee, tai palaute, jota käyttäjä ei ymmärrä. Määrittelyyn vaikuttaa myös konteksti, jossa ongelma esiintyy. *Suunnittelumuutos (design change)* korjaa ongelman, poistaa katkoksen sekä muuttaa seurausta. Suunnittelumuutos tarkoittaa arvioitavan kohteen jonkin piirteen uudelleen suunnittelua, joten se ei ole varsinaisesti käytettävyyssongelman osa. Kuten luvussa 20 (Kosonen) todetaan, suunnittelumuutos ei välttämättä poista ongelmia – se voi jopa tuottaa uusia suunnitteluvirheitä.

Tiivistetysti sanottuna käytettävyyssongelman kriteereissä on otettava huomioon syy, katkos-

tilanne, seuraus sekä konteksti. Myös Cockton ja Woolrych (2002) painottavat kontekstin merkitystä käytettävyyssongelman määrittelyssä. Havaittu käytettävyyssongelma voi oikeaan kontekstiin sijoitettuna osoittautua harmittomaksi kosmeettiseksi haitaksi.

Käytettävyyssongelmien vakavuusluokitus voidaan tehdä arviointitapahtuman aikana tai myöhemmin ongelmankuvausten perusteella. Vakavuutta voidaan mitata sen mukaan, miten usein ongelma esiintyy, miten paljon siitä on haittaa käyttäjän suoriin tai miten paljon



Kuva 1: Käytettävyyssongelman osatekijät (Lavery et al., 1997).

aikaa käyttäjältä kuluu siitä selviämiseen. (Hassenzahl, 2000; Nielsen, 1993). Varsinkin analyyttisten menetelmien yhteydessä vakavuusluokitus perustuu mittaamisen sijasta arvioijan päättelyyn.

Käytettävyysohjelman vakavuusaste voidaan ilmaista esimerkiksi Nielsenin (1993) luokiteluasteikolla, jossa 0 ei ole varsinainen käytettävyysohmella ollenkaan ja 4 on käytettävyysohmakatastrofi, joka pitää korjata välittömästi. Edellä kuvatussa Jacobsenin ja muiden (1998) tutkimuksessa käytettävyysohmella määriteltiin vakavaksi, jos se täytti vähintään kolme käytettävyysohmalle asetettua kriteeriä. Käytettävyysohmien vakavuusluokituksesta on kerrottu enemmän luvussa 13 (Koskinen).

19.2.3. Arviointimenetelmien laatu

Käytettävyysohmien arviointimenetelmien laadun kriteereinä käytetään reliabiliteetin ja validiteetin lisäksi mm. vaikuttavuutta (effectiveness) ja kattavuutta (thoroughness). Kattavuus tarkoittaa sitä, että menetelmällä löydetään mahdollisimman monta sovelluksessa olevaa ohmella. Vaikuttavuus on kattavuuden ja validiteetin yhdistelmä. Koska arvioijan vaikutus koskee pääasiassa reliabiliteettia ja validiteettia, tässä työssä keskitytään niihin. Arviointimenetelmien laadun kriteereihin on perehdytty tarkemmin luvussa 20 (Kosonen).

Reliabiliteetti kuvaa tutkimuksen luotettavuutta eli miten hyvin ohmella antaa samanlaisia tuloksia käyttökerrasta riippumatta. Käytettävyysohmien se merkitsee arviointia tekevien henkilöiden yhdenmukaista ohmien löytämistä. Validiteetilla taas kuvataan tutkimuksen pätevyttä. Se tarkoittaa, että tutkimuksen havainnot ovat päteviä myös reaali-maailmassa eli raportoidut ohmella ovat käyttäjänkin kannalta ohmella ja siten mielekkäitä korjata. Luvussa 2 (Vanhala) on kuvattu reliabiliteettia ja validiteettia yksityiskohtaisemmin.

Käytettävyysohmien validiteettia heikentäviä tekijöitä ovat *väärät hälytykset (false alarms)* sekä havaitsematta jääneet todelliset käytettävyysohmella. Jos todellinen ohmella jää havaitsematta, syynä on joko se, ettei sitä ole huomattu lainkaan tai sitten sitä ei ole pidetty käytettävyysohmella. Väärällä hälytyksellä tarkoitetaan ohmella, joka on analyyttisellä menetelmällä todettu käytettävyysohmella, mutta esimerkiksi empiirisillä testeillä ei osoittaudukaan käytettävyysohmella. Väärät hälytykset ja havaitsematta jääneet todelliset ohmella ovat nimenomaan analyyttisten menetelmien heikkous. Ohmella voi olla käytettävyysohmella vain jossakin tietyssä kontekstissa tai käyttäjät ovat oppineet kiertämään ohmella ja pitävät sitä ohjelmaan kuuluvana ominaisuutena. (Cockton & Woolrych, 2002; Hertzum & Jacobsen, 2001).

19.3. Arvioijan vaikutuksen tutkiminen

Arvioijan vaikutuksen tutkiminen on keskittynyt heuristisella arvioinnilla, kognitiivisella läpikäynnillä ja käytettävyysohmella saatuihin arvioinnin tuloksiin. Tässä kohdassa tarkastellaan näistä tehtyjä tutkimuksia.

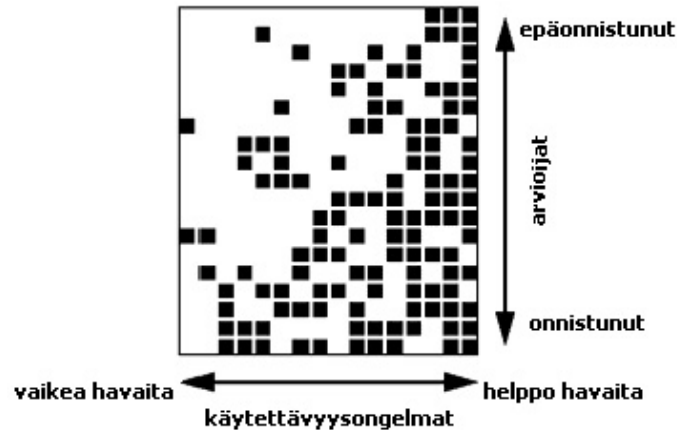
19.3.1. Heuristisessa arvioinnissa

Nielsenin (1992) tutkimuksessa arvioinnin kohteena oli puhekäyttöinen pankkijärjestelmä ja arviointimenetelmänä heuristinen arviointi. Tutkimuksen eräänä tavoitteena oli selvittää, minkälainen vaikutus arvioijan käytettävyysohmella tasolla on arvioinnin tuloksiin. Arvioijina toimi kolmenlaisia henkilöitä, aloittelijat, käytettävyysohmella sekä ns. *kahden alan asiantuntijat (double expert)*. Kahden alan asiantuntijoilla tarkoitetaan sitä, että arvioija on sekä käytettävyysohmella että arvioitavan sovelluksen kohdealueen asiantuntija. Aloittelijoita oli 31 henkeä, käytettävyysohmella 19 ja kahden alan asiantuntijoita 14 henkeä.

