

# Kyselylomakkeet käytettävyytutkimuksessa

# 2

*Toni Vanhala*

Kyselylomakkeita voidaan käyttää monien erilaisten tutkimusten osana, yhtenä useista tai ainoana tiedonkeruumenetelmänä. Kyselyn soveltuminen käytettävyytutkimuksen osaksi riippuu tutkimuksen tavoitteista ja valitusta tutkimusstrategiasta. Tutkimuksen tavoitteet ja tarkastelukohde vaikuttavat myös siihen, kannattaako käyttää valmista, muokattua vai täysin itselaadittua lomaketta. Omien kysymysten laadinta on haastavaa ja mahdollisia ongelmia on lukuisia. Lomakkeen onnistunut laadinta edellyttääkin huolellista suunnittelua ja toistuvaa lomakkeen arviointia ja muokkausta. Aina ennen oman lomakkeen laatimista kannattaakin perehtyä valmiisiin lomakkeisiin ja niiden kysymyksiin. Myös käytettävyyden arviointiin on jo olemassa lukuisia valmiita lomakkeita. Lomakkeen suunnittelun lisäksi myös kyselyn suorittamiseen on aiemmassa tutkimuksessa annettu monia ohjeita, joiden avulla voidaan taata tulosten luotettavuus ja pätevyys. Luotettavuuden ja pätevyyden lisäksi kannattaa huomioida jo kyselyä suunniteltaessa tulosten analyysin vaatima työmäärä, johon yksittäisistä seikoista ehkä eniten vaikuttaa kysymysten tyyppi. Avoimet kysymykset synnyttävät laadullista aineistoa, jonka käsittely vaatii enemmän resursseja kuin määrällisen aineiston, joka muodostuu vaihtoehtokysymysten vastauksista. Lisäksi lomakkeiden käyttöön käytettävyytutkimuksissa liittyy monia erityisesti huomioitavia seikkoja ja valintoja, joiden merkitys on syytä ymmärtää. Esimerkiksi lomakkeiden avulla suoritettun kyselyn ja arvioitavan sovelluksen käyttämisen välinen aika saattaa vaikuttaa vastauksiin merkittävästi.

## 2.1. Johdanto

Kyselylomakkeiden avulla kerätään tietoa vastaajien ajatuksista, mielipiteistä ja tuntemuksista. Lomakkeen avulla tiedot voidaan kerätä yhtenäisellä tavalla, jolloin tutkimuksen kannalta epäoleelliset tekijät ovat hallittavissa.

Kysely mahdollistaa tutkimuksen osallistujajoukon kasvattamisen suhteellisen vähäisellä resurssien lisäämisellä. Aina suuri vastaajien määrä ei kuitenkaan takaa onnistunutta tutkimusta vaan saattaa jopa haitata tutkimuksen tavoitteita.

Kyselylomakkeita voidaan käyttää useimpien tutkimusten osana. Kyselylomakkeita voidaan käyttää muita menetelmiä täydentävänä tiedonkeruumenetelmänä, mutta ne soveltuvat myös tutkimuksen ainoaksi tiedonkeruumenetelmäksi. Tutkimuskysymykset, tutkittava kohde ja taustalla oleva teoria määräävät, kuinka hyvin kyselylomakkeet lopulta soveltuvat tutkimuksen osaksi. Eniten lomakkeiden käyttöä rajoittaa niillä kerätyn tiedon subjektiivinen luonne.

Käytettävyytutkimuksen tarpeisiin on suunniteltu lukuisia vapaasti käytettäviä ja kaupallisia lomakkeita. Selvin rajoite valmiiden lomakkeiden käytölle suomalaisessa tutkimuksessa on

niiden englanninkielisyys. Valmiit lomakkeet eivät myöskään välttämättä anna riittävän yksityiskohtaista tietoa tutkimuksen kohteesta. Muun muassa näiden syiden vuoksi on lomakkeiden muokkaaminen tai täysin uuden lomakkeen laadinta usein tarpeen.

Lomakkeen laatiminen on virhealtis prosessi, joka vaatii asiaan paneutumista. Joitakin yleisiä ohjeita ja neuvoja laadintaan voidaan antaa, mutta pätevän ja luotettavan lomakkeen laatiminen edellyttää toistuvaa testausta ja muokkaamista. Lomakkeen laadinta voidaankin rinnastaa käytettävyystudkimukseen, joka tuotekehityksen osana on useimmiten iteratiivista: prototyyppeä arvioidaan ja muokataan toistuvasti.

Kyselyn toteuttamiseen liittyy lomakkeen valinnan ja suunnittelun lisäksi monia päätöksiä. Vastajien valinta, lomakkeiden toimitustapa ja vastausten analysointimenetelmä voivat kaikki vaikuttaa tutkimuksen vaatimaan työmäärään ja siitä saatujen tulosten pätevyyteen ja luotettavuuteen merkittävästi. Tässä työssä ei kuitenkaan kyetä tyhjentävään esitykseen eri ratkaisuihin liittyvistä ongelmista, hyödyistä ja haitoista. Tarkoituksena on antaa lukijalle yleinen käsitys kyselymenetelmistä ja valmiudet perehtyä oman tutkimuksen kannalta merkittäviin aiheisiin.