Tulokset osoittavat, että eniten ohmella havaitsivat kahden alan asiantuntijat (60 %), toiseksi eniten käytettävyysohmella (41 %) ja vähiten aloittelijat (22 %). Prosenttiluku on keskiarvo yhden arvioijan havaintojen lukumäärän suhteesta arvioijien tuottamien ohmella.

kuvausten kokonaismäärään. Lukua käytetään arvioijan vaikutuksen kuvaamiseen. Hertzum ja Jacobsen (2001) käyttävät siitä nimitystä *havaitsemisaste* (*detection rate*). Nielsen (1992) puhuu vain kaikkien ongelmien keskiarvosta. Mitä lähempänä 100 %:ia luku on, sitä vähäisempi on arvioijan vaikutus. Toinen Hertzumin ja Jacobsenin käyttämä arvioijan vaikutusta kuvaava luku on kahden arvioijan välisen *yksimielisyyden aste* (*any-two agreement*). Siinä vertaillaan aina kahden arvioijan välisiä havaintoja keskenään.

Kuvan 2 matriisissa on havainnollistettu Nielsenin (1992) tutkimuksen käytettävyysasiantuntijoiden tekemät havainnot. Matriisin sarakkeet kuvaavat käytettävyysongelmia ja rivit arvioijia. Musta neliö vastaa yhden henkilön löytämää yhtä ongelmaa. Tässä matriisissa vaikeimmin havaitut ongelmat on ryhmitelty vasempaan reunaan ja vähiten ongelmia havainneet arvioijat yläosaan. Vähiten ongelmia havainnut arvioija löysi kolme ongelmaa ja eniten havainnut kymmenen. Kukaan ei löytänyt kaikkia ongelmia. Merkillepantavaa on, että eniten ongelmia löytäneet arvioijat eivät kuitenkaan havainneet kahta vaikeimmin havaittavissa ollutta ongelmaa.



Kuva 2: Käytettävyysasiantuntijoiden heuristisessa arvioinnissa havaitsemat ongelmat (Nielsen, 1994).

Koska Nielsenin tutkimuksessa käytettiin vertailulukuna havaitsemisastetta, voidaan tuloksia tarkastella myös arvioijan vaikutuksen näkökulmasta. Arvioijan vaikutus näyttäisi olevan sitä matalampi mitä korkeampi asiantuntijuus arvioijalla on. Kahden alan asiantuntijoiden havaitsemisaste oli 60 %, kun aloittelijoiden vastaava luku oli 22 %. Kahden alan asiantuntijoiden ryhmä löysi myös eniten käytettävyysongelmia, kun taas aloittelijoiden ryhmä löysi vähiten. Ryhmien välisiä eroja tarkasteltaessa voidaan siis havaita, että löydettyjen ongelmien lukumäärällä on merkitystä arvioijan vaikutukseen. Mitä enemmän ongelmia ryhmä löysi, sitä matalampi oli arvioijan vaikutus. Vaikutusta on tietysti siltäkin, että kahden alan asiantuntijat tunsivat arvioitavan kohteen muita paremmin.

19.3.2. Käytettävyystestissä

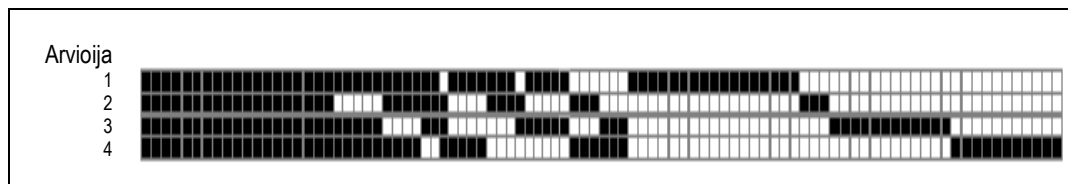
Jacobsen ja muut (1998) tutkivat arvioijan vaikutusta käytettävyystestauksen yhteydessä. Arvioinnin kohteena oli multimediaa sisältävä tekstinkäsittelyohjelma. Neljä käytettävyysasiantuntijaa katsoivat kukin neljä videolle nauhoitettua käytettävyystestiä ja raportoivat löytämänsä ongelmat annettujen ohjeiden mukaan. Raportissa tuli esittää todiste ongelman olemassaolosta joko käyttäjän toimintosarjan ja/tai suullisen ilmauksen perusteella, vapaa-muotoinen ongelman kuvaus sekä ongelman yksilöivä kriteeri. Käytettävänä oli kahdeksan etukäteen määritettyä kriteeriä sekä lisäksi mahdollisuus muodostaa jo havaittujen ongelmien ryhmistä uusi yleisempi ongelma.

Ongelmia havaittiin yhteensä 93, joista vain 20 % oli sellaisia, jotka kaikki neljä havaitsivat (taulukko 1). Arvioijan vaikutus ilmeni myös vakavuusluokituksessa.

Taulukko 1: Ongelman raportointien lukumäärät (Jacobsen et al., 1998).

Ongelman raportointien arvioijien lukumäärä	1	2	3	4
Vakavat ongelmat	8 (22 %)	7 (19 %)	7 (19 %)	15 (41 %)
Kaikki ongelmat	43 (46 %)	19 (20 %)	12 (13 %)	19 (20 %)

Arvioinnin tulosta on havainnollistettu kuvan 3 matriisissa, jossa rivit kuvaavat arvioinnin tehneitä henkilöitä ja sarakkeet havaittuja ongelmia. Matriisin vasempaan reunaan on ryhmitelty ne 19 ongelmaa, jotka kaikki neljä käytettävyyssiantijaa löysivät. Havaitsemisaste oli 52 %. Vakavien käytettävyysongelmien arvioinnin havaitsemisaste oli 72 %.



Kuva 3: Neljän käytettävyyssiantuntijan testinauhoilta löydämät ongelmat (Jacobsen et al., 1998).

19.3.3. Kognitiivisessa läpikäynnissä

Hertzumin ja Jacobsenin (1999) tutkimuksessa arvioinnin kohteena oli Internetissä toimiva kirjasto ja arviointimenetelmänä kognitiivinen läpikäynti. Tutkimuksen tavoitteena oli selvittää, miten paljon aloittelijat (opiskelijat, joilla kahden tunnin perehdytys menetelmään) löytävät samoja ongelmia arvioitavasta käyttöliittymästä kognitiivisella läpikäynnillä, kun kaikki käyttävät samoja tehtäviä.

Opiskelijat löysivät yhteensä 33 käytettävyysongelmaa. Kuvassa 4 on esitetty havaitut ongelmat arvioijittain. Vähiten ongelmia havainneet arvioijat on ryhmitelty matriisin yläosaan. Matriisista ilmenee, että vähiten ongelmia havainnut opiskelija löysi vain kaksi ongelmaa eikä kukaan löytänyt kaikkia.

Tutkimuksessa havaitsemisaste oli 18 %, mikä on merkki korkeasta arvioijan vaikutuksesta. Syitä siihen voidaan etsiä arviointia tekevien henkilöiden kokemattomuudesta tai vähäisestä asiantuntemuksesta. Kenelläkään arvioijista ei ollut aiempaa kokemusta menetelmän käytöstä



Kuva 4: Opiskelijoiden kognitiivisella läpikäynnillä löydämät ongelmat (Hertzum & Jacobsen, 1999).

parhaimmillaan 73 %. Mitä lähempänä havaitsemisaste on 100 %:ia, sitä vähäisempi on arvioijan vaikutus.

Hertzumin ja Jacobsenin (1999) mukaan suuri syy tutkimuksessa ilmenneeseen arvioijan vaikutukseen on se, että arvioijat mallinsivat oletetun käyttäjän toiminnot liiaksi omien kokemustensa mukaisiksi. Tällaisesta omiin kokemuksiin kiinnittymisestä käytetään nimitystä *ankkurointi* (*anchoring*). Ankkurointi vahvistuu, jos oletettu käyttäjä kuvataan epätarkasti ja liikaa yleistäen.

19.3.4. Havainnot kootusti – tulosten vertailua

Yksittäisen tutkimuksen sisällä havaintojen vertailu on suhteellisen helppoa – onhan kaikille arvioijille annettu samanlaiset ohjeistukset menetelmän käyttöön. Silti tutkijan tehtäväksi jää esimerkiksi miettiä, tarkoittavatko eri tavoin kuvatut ongelmat itse asiassa eri ongelmia vai samaa ongelmaa, joka vain on määritelty muista poikkeavasti.

Hertzum ja Jacobsen (2001) ovat vertailleet yhdentoista eri käytettävyyssarvioinnin tuloksia, joihin myös edellä esitetyt sisältyvät. Käytetyt menetelmät olivat käytettävyyssarvioinnin, kogni-

tiivinen läpikäynti sekä heuristinen arviointi. Osa arvioinneista tehtiin ryhmätyönä. Arvioijien asiantuntemus vaihteli kahden alan asiantuntijasta aloittelijaan. Kaikista yhdestätoista arvioinnista on laskettu arvioijan vaikutusta kuvaava havaitsemisaste, mikä vaihteli välillä 18–73 %. On kuitenkin huomattava, että arvioinnit eivät ole suoraan vertailukelpoisia, koska esimerkiksi arvioitava kohde on jokaisessa ollut eri.

Arvioijan vaikutuksen tutkimus on painottunut käytettävyydestäukseen, kognitiiviseen läpikäyntiin ja heuristiseen arviointiin. Ainakin näiden menetelmien välillä arvioijan vaikutus esiintyy melko tasaisesti kaikissa. Eniten arvioijan vaikutuksen esiintyminen hämmästyttää empiirisen käytettävyydestäuksen yhteydessä, jota on perinteisesti pidetty hyvin luotettavana menetelmänä. Arvioijan vaikutusta ei tietävästi ole tutkittu muitten kuin mainittujen menetelmien yhteydessä. Se ei kuitenkaan merkitse sitä, etteikö arvioijan vaikutusta niissäkin olisi. Haastattelussa tai fokusryhmissä arviointia tekevän henkilön eli moderaattorin (haastattelijan) rooli on hyvin keskeinen. Aihetta olisi selvittää, miten paljon esimerkiksi moderaattorin ohjailu tai kehon kielen viestit vaikuttavat tulokseen. Haastattelusta arviointimenetelmänä on kerrottu luvussa 3 (Vuorela) ja fokusryhmistä luvussa 4 (Parviainen).

19.4. Arviointien eroihin johtavia tekijöitä

Nielsenin (1993) mukaan syy arviointia tekevien henkilöiden välisiin eroihin ongelmien löytämisessä on siinä, että eri ihmisillä on erilainen käsitys siitä, mitä käytettävyydellä tarkoitetaan. Hertzumin ja Jacobsenin (2001) mukaan arvioijan vaikutuksen pääasiallinen syy on siinä, että käytettävyyden arviointi on *kognitiivista toimintaa* (*cognitive activity*). Kognitiivinen toiminta liittyy ihmisen tiedonkäsittelyyn kuten havaitsemiseen, ajatteluun, ongelmanratkaisuun ja muistiin. Koska käytettävyyсарviointi perustuu suurelta osin käytettävyyssasiantuntijan tekemiin havaintoihin ja päätelmiin, korostuvat yksilöiden väliset erot.

Myös arviointia tekevän henkilön asiantuntijuuden tasolla on merkitystä arvioijan vaikutukseen. Kuten kohdassa 19.3.1 esitellyssä tutkimuksessa (Nielsen, 1992) todettiin, arvioijan vaikutus oli vähäisempi kahden alan asiantuntijoiden kohdalla kuin ”tavallisten” käytettävyyssasiantuntijoiden kohdalla ja huomattavasti matalampi kuin aloittelijoiden kohdalla. Kokeneet käytettävyyssasiantuntijat löytävät enemmän ongelmia kuin esimerkiksi arviointia tekevät opiskelijat, jolloin myös arvioijan vaikutus on matalampi.

Hertzum ja Jacobsen (2001) ovat tutkineet *arvioinnin ohjeistuksen* (*guidelines*) vaikutusta arviointitapahtumaan ja nostavat esille kolme tärkeintä arvioijan vaikutusta lisäävää ohjeistuksen puutetta. Nämä ovat

- puutteellinen *tavoitteiden analysointi* (*goal analysis*), jonka seurauksena arvioijat laativat huomattavasti toisistaan poikkeavia tehtäväskenaarioita.
- puutteellisesti kuvattu arviointiprosessi, jonka seurauksena arvioijat mallintavat käyttäjän omien kokemustensa perusteella itsensä kaltaiseksi ja
- puutteellinen ongelman kriteerien määrittely, jonka seurauksena valituksi tulee ongelmia, jotka eivät ole käytettävyysongelmia lainkaan.

Cockton ja Woolrych (2002) pitävät arviointimenetelmien puutteita suurimpana syynä arvioijan vaikutukseen sekä siihen, miksi havaituissa ongelmissa on niin paljon vääriä hälytyksiä ja todellisia ongelmia jää havaitsematta. He kritisoiivat varsinkin ns. kevennettyjä menetelmiä. Kevennettyjen menetelmien kuten esimerkiksi heuristisen arvioinnin heikkoutena on niiden pintapuolisuus, johon pyrkimys kustannussäästöihin on johtanut.