## 2.2. Kyselyn soveltuminen tutkimukseen

Ennen kuin kysely liitetään osaksi tutkimusta, on syytä arvioida voidaanko tutkimuksen tavoitteet saavuttaa kyselyn avulla. Tähän vaikuttavat tutkimustavoitteiden lisäksi valittu tutkimusstrategia ja tutkimuksen sijoittuminen tuotekehityksen tiettyyn vaiheeseen.

### 2.2.1. Tutkimuksen tavoite ja tutkimusstrategia

Tutkimukset voidaan yleisesti jakaa *eksploratiivisiin (explorative)*, *kuvaileviin (descriptive)* ja *selittäviin (explanatory)*. Eksploratiivisessa tutkimuksessa tutkija pyrkii löytämään uusia ilmiöitä tai kartoittamaan nykyisen tilanteen, jota tunnetut teoriat eivät kata. Selittävä tutkimus pyrkii selittämään tilanteen tai ongelman, useimmiten syy- ja seuraussuhteiden avulla. Kuvaileva tutkimus edellyttää jonkinlaista tietämystä tutkittavasta ilmiöstä, jolloin tutkimuksen kohteesta voidaan kerätä riittävästi tietoa tarkan kuvauksen muodostamiseksi.

Kyselylomakkeiden käyttö tuntuu soveltuvan erityisesti kokeellisten tutkimusten, *laajojen kyselyiden (survey)* ja tapaustutkimusten yhteyteen. Erityisesti laajassa kyselyssä lomake soveltuu tutkimuksen ainoaksi tiedonkeruumenetelmäksi, korvaten esimerkiksi haastattelun, joka on toinen kyselytutkimuksen yleisimmistä aineiston keruumenetelmistä. Lomakekysely mahdollistaa laajemman osallistujajoukon käyttämisen suhteellisen pienellä resurssien lisäämisellä.

Kokeelliset tutkimukset voidaan jakaa konrolloituihin kokeisiin ja kenttäkokeisiin. Kontrolloidut kokeet tapahtuvat kokeen järjestäjän konrolloimassa ympäristössä, kuten käytettävyysslaboratoriossa. Kenttäkokeet tapahtuvat ”oikeassa” ympäristössä, jota kokeen järjestäjä pystyy konrolloimaan hyvin rajoitetusti. Kenttäkokeen erikoistyyppi on luonnollinen koe, jossa koeasetelma syntyy ilman tutkijan tekemiä järjestelyjä.

Kaikille kokeellisille tutkimuksille on yhteistä syy- ja seuraussuhteiden hakeminen; kokeellinen tutkimus pyrkii selvittämään, mitkä muuttujat ovat yhteydessä toisiinsa. Kokeellinen tutkimus onkin lähes aina selittävää tutkimusta. Kokeellisesti voidaan esimerkiksi tutkia, lisääkö hypermediateos opiskelijoiden motivaatiota. Tässä esimerkissä hypermediateos on *riippumaton muuttuja*, jonka käyttöä tai saatavuutta kokeen järjestäjä konrolloi, ja motivaatio on *riippuva muuttuja*, jonka muutoksia pyritään arvioimaan suhteessa riippumattomaan muuttujaan.

Kysely on halpa tapa saada nopeasti tietoa laajasta joukosta (Järvinen & Järvinen, 2000). Yleiskatsauksessa (survey) sama kysely toimitetaan laajalle osallistujajoukolle. Myös verkkosivuilla esitetyt, kaikille sivuston käyttäjille suunnatut kyselyt ovat yleiskatsauksia.

Kyselylomakkeet ovat ainoa vartenotettava menetelmä näin laajaa aineistoa kerätessä. Osallistujien suuren määrän ja heterogeenisuuden vuoksi useat lomakkeiden suunnittelukysymyksistä ovat erityisen tärkeitä yleiskatsauksissa. Erityisesti kyselyyn vastaamattomuus ja tästä aiheutuva vastaajien valikoituminen ovat merkittäviä riskejä, joihin on vaikeaa varautua.

Tapaustutkimus pyrkii keräämään suuren määrän tietoa yksittäisestä tapauksesta. Lomake on houkuttelevan helppo tapa kerätä suuri määrä aineistoa pienellä työmäärällä. Yleensä käytössä on useita rinnakkaisia tiedonhankintamenetelmiä, kuten kyselyt, haastattelut, havainnointit ja dokumenttianalyysit (Robson, 1994).

Käytettävyydestit sopivat tapaustutkimusten määritelmään, sillä niissä usein kerätään tietoa useilla rinnakkaisilla menetelmillä, kuten käyttäjän toimintaa tarkkailemalla, haastattelulla ja kyselylomakkeilla. Tarkoituksena sekä tapaustutkimuksessa että käytettävyydestissä on tutkittavan kohteen tarkka kuvaaminen, esimerkiksi kaikkien käytettävyyso Ongelmien löytäminen. Useimmiten testaus on kuvailevaa tutkimusta, sillä aiemman tutkimuksen perusteella voidaan määritellä kiinnostuksen kohteet, joista tietoa pyritään keräämään. Esimerkiksi vakiintuneita heuristiikkoja voidaan käyttää etsittäessä mahdollisia ongelmia, mikä ohjaa käytettävyydestissä suoritettavien tehtävien laadintaa.

Taulukko 1 esittää yleisimmät tutkimusstrategiat ja niitä noudattavan tutkimuksen yleiset ominaisuudet. Taulukko on koottu yleisistä tutkimusoppaista (mm. Robson, 1994).