Huomattava kuitenkin on, että vaikka käytettävyyssasiantuntijat yksilöinä havaitsevat niin oleellisesti erilaisia käytettävyysongelmia, he ryhmätyöskentelyssä saavuttavat nopeasti keskinäisen yksimielisyyden. Ryhmätyö tarjoaa lisäksi käytettävyyssasiantuntijoille tilaisuuden kehittää arviointitaitojaan tutkimalla omien ja kollegojen havaintojen välisiä eroja. (Hertzum & al., 2002)

19.5. Arvioijan vaikutuksen hallinta

Arvioijan vaikutuksen hallintaan ei liene olemassa yleispätevää ratkaisua. Jos tavoitteena olisi vain tutkimuksen reliabiliteetin parantaminen, arviointia tekevien henkilöiden määrän kasvattaminen voisi olla ratkaisu. Sillä on kuitenkin kääntöpuolensa, sillä mitä enemmän arvioijia lisätään, sitä enemmän lisääntyvät väärät hälytykset. Toisin sanoen tutkimuksen validiteetti kärsii. (Cockton & Woolrych, 2002; Cockton et al., 2003; Sears, 1997).

Jos arvioijan vaikutuksen hallintaa tarkastellaan käytettävyyssiantuntijan näkökulmasta, tärkeäksi nousee arviointiin sisältyvän inhimillisen tekijän osuuden vähentäminen. Keinoja voivat olla erilaiset arvioinnin apuvälineet kuten standardoitu raportointi, mutta myös käytettävyyssiantuntijan taitojen kehittäminen.

Myös arviointimenetelmiä voidaan kehittää. Kehittäminen voi olla menetelmän ominaisuuksien parantamista tai menetelmien hyväksi havaittujen osien yhdistämistä uudeksi menetelmäksi.

19.5.1. Arvioinnin ohjeistus

Hertzum ja Jacobsen (2001) korostavat ohjeistuksen merkitystä arvioijan vaikutuksen hallinnassa. Varsinkin on tärkeää, että tehdään riittävän tarkka tavoitteiden analysointi ennen arviointia ja että sekä arviointiprosessista että käytettävyysongelman kriteereistä on olemassa riittävän yksityiskohtaiset määrittelyt.

Tavoitteiden tarkalla analysoinnilla pyritään takaamaan, että käytettävyyssiantuntijoiden laatimat tehtäväskenaariot olisivat yhdenmukaisia ja riittävän kattavia, jotta kaikki oletettujen käyttäjien kannalta relevantit sovelluksen toiminnot ja ominaisuudet tulisi arvioiduksi. Aina ei ole tarkoituksenmukaista eikä edes mahdollista, että arviointiin otetaan mukaan kaikki sovelluksen ominaisuudet. Jos analysointi on puutteellista, käytettävyyssiantuntijat joutuvat itse päättämään, mitkä käyttäjän ja arvioinnin tavoitteet ovat.

Toinen tärkeä parannuskohde on arviointiprosessin kuvaus. Sillä on merkitystä varsinkin heuristisessa arvioinnissa, jossa ei ole tarkasti määriteltyä etenemisprosessia, mutta myös kognitiivisessa läpikäynnissä. Jos arviointiprosessin kuvaus on puutteellinen, käytettävyyssiantuntija joutuu turvautumaan esimerkiksi käyttäjän mallintamisessa omiin kokemuksiinsa. Tällainen omiin kokemuksiin kiinnittyminen eli ankkurointi johtaa siihen, että oletetusta käyttäjästä tulee liiaksi arviointia tekevän henkilön kaltainen, mikä taas vaikuttaa arvioinnin kulkuun.

Kolmas ohjeistuksen kannalta tärkeä kohde on ongelmien kriteerien määrittely. Arviointia tekevän henkilön tulisi tarkasti tietää, mitkä kriteerit ongelman on täytettävä, että se voidaan luokitella käytettävyysongelmaksi. Puutteellinen ohjeistus lisää käytettävyyssiantuntijan päätösvastuuta – siitä seuraa väärä hälytyksiä ja havaitsematta jääneitä käytettävyysongelmia.

19.5.2. Menetelmien kehittäminen

Searsin (1997) ratkaisuehdotus arviointimenetelmien laadun parantamiseen on niiden edelleen kehittäminen. Hän esittelee heuristisesta arvioinnista ja kognitiivisesta läpikäynnistä kehitetyn *heuristisen läpikäynnin* (*heuristic walkthrough*). Heuristinen läpikäynti on vapaamuotoisempi kuin kognitiivinen läpikäynti, mutta rakenteellisempi kuin heuristinen arviointi. Heuristiikkalistojen lisäksi menetelmässä käytetään kognitiivisesta läpikäynnistä tuttuja neljää kysymystä sekä arvioitavia tehtäviä, jotka on määritelty oletetun käyttäjäryhmän perusteella.

Sears on verrannut heuristisella arvioinnilla, kognitiivisella läpikäynnillä ja heuristisella läpikäynnillä tehtyjen arviointien tuloksia. Vertailumittareina hän käytti reliabiliteettia, validiteettia ja kattavuutta. Heuristisen läpikäynnin arviointituloksen reliabiliteetti ja kattavuus olivat korkeampia kuin kognitiivisessa läpikäynnissä ja validiteetti parempi kuin heuristisessa arvioinnissa.

Merkillepantavaa on, että eri menetelmien välillä oleva reliabiliteettitaso ero väheni, kun arvioijien määrää lisättiin. Arvioijien määrän lisääminen heikensi kaikkien menetelmien ja varsinkin heuristisen arvioinnin validiteettia.

19.5.3. Käytettävyysohjelmien priorisointi

Koska arvioijan vaikutus ilmenee myös vakavuusluokituksessa, voidaan arviointimenetelmien laatua parantaa myös kehittämällä käytettävyysohjelmien priorisointia. Vakavuusasteluokitus on nojautunut enemmän käytettävyyssiantuntijan kokemukseen kuin johonkin menetelmään tai teoreettiseen viitekehukseen.

Nielsenin (1993) mukaan arvioijien tekemä vakavuusluokitus on niin subjektiivista toimintaa, ettei koskaan saisi luottaa yhden asiantuntijan tekemään määrittelyyn. Luotettavampaa on kerätä vakavuusluokitus usealta, vähintään kolmelta neljältä arvioijalta ja käyttää niiden keskiarvoa.

Hassenzahl (2000) esittelee kaksi erilaista lähestymistapaa vakavuusluokitukseen. Nämä ovat *aineistoon perustuva (data-driven)* ja arviointia tekevän henkilön *päätelyyn perustuva (judgement-driven)* vakavuusluokitus. Nämä lähestymistavat noudattavat Nielsenin (1993) jaottelua. Sen mukaan käytettävyysohjelmien priorisointi voidaan tehdä joko keräämällä dataa empiirisissä testeissä tai käytettävyyssiantuntijoiden päätelyyn avulla.