**Taulukko 1:** Tutkimusstrategioihin liittyvät ominaisuudet.

tutkimusstrategia	osallistujien määrä	tavoite	kysymykset	ongelmat	erityistä
kyselytutkimus	satoja	kuvaaminen	kuka, mitä, missä, kuinka monta, kuinka paljon	vastaajien valikoituminen	vastaajien valinta tärkeää, kerätyn tiedon määrä ei takaa yleistettävyyttä ja hyödyllisyyttä
kokeellinen tutkimus	kymmeniä	selittäminen	kuinka, miksi	harhat, väliintulevat muuttujat	vaatii tilanteen ja koeasetelman hallinnan
tapaustutkimus	yksi tai muutamia	uusien ideoiden tuottaminen	kuinka, miksi	heikko yleistettävyyttä, työläs	käytössä useita rinnakkaisia tiedonkeruumenetelmiä

Yleiskatsauksessa tutkimuksen tavoitteena on useimmiten kuvata nykyinen tilanne, kuten vallitseva mielipide. Kokeellisessa tutkimuksessa tavoitteena on ilmiön selittäminen syy- ja seuraussuhteiden avulla. Tapaustutkimuksella pyritään useimmiten yksittäisen tapauksen tarkan kuvauksen avulla tuottamaan uusia ideoita, esimerkiksi muissa vastaavissa tapauksissa käytettäväksi.

### 2.2.2. Kysely eri tuotekehitysvaiheissa

Koska kyselylomakkeet soveltuvat tiedonkeruumenetelmäksi lähes jokaiseen tutkimukseen, voidaan niitä käyttää myös kaikissa tuotekehityksen vaiheissa, ehkä varsinaista toteutusvaihetta lukuun ottamatta.

Esitutkimusvaiheessa lomakkeilla voidaan kerätä tietoa kohderyhmän ominaisuuksista ja tarpeista. Suunnitteluvaiheessa erilaisten vaihtoehtojen hyviä ja huonoja puolia voidaan vertailla, esimerkiksi pyytämällä ihmisten mielipiteitä tuotteen tai käyttöliittymän erilaisista ominaisuuksista ja toteutusvaihtoehdoista. Näissä vaiheissa lomakkeiden mahdollistamasta suuresta osallistujajoukosta on erityisesti hyötyä. Suurella osallistujien määrällä kaikki käyttäjärühmät tulevat varmemmin katetuksi, minkä ansiosta tulokset ovat luotettavampia.

Valmiiden tuotteiden ja prototyyppien arvioinnissa kyselyiden hyödyt liittyvät lähinnä menetelmän yhdenmukaisuuteen. Kysymykset esitetään kaikille vastaajille samalla tavalla, jolloin vastauksiin vaikuttavien epäoleellisten tekijöiden kontrollointi on helpompaa kuin useimmilla muilla menetelmillä. Käytettävyysongelmista ja tutkittavan kohteen ominaisuuksista ei kuitenkaan lomakkeella välttämättä saa yksityiskohtaista tietoa, joten muiden täydentävien menetelmien käyttö on usein tarpeellista.

Tuotteen valmistumisen ja käyttöönoton jälkeen kyselyillä voidaan kartoittaa tuotetta pidemmän aikaa käyttäneiden vastaajien ongelmia ja mielipiteitä. Kyselyn tuloksia voidaan käyttää kyselylomakkeen itsensä ja tuotteen seuraavien versioiden pohjana ja koulutuksen, ohjeistuksen ja käyttötuen tarpeen arvioimisessa.

### 2.2.3. Muut tiedonkeruumenetelmät

Kyselyn ehdoton etu on sen tarjoama mahdollisuus kerätä kohtuullisella vaivalla suuri määrä tietoa suurelta vastaajajoukolta. Kyselyllä kerätyn tiedon merkitystä ei kuitenkaan pidä yliarvioida, eikä tiedon määrä korvaa sen laatua.

Kyselyllä kerätty tieto arvioinnin kohteesta on aina subjektiivista eli vastaajan tulkintaan perustuvaa. Esimerkiksi kyselyllä kerätty arvio tuotteesta ei suoraan kerro sen käytettävyydestä, eikä arvion taustalla vaikuttavien tekijöiden erittely ole helppoa. Miellyttävyyden arvioiminen edellyttää vastaajalta vain hänen omien subjektiivisten tuntemustensa erittelyä, mikä vaatii vastaajalta vähemmän tulkintaa kuin esimerkiksi käytettävyyden arviointi. Näin ollen käyttäjän subjektiivisten tuntemusten mittaaminen lomakkeella voisi olla suositeltavampaa kuin kohteen objektiivisten ominaisuuksien.

Kohteen objektiivisten ominaisuuksien, kuten tehokkuuden ja toiminnallisuuden, mittaamisessa muut menetelmät ovat ylivoimaisia. Esimerkiksi tehtävien suoritusajan mittaaminen ja muilla tavoin toteutettu suora havainnointi antavat kyselyä luotettavampaa tietoa varsinaisesta tehokkuudesta. Joskus vaikutelma tehokkuudesta on kuitenkin varsinaista tehokkuutta tärkeämpi, jolloin käyttäjän tuntemukset ovat jälleen kiinnostuksen kohteena. Kyselyllä mitatun tiedon subjektiivisuus voi siis tutkimustavoitteen mukaan olla joko toivottavaa tai haitallista.

Haastatteluun verrattaessa kyselyn etuna on vähäisempi resurssien tarve: varsinaisen haastattelun lisäksi aikaa saattaa kuluu myös esimerkiksi vastausten kirjaamiseen nauhoitetuista haastatteluista. Lisäksi haastattelija voi vaikuttaa käyttäytymisellään ja olemuksellaan haastattelussa annettuihin vastauksiin. Toisaalta henkilökohtaisen haastattelun ympäristö voidaan pitää melko kontrolloituna, verrattuna esimerkiksi verkkokyselyyn, johon vastattaessa saataan vaikkapa puhua puhelimesta tai kuunnella musiikkia.

Kyselyn suorittaminen edellyttää melko hyvää tuntemusta aihealueesta ja jonkinlaista kyselyä ohjaavaa teoriaa tai viitekehystä, sillä kysymykset ja vaihtoehdot on kyettävä laatimaan kiinnostuksen kohdetta mittaaviksi ja kattaviksi. Haastattelun etuna on mahdollisuus tarkentaa vastauksia ja korjata väärinymmärryksiä.

### 2.2.4. Kysely tutkimuksen osana

Kysely soveltuu useisiin tarkoituksiin ja näin ollen myös erilaisiin tutkimuksiin. Tässä kohdassa kuvataan muutamaa esimerkkitutkimusta, joissa lomakkeita on käytetty ainoana tiedonkeruutapana tai yhtenä tapana muiden joukossa.

Georgian teknillisen korkeakoulun (Georgia Institute of Technology) grafiikan, visualisoinnin ja käytettävyyden tutkimuskeskuksen toteuttama web-käyttöä kartoittanut verkkolomakkeella suoritettu kysely (GVU's 10th WWW user survey, 1998) on hyvä esimerkki laajaa aineistoa hyödyntävästä kuvailevasta kyselytutkimuksesta. Esimerkiksi eräs tutkimuksen kysymyksistä kartoitti Internetin käytössä kohdattuja ongelmia ("What do you find to be the biggest problems in using the Web?"). Lomakkeella annettiin 18 vastausvaihtoehtoa, joista kolme useimmin valittua olivat hitaasti latautuvat mainokset, sivujen hidas latautuminen ja toimimattomat hypertekstilinkit sivujen välillä.

Tutkimuksen tulokset kuvaavat silloisen web-käytön ominaisuuksia, ja koska vastaava kysely oli aiemmin suoritettu lukuisia kertoja, saadaan myös ilmiöiden kehittymisestä tietoa. Toisaalta peräkkäisten kyselyiden tuloksia vertaamalla voidaan myös tehdä päätelmiä kyselyn luotettavuudesta. Lomake sisälsi myös kysymyksen, jonka avulla selvitettiin, mistä vastaaja sai tietää järjestetystä kyselystä. Tämän tiedon avulla eroja peräkkäisiin kyselyihin vastaajien välillä ja vastausten mahdollista vinoutumista voitiin arvioida.

Ennen kokeen tai käytettävyydestin suorittamista kerätään usein taustatietoa osallistujista kyselylomakkeella (tarkemmin raportin luvuissa 13 ja 18). Tieto sukupuolesta, iästä ja aiemmasta käyttökokemuksesta auttaa käyttäjien ryhmittelyssä. Kerätyllä tiedolla voidaan pyrkiä varmistamaan, että kokeeseen osallistunut ryhmä vastaa suurempaa ryhmää, jota koskeviksi tulokset halutaan yleistää. Tämä ryhmä koostuu kaikista kiinnostuksen kohteena olevista yksilöistä, ja sitä kutsutaan *populaatioksi*.

Osallistujista kerätyn tiedon avulla voidaan myös pyrkiä kontrolloimaan ja arvioimaan *väliintulevia muuttujia*. Väliintulevat muuttujat vaikuttavat kiinnostuksen kohteena olevaan riippuvaan muuttujaan ja häiritsevät näin riippumattoman muuttujan vaikutuksen havaitsemista. Esimerkiksi aiemmat käyttökokemukset vastaavista järjestelmistä ja käyttöliittymistä voivat helpottaa tai hankaloittaa uuden omaksumista ja näin ollen vaikuttaa käytettävyydestin tehtävistä suoriutumiseen.

Bernardin ja muiden (2002) tutkimuksessa vertailtiin erityyppisten kirjasinten luettavuutta verkkosivuilla. Lomaketta käytettiin koulutustason, näkökyvyn, iän ja verkkosivujen lukemistiheyden keräämisessä. Näillä ominaisuuksilla voi selvästikin olla vaikutusta lukukykyyn. Kalbach ja Bosenick (2003) tutkivat verkkosivujen navigointivalikon sijainnin vaikutusta käytettävyyteen, ja keräsivät pohjatietoa osallistujien ammatista, Internetin käyttökokemuksesta ja kärsivyydestä.