Aineistoon perustuvassa vakavuusluokituksessa Hassenzahl käyttää kriteerinä *yhteenlaskettua ongelmankäsittelyaikaa (total problem handling time)*, joka sisältää sekä ongelman käsittelyyn kuluvaan aikaan että ongelman esiintymistiheyden. Sekä ongelman käsittelyyn kuluva aika että esiintymistiheys ovat muuttujia, jotka lasketaan empiiristen testien tuloksista. Yhteenlasketulla ongelmankäsittelyajalla ei kuitenkaan osoiteta vain taloudellisia kustannuksia vaan myös psykologiset kustannukset. Psykologisilla kustannuksilla tarkoitetaan käytettävyysohjelman käyttäjälle aiheuttamaa stressiä.

Aineistoon perustuvaa priorisointia pidetään laadultaan korkeatasoisena, mutta kalliina. Empiiriset testit ja nauhoitettujen ongelmien analysointi vievät aikaa. Jos käytettävyyssarviointi on tehty jollakin analyttisellä menetelmällä kuten heuristisella arvioinnilla, ei empiiristä aineistoa vakavuusarviointiin ole saatavilla.

Vaihtoehtona aineistoon perustuvalla priorisoinnilla Hassenzahl esittelee päätelyyn perustuvan mallin, joka toteutetaan *intressiryhmän (interest group)* avulla. Intressiryhmän jäsen voi olla esimerkiksi käyttäjä, sovelluskehittäjä tai ohjelmistoprojektin päällikkö. Hassenzahl perustelee intressiryhmän valintaa sillä, että intressiryhmän jäsenellä on usein esimerkiksi oman työnsä kautta parempi tuntuma siihen, miten paljon ongelmasta tulisi olemaan haittaa oikeassa työtilanteessa kuin käytettävyyssiantuntijalla.

Päätelyyn perustuvassa mallissa vakavuusluokitusta varten intressiryhmän jäsenille annetaan valmiina asiantuntijoiden laatimat käytettävyysohjelmakuvaukset sekä vakavuuden arvioinnissa käytettäviä kriteereitä, joita käytetään perusteluina vakavuusluokitukseen. Kriteerejä on kahdenlaisia, funktionaalisia ja dysfunktionaalisia. Funktionaaliset kriteerit ovat käyttäjälle aiheuttavia konkreettisia seurauksia käytettävyysohjelmasta, kun taas dysfunktionaaliset kriteerit ovat arviointia tekevää henkilöä koskevia. Funktionaalisia kriteerejä ovat ongelman esiintymistiheys, ongelman käsittelyyn kuluva aika sekä psykologiset kustannukset.

Dysfunktionaalisia kriteerejä ovat

- vaikeus kuvitella, mitä seurauksia ongelmasta aiheutuu,
- vaikeus löytää syytä, mikä ongelman aiheuttaa,
- vaikeus arvioida ongelman vakavuutta muuten kuin henkilökohtaisesta näkökulmasta ja
- vaikeus ymmärtää ongelmakuvauksen perusteella, mistä siinä tarkasti ottaen on kysymys.

Dysfunktionaaliset kriteerit ovat eräänlaisia häiriötekijöitä, jotka vaikuttavat vakavuusarviointiin. Dysfunktionaalisilla kriteereillä pyritään kartoittamaan arvioijien välistä vaihtelua. Vakavuusarviointi voidaan tehdä yläkanttiin, jos arviointia tekevä henkilö kokee, että ongelmasta

olisi erityisen paljon haittaa nimenomaan hänen työlleen. Tällöin tarkastelukulma on helposti liian henkilökohtainen eikä esimerkiksi kaikkia samaan ryhmään kuuluvia (esim. käyttäjiä) koskeva. Dysfunktionaalisten kriteerien mukaanotto vakavuuden arvioinnissa auttaa myös huomaamaan, mitkä ongelmakuvaukset pitäisi ehkä muotoilla toisin, kun niitä ei ymmärretä, ja mitkä kuvaukset kaipaisivat tuekseen käytännön testitilanteista kerättyjä havaintoja.

Käytännössä päättelyyn perustuva vakavuusarviointi Hassenzahlin tutkimuksessa toteutettiin siten, että intressiryhmän jäsenille jaettiin vihkonen, jossa oli yksi ongelmakuvaus yhdellä sivulla. Ongelmakuvauksen lisäksi sivulla oli kriteerit, joilla ongelman vakavuus arvioitiin. Esimerkiksi psykologiset kustannukset –kriteeri esitettiin väittämänä ”tällainen ongelma saisi minut hermostumaan” ja vastausta varten annettiin Likertin viisiportainen asteikko, jossa vastausvaihtoehdot olivat väliltä ”täysin eri mieltä” ja ”täysin samaa mieltä”. Myös dysfunktionaaliset kriteerit esitettiin väittäminä jokaisen ongelmakuvauksen yhteydessä.

Hassenzahl vertaa intressiryhmän antamia vastauksia tilastollisesti keskenään käyttäen pohjana ongelmien vakavuudesta käytettävyytysteissä tehtyjä aitoja, aineistopohjaisia havaintoja. Mainittakoon, että Hassenzahlin tutkimuksessa parhaiten ongelman vakavuutta kuvasivat psykologiset kustannukset, ongelman käsittelyyn kuluva aika sekä kyvyttömyys kuvitella, mitä seurauksia ongelmasta aiheutuu. Vain näissä kriteereissä vastausten tilastollinen analyysi osoitti tilastollista merkitsevyyttä.

Vaikka Hassenzahlin (2000) tutkimuksen pääasiallinen tavoite oli vertailla, miten aineistoon perustuva ja päättelyyn perustuva vakavuusarviointi vastaavat toisiaan, se on myös oivallinen esimerkki siitä, miten asiantuntijuuteen perustuvaa vakavuusarviointia voidaan kehittää.