Suoritetulla kyselyllä voidaan myös pyrkiä mittaamaan jotakin riippuvaa muuttujaa, josta muilla menetelmillä on vaikeaa tai mahdotonta kerätä tietoa. Tällaiset muuttujat liittyvät useimmiten osallistujan sisäisiin ominaisuuksiin ja prosesseihin, kuten ajatuksiin, asenteisiin, mielipiteisiin ja tunnetiloihin. Esimerkiksi kahdesta erilaisesta käyttöliittymästä saatujen vaikutelmien erot voidaan yhdistää käyttöliittymän ominaisuuksiin, kuten Björk ja muut (1999) tekivät vertaillessaan kahta erilaista kämmentietokoneen verkkoselainta. Osallistujia pyydettiin arvioimaan selaimen yleisnäkymän (overview) laatua ja hakemisen helpoutta.

Tällainen asetelma on kuitenkin lähempänä kahta erillistä tapaustutkimusta, joiden tuloksia myöhemmin verrataan. ”Puhdas” kokeellinen tutkimus edellyttäisi, että lomakkeella mitattu muuttuja olisi suoraan yhteydessä kiinnostuksen kohteena oleviin käyttöliittymän ominaisuuksiin. Björkin ja muiden tutkimuksessa selaimet erosivat toisistaan monilta osin, jolloin ei pystytä varmasti sanomaan, mitkä ominaisuudet tai käyttöliittymän piirteet todella vaikuttivat arvioihin.

Lomakkeiden käytössä tapaustutkimusten yhteydessä on usein ongelmana taustalla olevan teorian puutteellisuus tai puuttuminen (Hufnagel & Conca, 1991). Mitatut muuttujat ja lomakkeen kysymysten yhteys näihin muuttujiin olisi kuitenkin kyettävä perustelemaan. Tämä voi olla ongelmana etenkin eksploraatiivisissa tapaustutkimuksissa, joissa pyritään löytämään uusia ilmiöitä tai muodostamaan uusia teorioita.

Yksittäisestä tapauksesta voidaan kuitenkin kyselyn avulla saada suuri määrä tietoa. Tieto kerätään yhtenäisellä tavalla, mikä helpottaa sen analysointia. Root ja Draper (1983) keräsivät Pascal-ohjelmointikielellä kirjoitettujen ohjelmien muokkaamiseen tarkoitetun editorin jokaisen yksittäisen toiminnon käytöstä tietoa, jonka avulla editorin ongelmakohdat pystyttiin tunnistamaan.

Esimerkkinä eksploraatiivisesta käytettävyydestutkimuksesta voidaan mainita Nielsenin (1994) tutkimus, jossa kuusiportaisella asteikolla arvioitiin, kuinka hyvin heuristiikat soveltuivat aiemmista testeistä kerättyjen 249:n ongelman selittämiseen. Arvioituja heuristiikkoja oli noin sata, joista tutkimuksen tuloksiin sovelletun faktorianalyysin avulla muodostettiin yhdeksän heuristiikkaa. Varsinaista kyselytutkimusta tutkimuksessa ei kuitenkaan suoritettu, sillä ainoastaan Nielsen itse arvioi heuristiikkojen soveltuvuutta. Vastaavanlainen kartoitus heuris-

tiikkojen selittävydestä olisi kuitenkin voitu toteuttaa yleiskatsauksena, jossa kyselylomake olisi toimitettu laajalle joukolle vastaajia.

Laajalle vastaajajoukolle toimitettu kysely voi tarjota tietoa ongelmien yleisyydestä ja merkittävydestä, mutta ongelmista tarkemman kuvan muodostaakseen kannattaa käyttää kyselyn rinnalla muita menetelmiä, kuten käytettävyydestä. Ongelmien kuvaaminen useimmiten edellyttäisi avoimia kysymyksiä, joihin voi antaa vapaasti muotoillun vastauksen. Tämä aiheuttaa lisävaatimuksia tuloksen analyysille. Vapaamuotoisten vastausten analyysin vaatimat lisäresurssit heikentävät kyselyn vahvuuksia, tiedon keräämisen ja analysoinnin helppoutta.

## 2.3. Valmiit kyselylomakkeet

On olemassa lukuisia käytettävyyden arviointiin kehitettyjä kyselylomakkeita. Valmiit lomakkeet mittaavat tuotteen tai käyttöliittymän yleisiä ominaisuuksia. Näin ollen ne eivät sovellu tietyn ohjelmiston tai käyttöliittymän piirteiden yksityiskohtaiseen arviointiin. Yleensä lomakkeilla mitattu käytettävyys on jaettu muutamaaan ulottuvuuteen, jotka on saatettu muodostaa lomakkeen aiempien versioiden testaamisen ja esimerkiksi faktorianalyysin avulla.

Standardoitujen lomakkeiden kehittäminen on pitkä iteratiivinen prosessi, jossa lomaketta muokataan ja testataan toistuvasti. Standardoidun lomakkeen käyttö takaa tulosten vertailukelpoisuuden. Lomakkeiden vertaaminen yleisesti on vaikeaa, sillä täysin objektiivista ja luotettavaa käytettävyyden mittaria ei tunneta. Joitakin vertailuja on kuitenkin tehty, ja näiden tulosten perusteella voidaan sanoa, että ainakin jotkut lomakkeet antavat yhteneväisiä tuloksia. Kirakowski (1994) kuvaa kaksi tällaista tutkimusta (Lucey, 1991; Wong & Rengger, 1991).

Eri lomakkeiden ulottuvuuksien välillä vaikuttaa olevan joitakin yhtäläisyyksiä, mutta toisaalta ne myös poikkeavat toisistaan merkittävästi. Ivesin ja muiden (1983) mukaan julkisten testitulosten puuttuminen estää lomakkeiden vertailun ja verifiointin, mikä olisi välttämätöntä yleisen käytettävyyden mittarin kehittämiseksi.

Seuraavassa esitellään lyhyesti muutama käytettävyyden tai sen osatekijöiden mittaamiseen tarkoitettu lomake.

### 2.3.1. Software Usability Measurement Inventory (SUMI)

SUMI on pitkän kehitystyön tuloksena muotoutunut lomake, joka on saatavilla myös suomenkielisenä. Lomake koostuu väitteistä, joihin vastataan ”samaa mieltä”, ”eri mieltä”, tai ”en tiedä”. Ammatillaisen käyttöön tarkoitettu aloituspakkaus sisältää ohjekirjan, tulosten laskentaan tarkoitetun Windows-ohjelman, ja 50 kyselylomaketta. Pakkauksen hinta on nykyisin (1.3.2004) hiukan alle 1000 euroa. Lomakkeita on saatavissa lisää 50 lomakkeen erissä, joista jokainen maksaa noin 60 euroa.

SUMI jakaa ohjelmiston *käytettävyyden* (*usability*) viiteen ulottuvuuteen, jotka ovat tehokkuus (efficiency), tunne (affect), avuliaisuus (helpfulness), hallinta (control) ja opittavuus (learnability) (Kirakowski, 1994). Jokaista ulottuvuutta kohti lomakkeella on 10 väitettä, joten lomake sisältää yhteensä 50 väitettä. Näiden lisäksi lomake mittaa yleistä (global) käytettävyyttä, joka lasketaan 25 väitteeseen annetuista vastauksista. Nämä väitteet sisältyvät käytettävyyden osatekijöitä mittaaviin 50 väitteeseen. Lomakkeen sisältö on muotoutunut iteratiivisessa kehityksessä, jossa lomaketta on toistuvasti testattu ja sisältöä muokattu tulosten luotettavuuden ja kuvaavuuden parantamiseksi.

SUMI on standardoitu lomake. Lomakkeella saadut tulokset noudattavat tiettyä jakaumaa, joten saadun tuloksen perusteella testatun järjestelmän käytettävyyttä voi verrata markkinoiden yleiseen tilanteeseen. Toisin sanoen tuloksen perusteella voi päätellä, kuinka suuri osa järjestelmistä ja tuotteista on käytettävyydeltään parempia tai huonompia. Lomakkeen vertailukyvyn säilyttäminen edellyttää standardoinnin toistamista säännöllisin väliajoin. SUMI-lomake standardoidaan vuosittain.

### 2.3.2. Questionnaire for User Interaction Satisfaction (QUIS)

QUIS-lomakkeen versio 5.0 löytyy Chinin ja muiden (1988) artikkelista. Lomake arvioi käytettävyyttä viidellä ulottuvuudella: yleisvaikutelmaa (overall reactions to the software), näyttöä (screen), terminologiaa ja järjestelmän tietoja (terminology and system information), oppimista (learning) ja järjestelmän kyvykkyyttä (system capabilities).

Lomake koostuu kymmenportaisista asteikoista, joiden avulla järjestelmää koskevat väitteet ja järjestelmän ominaisuudet arvioidaan. Asteikolla nolla voi merkitä esimerkiksi vaihtoehtoa ”ei koskaan” tai ”vaikea”, ja yhdeksän merkitsee vaihtoehtoa ”aina” tai ”helppo”. Suurempi lukuarvo on siis aina pienempää parempi. Yleisvaikutelmaa mittaava osa lomakkeen versiosta 5.0 on esitetty taulukossa 2.

Kaupallinen QUIS 7.0 sisältää edellisten ulottuvuuksien lisäksi myös muutaman vapaaehtoisen osan, joiden avulla arvioidaan muun muassa opastuksen onnistuneisuutta ja ohjelman asennuksen helppoutta. Näitä osia mukaan ottamalla tai pois jättämällä lomaketta voidaan jonkin verran muokata tutkimukseen soveltuvaksi. Lomakkeen saa akateemiseen käyttöön 300 dollarilla, kaupallinen käyttö maksaa 1000 dollaria.

**Taulukko 2:** QUIS 5.0 lomakkeen yleisvaikutelmaa mittaavat kysymykset (Chin et al., 1988).

Yleiset reaktiot ohjelmistoon										
surkea										suurenmoinen
0	1	2	3	4	5	6	7	8	9	
vaikea										helppo
0	1	2	3	4	5	6	7	8	9	
turhauttava										tydyttävä
0	1	2	3	4	5	6	7	8	9	
tehoton										tehokas
0	1	2	3	4	5	6	7	8	9	
tylsä										innostava
0	1	2	3	4	5	6	7	8	9	
kankea										joustava
0	1	2	3	4	5	6	7	8	9	

### 2.3.3. Muut valmiit lomakkeet

Web site Analysis and MeasureMent Inventory (WAMMI) on SUMI-lomakkeen kehittäjän Jurek Kirakowskin ja Nigel Claridgen laatima, SUMIsta kuitenkin täysin erillinen lomake verkkosivustojen käytettävyyden arviointiin (WAMMI consortium, 2002). Myös WAMMI:sta on saatavana suomenkielinen versio, ja se on SUMIn tavoin standardoitu ja kaupallinen tuote.