19.5.4. Standardoitu raportointi

Laveryn ja muiden (1997) kehittämän käytettävyysongelmiin raportointimallin (kuva 5) rakenne perustuu käytettävyysongelman osatekijöihin (kohta 19.2.2). Käytettävyysongelman

- | | |
|---|--|
| <ol style="list-style-type: none">1. Konteksti<ol style="list-style-type: none">a) Kuvaile konteksti, jossa ongelma ilmeni.2. Ongelman aiheuttaja<ol style="list-style-type: none">a) Mikä ongelman aiheutti?b) Ongelman aiheuttajan tyyppi (esim. tiedon puute, suunnitteluvirhe)3. Katkos osallistujan vuorovaikutuksessa<ol style="list-style-type: none">a) Tapahtuiko vuorovaikutustilanteessa katkos (kyllä/ei)? Jos vastauksesi on ”ei”, siirry kysymykseen 5.b) Minkälaisia ikävyyksiä katkoksesta aiheutui?4. Katkoksen aiheuttamat seuraukset<ol style="list-style-type: none">a) Miten osallistuja katkoksen jälkeen käyttäytyi?b) Minkälainen vaikutus katkoksesta oli osallistujan suorituksen tai työhön? Siirry kohtaan 6.5. Seuraus (lopputulos)<ol style="list-style-type: none">a) Muuttiko ongelma osallistujan käyttäytymistä?b) Minkälainen vaikutus ongelmalla oli osallistujan suorituksen tai työhön?6. Ratkaisuehdotus<ol style="list-style-type: none">a) Suosituksesi ongelman ratkaisemiseksi. | osatekijät (ongelman aiheuttaja, ongelmasta johtuvat katkokset, seuraukset ja konteksti) on otettu mukaan ongelman kuvaukseen. |
|---|--|

Raportointimalli soveltuu käytettäväksi sekä empiiristen että analyttisen arviointimenetelmien yhteydessä. Analyttisissä menetelmissä kohtien 3, 4 ja 5 kysymykset muutetaan konditionaali-muotoon, koska niissä ei ole testikäyttäjää mukana.

Standardoidun raportoinnin ansiosta käytettävyyssiantuntijat analysoivat ongelmia entistä syvällisemmin, sillä heidän on otettava kantaa sekä ongelman syihin että seurauksiin. Lisäksi analy-

Kuva 5: Käytettävyysongelman raportointimalli (Lavery et al., 1997).

sointi on yhdenmukaisempaa määrämuotoisen rakenteen takia. Ongelmien analysoinnissa otetaan huomioon myös konteksti, missä ongelma esiintyy.

19.5.5. DARE-malliin perustuva raportointi

Myös Cockton ja muut (2003) painottavat raportoinnin merkitystä käytettävyyden arvioinnin laadun parantamisessa. He viittaavat edellä kuvattuun Laveryn ja muiden (1997) kehittämään standardoituun raportointimalliin yhtenä tärkeimmistä välineistä arvioinnin kehittämistyössä.

Cocktonin ja muiden lähestymistapa käytettävyyden laadun parantamiseen on käytettävyydsasiantuntijan eikä arviointimenetelmän näkökulma. Heidän kehittämänsä raportointi perustuu *DARE (The Discover and Analysis Resource)* –malliin, jossa käytettävyydsarviointi kuvataan käytettävyydsasiantuntijan tietovarastoina. Mallinnus koskee arvioinnin kahta keskeisimmistä vaihetta eli tiedon keräämistä ja ongelmien analysointia. Vastaavat tietovarastot liittyvät havaitsemiseen ja analysointiin.

DARE-malliin perustuvan raportoinnin rakenne sisältää sekä ongelmien havaitsemisen että niiden analysoinnin. Raportti jakaantuu neljään osaan eli ongelmien kuvaukseen, kuvaukseen siitä, miten ongelma havaittiin, ongelmien perusteluun jollakin heuristiikalla (heuristinen arviointi) sekä ongelmien seulontaan.

Ensimmäisessä osassa käytettävyydsasiantuntija kuvailee havaitut ongelmat sekä niiden käyttäjälle aiheuttamat vaikeudet. Ongelman kuvaus sisältää seuraavat kohdat:

- lyhyt ongelman kuvaus omin sanoin esitettyinä,
- ongelman käyttäjälle aiheuttamien todennäköisten ja todellisten vaikeuksien luettelo,
- kontekstin selvittäminen eli missä ympäristö(i)ssä ongelma todennäköisimmin ilmenee sekä
- ongelman syiden kuvailu.

Toisessa osassa kuvataan, miten ongelma havaittiin. Eri tavat helpoimmasta vaikeimpaan ovat

- järjestelmän silmäily ilman erityistä strategiaa,
- järjestelmän tutkiminen jotakin strategiaa käyttämällä (esimerkiksi linkkien, työkalupalkkien ja valikkojen tarkistaminen),
- tavoitteellinen järjestelmän läpikäynti (esimerkiksi etsitään tiettyä käyttäjän tavoitteeseen sisältyvää asiaa) sekä
- tavoitteellinen järjestelmän läpikäynti jotain tunnettua metodologiaa käyttämällä.

Kolmas osa koskee erityisesti heuristista arviointia ja siinä käytettävyydsasiantuntija perustelee, miksi käytetty heuristiikka on relevantti. Neljännessä raportin osassa ongelmien joukosta seulotaan pois epätodennäköiset. Jokainen ongelman poistaminen perustellaan.

Cocktonin ja muiden tutkimuksissa pystyttiin osoittamaan, että DARE-malliin perustuvan raportoinnin käyttäminen paransi arvioinnin laatua merkittävästi. Varsinkin väärin hälytysten määrä väheni, mikä parantaa validiteettia.

19.5.6. Käyttäjän toimintoihin perustuva teoreettinen viitekehys (UAF)

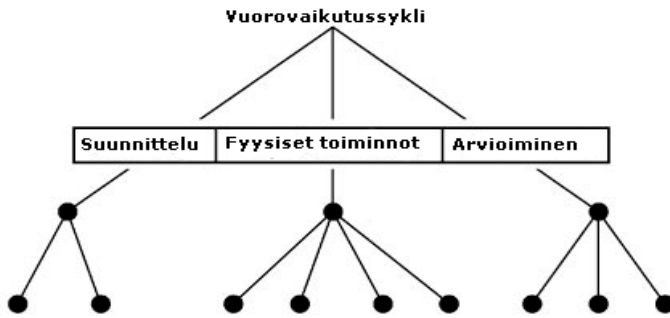
Hartsonin ja muiden (1999) kehittämän *käyttäjän toimintoihin perustuvan viitekehksen (The User Action Framework, UAF)* tarkoituksena on koota yhteen kaikki käytettävyyds tutkimukseen liittyvä toiminta eli ongelmien etsiminen käytettävyydsarvioinnissa, ongelmien analysointi ja luokittelu, raportointi, ongelmien korjaaminen sekä uudelleensuunnittelu. Lisäksi se toimii käsitteiden ja ohjeistusten tietovarastona. UAF on perustana ohjelmistototeutukselle¹, jonka avulla ongelmakuvauksia hallitaan iteratiivisen kehitystyön aikana (Andre et al., 2001).