Keinonen (1998) esittelee Lewisin (1995) kehittämät kolme kyselylomaketta. After Scenario Questionnaire (ASQ) on tarkoitettu käytettäväksi käytettävyydestin jokaisen tehtävän suorituksen jälkeen. Post Study System Usability Questionnaire (PSSUQ) on tarkoitettu käytettävyydestin jälkeisen arvion keräämiseen ja Computer System Usability Questionnaire (CSUQ) on suunnattu kenttätutkimuksiin.

Kaikki Lewisin lomakkeet käyttävät 7-portaista Likert-asteikkoa. ASQ mittaa tehtävästä suoriutumisen helppoutta, subjektiivista nopeusarviota ja saatavilla olevan tuen riittävyttä. PSSUQ ja CSUQ mittaavat järjestelmän hyödyllisyyttä, tiedon laatua, ja käyttöliittymän laatua. Likert-asteikkoa käsitellään tarkemmin kohdassa 2.4.1.

System Usability Scale (SUS; Brooke, 1996) koostuu kymmenestä väittämästä, joiden paikkansapitävyys arvioidaan 5-portaisella Likert-asteikolla. SUS-lomake on suomennettu

(taulukko 3). Lomakkeella kerättyjen vastausten analysointi on hyvin helppoa, sillä lomakkeella saatu arvio käytettävyydestä muodostuu väittämien saamien pisteiden summasta. Lomake antaa siis järjestelmän käytettävyydelle kokonaispistemäärän, joka sijoittuu nollan ja sadan väliin.

Muita tunnettuja käytettävyyden arviointiin tarkoitettuja lomakkeita ovat Usefulness, Satisfaction, and Ease of use (USE) –lomake (Lund, 2001), End-User Computing Satisfaction Instrument (EUCSI) (Doll & Torkzadeh, 1988), Technology Acceptance Model (TAM) (Davis, 1993) ja ISO 9241-10 -standardiin perustuva IsoMetrics (Gediga et al., 1999). Lomakkeiden välillä on suuria eroja, jotka vaikuttavat mitattaviin muuttujiin ja näin ollen myös mitattuihin käytettävyyden osatekijöihin. Esimerkiksi EUCSI, TAM ja PSSUQ mittaavat koettua hyödyllisyyttä (perceived usefulness), jota SUMI ja QUIS eivät suoranaisesti mittaa. Keinonen (1998) vertailee ansiokkaasti lomakkeita ja niitä oletuksia käytettävyydestä, joihin lomakkeet perustuvat.

### 2.3.4. Valmiin lomakkeen muokkaaminen

Kitchenham ja Pfleeger (2002b) mainitsevat neljä syytä valmiin lomakkeen muokkaamiseen: lomakkeen pituuden, vastaajien ja lomakkeen kohderyhmän eroavaisuudet, lomakkeen kielen ja suunnittelusta poikkeavan tiedonkeruumenetelmän.

Jos lomake on pitkä, epäoleelliset kysymykset ainoastaan turhauttavat vastaajia. Kaikissa tilanteissa ei pitkää kyselyä muutenkaan pystyttäisi suorittamaan, esimerkiksi silloin kun lomakkeet jaetaan ja täytetään kadulla, vaikkapa jonkinlaisen kenttätutkimuksen yhteydessä.

Kyselyyn vastaajat voivat erota siitä ryhmästä, jolle lomake on alun perin kehitetty. Esimerkiksi kokeneiden ja kokemattomien käyttäjien tai aikuisten ja lasten välillä on suuria eroja käsitteiden hallinnassa ja ajattelutavoissa. Tällöin kysymysten ja vaihtoehtojen sisältöä on syytä muokata.

Kaikista selvin syy lomakkeen muokkaamiseen on sen kääntäminen toiselle kielelle. Tämä harvoin onnistuu sanasta sanaan, eikä tällainen kääntäminen ole suositeltavaakaan. Valmiista lomakkeista ainoastaan SUMI-lomakkeesta on tiettävästi saatavilla virallinen suomenkielinen versio (SUMI, 2004).

Neljäntenä syynä Kitchenham ja Pfleeger (2002b) mainitsevat alkuperäisestä poikkeavan tiedonkeruumenetelmän käytön. Esimerkiksi konteksti, jossa kysely suoritetaan, voi vaikuttaa kysymysten tulkintaan ja vaatia niiden muotoilemista yksikäsitteisemmiksi. Lomaketta voidaan myös käyttää haastattelussa aineiston keräämiseen, kuten kuvataan raportin luvussa 3 (Vuorela). Tällöin kysymyksiä on ehkä muokattava puheenomaisemmiksi ja lyhyemmiksi.

Näiden syiden lisäksi lomaketta voidaan muokata, jotta tuotteen tai käyttöliittymän ominaisuuksista saataisiin yksityiskohtaisempi arvio. Valmiit lomakkeet mittaavat yleisiä ominaisuuksia, jotka ovat melko abstrakteja. Yksityiskohtaisempaa tietoa saadaan lisäämällä kysymyksiä tai muokkaamalla olemassa olevia yksityiskohtaisemmiksi, esimerkiksi yksittäisiä toimintoja tai näyttöruutuja koskeviksi.