Viitekehksen taustalla on vuorovaikutusprosessi, joka perustuu Normanin *seitsenvaiheiseen malliin (theory of action)* ihmisen ja koneen vuorovaikutuksesta (kuva 1 raportin luvussa 9). Vuorovaikutusprosessi kuvaa käyttäjän työvaihetta ohjelmistoa käyttäessään. Esimerkiksi käyttäjä haluaa liittää sähköpostiviestiinsä liitetiedoston – miten se tapahtuu ja mitä käyttäjän pitää tehdä? Suunnitteluvaiheessa käyttäjän on selvitettävä itselleen, mitä pitää tehdä, että sovelluksen käytön voi aloittaa ja miten se fyysisesti tapahtuu. Fyysiset toiminnot tarkoittavat

¹ <http://hemlock.cs.vt.edu/publicuafviewer/pages/default.asp> (24.5.2004)

suunnitelmien täytäntöönpanoa konkreettisina toimenpiteinä käyttöliittymässä. Käytettävyyden kannalta on keskeistä, kykeneekö käyttäjä tekemään haluamansa toiminnot ja miten helposti. Arviointivaiheessa käyttäjä tarkastelee saamaansa palautetta (jos sitä annettiin), ja yrittää ymmärtää sen merkityksen suhteessa toiminnalle asetettuun tavoitteeseen. (Andre et al., 2001)

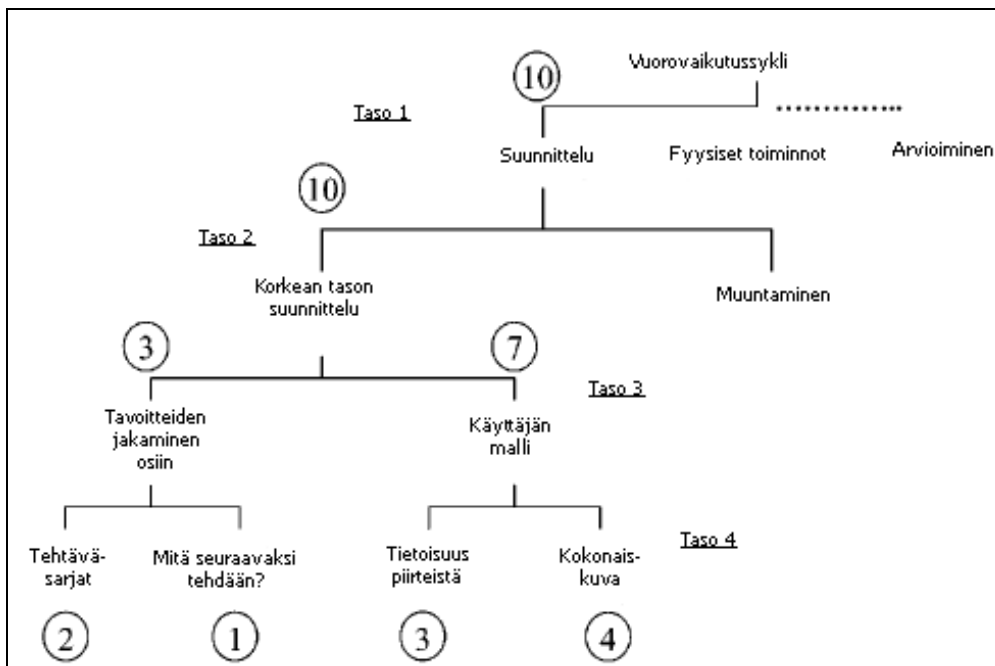
UAF:ssa käytettävyyso ongelmia tarkastellaan siitä näkökulmasta, miten ne vaikuttavat käyttäjän fyysisiin ja kognitiivisiin toimintoihin tehtävän suorituksen aikana Ongelmien analysointi ei rajoitu vain niiden kuvauksiin, vaan ongelmat luokitellaan ja sijoitetaan kontekstiin. Käytettävyyssiantuntija saa käyttöönsä analysoinnin ja luokittelun tueksi tarvittavan termistön UAF-järjestelmää käyttäessään. Termistö kattaa sekä käytettävyyteen että vuorovaikutukseen liittyvät käsitteet.



Kuva 6: The User Action Framework (Andre et al., 2001).

Kun käytettävyyssiantuntija ryhtyy analysoimaan ongelmaa, hänen on ensin ratkaistava, mihin käyttäjän toimintojen vaiheeseen (suunnittelu, fyysiset toiminnot, arvioiminen, kuva 6) ongelma vaikuttaa. Sen jälkeen hän etenee hierarkkista polkua (kuva 7) ja jokaisella tasolla valitsee ongelmaan parhaiten sopivan kuvauksen. (Andre et al., 2001; Hartson et al., 1999)

UAF tarjoaa esimerkiksi heuristiseen arviointiin verrattuna huomattavasti monipuolisemmat työkalut ongelman kuvausta ja määrittelyä varten. Hierarkkisesti rakennettu UAF ohjaa arvioijaa tasolta toiselle sen jälkeen, kun hän on valinnut, liittyykö käsiteltävä ongelma käyttäjän suunnitteluvaiheeseen, fyysisiin toimintoihin vai arvioimiseen. UAF:n hierarkkinen rakenne käy ilmi kuvan 7 esimerkistä, jossa käytettävyyso ngelma liittyy suunnitteluun. Kuvan ympyröissä on arvioijien lukumäärä. Ensimmäisellä ja toisella tasolla arvioijat ovat olleet yksimielisiä käytettävyyso ngelman kategoriasta. Kolmannella tasolla mielipiteet ovat jakaantuneet siten, että kolme arvioijaa



Kuva 7: Esimerkki käytettävyyso ngelman luokittelupolusta (Andre et al., 2001).

sijoittaa ongelman ”tavoitteiden jakaminen osiin” –kategoriaan ja seitsemän ”käyttäjän malli”-kategoriaan. Neljännellä tasolla mielipiteet jakaantuvat edelleen. Arvioijien välistä yksimielisyyttä tarkasteltaessa otetaan kuitenkin huomioon koko polku eikä vain viimeisen tason tilanne. Kuvan 7 esimerkissä käytettävyysongelman voitaisiin kuvata liittyvän korkean tason suunnitteluun. Korkean tason suunnittelu sisältää käyttäjän mallin järjestelmästä, minkä perusteella käyttäjän tulisi saada järjestelmästä kokonaiskuva. (Andre et al., 2001)

Andren ja muiden (2001) tutkimukset osoittavat, että UAF:n avulla oli mahdollista vähentää arvioijan vaikutusta. Vaikka kaikkia käytettävyyssiantuntijoiden välisiä eroavaisuuksia ei voitukaan poistaa, parani tutkimusten reliabiliteettitaso kuitenkin merkittävästi. Myös raportoinnin laatu parani.

19.6. Keskustelua

Yhteenvetona voidaan todeta, että vaikka arvioijan vaikutus tuntuu mahdottomalta poistaa ainakaan kokonaan, sen hallintaan on olemassa useita erilaisia työkaluja. Käytännön arviointityön taustalle on kehitetty teoreettisia viitekehyksiä. Sekä käyttäjän että käytettävyyssiantuntijan toimintoja on mallinnettu. Käytettävyyssiantuntijoiden käytössä on erilaisia apuvälineitä kuten raportointimalleja, priorisointimalleja sekä arvioinnin ohjeistuksia.

Tutkimuksilla on kyetty osoittamaan, että arvioijan vaikutusta voidaan vähentää. Esimerkiksi Andren ja muiden (2001) kehittämän käyttäjän toimintoihin perustuvan viitekehyksen UAF:n käyttö vähensi merkittävästi arvioijan vaikutusta. Arvioijien välisiä eroja ei senkään käyttö kuitenkaan kokonaan poistanut.

Arvioijien erilaisuudella on myös positiivinen kääntöpuolensa. Jos kaikki arvioijat löytäisivät aina täysin samat huonosti suunnitellut kohdat ohjelmasta, jotkut ongelmat jäisivät kokonaan löytymättä. Kukaan käytettävyyssiantuntija tuskin kykenee löytämään kaikkia olemassa olevia suunnitteluvirheitä, jotka potentiaalisesti ovat käytettävyysongelmia.

Käytettävyyden arvioinnin tavoitteena on siis löytää mahdollisimman monta arvioitavassa kohteessa olevista käytettävyysongelmista. Yksikään ongelma ei saisi jäädä huomaamatta, mutta toisaalta havaittujen ongelmien tulisi olla todellisia myös reaali maailmassa.



Raija Perälä, FM. Valmistuin tietojenkäsittelytieteiden laitokselta vuonna 2004 pääaineena vuorovaikutteinen teknologia. Pro gradu -työni aiheena oli digitaalisen kirjaston käytettävyys ja tutkimuskohteena netLibrary. Tällä hetkellä työskentelen DOIP-Center Oy:ssä ICT-suunnittelijan ja kouluttajan tehtävissä.

Ohjaaja: Saila Ovaska
Opponentit: Leena Helin ja Sanna Ranne

Lähteet

Andre, T.S., Hartson, H.R., Belz, S.M., & McCreary, F.A. (2001) The User Action Framework: A reliable foundation for usability engineering support tools. *International Journal of Human-Computer Studies*, 54, 107–136.

Cockton, G. & Woolrych, A. (2002) Sale must end: Should discount methods be cleared off HCI's shelves? *interactions* 9(5) September & October 2002, 13–18.

Cockton, G., Woolrych, A., Hall, L. & Hindmarch, M. (2003) Changing analysts' tunes: the surprising impact of a new instrument for usability inspection method assessment. In *People and Computers XVII: Proceedings of HCI'03*, Springer, 145–161.

Hartson, H.R., Andre, T.S., Williges, R.C. & van Rens, L. (1999) The User Action Framework: A theory-based foundation for inspection and classification of usability problems. In Bullinger, H.-J. & Ziegler, J. (Eds.) *Human-Computer Interaction: Ergonomics and user interfaces. Proceedings of HCI International 99*, Volume 1, 1058–1062. Mahwah, NJ: Lawrence Erlbaum.

Hassenzahl, M. (2000) Prioritizing usability problems: data-driven and judgement-driven severity estimates. *Behaviour & Information Technology* January-February 2000, 19(1), 29–42.

Hertzum, M & Jacobsen, N.E. (1999) The evaluator effect during first-time use of the cognitive Walkthrough technique. *Human-Computer Interaction: Ergonomics and User Interfaces. Proc. of HCI International '00*, 1063–1067.

Hertzum, M. & Jacobsen, N.E. (2001) The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction*, 13(4), 421–443.

Hertzum, M., Jacobsen, N.E. & Molich, R. (2002) Usability inspections by groups of specialists: perceived agreement in spite of disparate observations. Proc. CHI '02 extended abstracts on *Human factors in computing systems*, 662–663.

Jacobsen, N.E., Hertzum, M. & John, B.E. (1998) The evaluator effect in usability tests. *Proc. Human Factors in Computing Systems (CHI98)*, ACM Press, 255–256.

Lavery, D., Cockton, G. & Atkinson, M.P. (1997) Comparison of evaluation methods using structured usability problem reports. *Behaviour & Information Technology*, 16 (4/5), 246–266.

Nielsen, J. (1992) Finding usability problems through heuristic evaluation. *Proc. Human Factors in Computing Systems (CHI'92)*, ACM Press, 373–380.

Nielsen, J. (1993) *Usability engineering*. Academic Press, Boston.

Nielsen, J. (1994) Heuristic evaluation. In Nielsen, J., & Mack, R. L. (Eds.), *Usability Inspection Methods*, 25–62. New York, NY: John Wiley & Sons.

Sears, A. (1997) Heuristic walkthroughs: Finding the problems without the noise. *International Journal of Human-Computer Interaction*, 9 (3), 213–234.

Kuvien lähteet

Kuva 1: Uudelleen piirretty ja käännetty suomeksi. Alkuperäinen kuva löytyy lähteestä: Lavery, D., Cockton, G. & Atkinson, M.P. (1997) Comparison of evaluation methods using structured usability problem reports. *Behaviour & Information Technology*, 16 (4/5), 246–266.

Kuva 2: Käännetty suomeksi. Alkuperäinen kuva: Nielsen, J. (1994) Heuristic evaluation. In Nielsen, J., & Mack, R.L. (Eds.), *Usability Inspection Methods*, 25–62. New York, NY: John Wiley & Sons. Se löytyy myös lähteestä Nielsen, J. How to conduct a heuristic evaluation. http://www.useit.com/papers/heuristic/heuristic_evaluation.html (19.5.2004)

Kuva 3: Jacobsen, N.E & John, B.E. (1998) The evaluator effect in usability studies: problem detection and severity judgements. In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, 136–1340.

Kuva 4: Hertzum, M & Jacobsen, N.E. (1999) The Evaluator Effect during First-Time Use of the Cognitive Walkthrough Technique, *Human-Computer Interaction: Ergonomics and User Interfaces. Proceedings of HCI International '00*, 1063–1067.

Kuva 5: Uudelleen piirretty ja käännetty suomeksi. Alkuperäinen kuva löytyy lähteestä: Lavery, D., Cockton, G. & Atkinson, M.P. (1997) Comparison of evaluation methods using structured usability problem reports. *Behaviour & Information Technology*, 16 (4/5), 246–266.

Kuvat 6 ja 7: Käännetty suomeksi. Alkuperäiset kuvat löytyvät lähteestä: Andre, T.S., Hartson, H.R., Belz, S.M. & McCreary, F.A. (2001) The User Action Framework: A reliable foundation for usability engineering support tools. *International Journal of Human-Computer Studies*, 54, 107–136.