## 2.4. Oman lomakkeen laadinta

Oman lomakkeen suunnittelu on tarpeen, jos valmiit lomakkeet eivät tunnu mittaavan tutkimuksen kannalta kiinnostavia ominaisuuksia. Toinen perustelu oman lomakkeen laatimiseksi on tarve kerätä yksityiskohtaisempaa tietoa kuin tutkimuksen kohteen sijoitusta joillakin käytettävyyden osaluottuvuuksilla. Tähän voi soveltaa täysin uuden lomakkeen sijasta myös muokattu lomake, jonka pohjana on käytetty jotakin valmista lomaketta. Erityisesti suomalaisten tutkimusten kannalta valmiin lomakkeen käyttö on ongelmallista, sillä suurin osa lomakkeista ei ole suomeksi saatavilla. Joka tapauksessa, aina omaa lomaketta laadittaessa on syytä tutkia valmiita lomakkeita, joiden sanamuotoja ja kysymysten sisältöä on varmasti mietitty perusteellisesti.

Jos tutkimuksen taustalla oleva teoria ei ohjaa muodostamaan varmasti kaikki olennaiset vastausmahdollisuudet kattavia vaihtoehtoja, on haastattelu luultavasti kyselyn sijasta paremmin soveltuva tiedonkeruumenetelmä. Haastattelussa voi välittömästi pyytää tarkennusta kerättyyn tietoon, ja haastattelu voikin tuoda muita menetelmiä paremmin uusia näkökohtia esiin (Järvinen & Järvinen, 2000).

Kyselyä suunniteltaessa on kannattavaa hakea ja tunnistaa mahdollisia ongelmia tutkimuksen validiteetissa. Näitä ongelmia kutsutaan myös harhoiksi. Tunnistamalla nämä harhat lomakkeen suunnitteluvaiheessa tutkija voi laatia ylimääräisiä kysymyksiä, joiden avulla vastaajat voidaan jakaa vertailuryhmiin (Hufnagel & Conca, 1991). Esimerkiksi tietojärjestelmän hyödyllisyyttä koskevaan kyselyyn vastaavat käyttäjät voidaan jakaa järjestelmän käyttökokemuksen mukaan aloitteleviin ja kokeneisiin käyttäjiin. Kokemuksen merkitys arvion muodostamisessa voidaan sulkea pois vertaamalla näiden kahden ryhmän vastausten eroja.

Seuraavaksi käydään läpi lomakkeen suunnittelun perusteita: kysymysten muotoiluja, asteikon valintaa ja lomakkeen arviointia – myös lomakkeen täyttäjän näkökulmasta.

### 2.4.1. Avoimet ja vaihtoehtokysymykset

Kysymykset voidaan laatia kahdella tavalla. Vastaajalle voidaan antaa valmiit vastausvaihtoehdot tai sallia vastaaminen omin sanoin. Ensimmäinen vaihtoehto on suositeltava, sillä se yksinkertaistaa vastausten analysointia (Robson, 1994). Avointen kysymysten synnyttämän laadullisen aineiston käsittely vaatii usein otoksen suuruudesta johtuen liikaa resursseja. Avoimia kysymyksiä kannattaa käyttää silloin, kun kysymyksen kohteena olevasta asiasta ei ole esimerkiksi teoriaan pohjautuvaa yleisesti hyväksyttävää luokitusta (Järvinen & Järvinen, 2000). Myös henkilökohtaisen ja paheksuttavan käyttäytymisen tiedusteleminen avoimilla kysymyksillä on suositeltavaa (Robson, 1994).

Kysymyksen vaihtoehdot voivat noudattaa nominaali-, ordinaali-, intervalli- tai suhdeasteikkoa. Nominaaliasteikon luokat eroavat laadullisesti, ordinaaliasteikossa luokat ovat järjestyssuhteessa, intervalliasteikossa järjestyksessä peräkkäisten luokkien erot ovat samat ja suhdeasteikossa voidaan tämän lisäksi määrittää nollakohta.

Ordinaali- tai intervalliasteikkoa noudattavia kysymyssarjoja käytetään usein kyselyiden yhteydessä, ja niihin viitataan termillä *asteikko* (*scale*). Asteikot poikkeavat toisistaan vaihtoehtojen määrän ja asteikon laadinnassa käytetyn menetelmän mukaan (Robson, 1994). Tunnettuja asteikkoja ovat *Likert-asteikko* ja *semanttinen differentiaali* (*semantic differential*)-asteikko.

Likert-asteikollinen lomake koostuu väittämistä, joita arvioidaan useimmiten viisiportaisella asteikolla. Taulukossa 3 on SUS-lomake<sup>1</sup>, jonka väittämiin vastataan Likert-asteikolla. Likert-asteikon huippu ja pienin arvo merkitsevät vastakkaisia mielipiteitä, kuten ”täysin eri mieltä” ja ”täysin samaa mieltä”.

Likert-asteikkoa noudattavassa lomakkeessa jokainen vastaus on tuloksen kannalta yhtä merkitsevä. Tämän vuoksi lomakkeella kerättyjen vastausten pisteyttäminen ja tulosten laskeminen on helppoa, sillä pistemäärä voidaan laskea summaamalla yksittäisten kysymysten vastaukset. SUS-lomaketta käytettäessä parittomien kysymysten pistemäärä lasketaan vähentämällä annetun vastauksen pistemäärästä yksi piste. Parillisten kysymysten pistemäärä saadaan vähentämällä vastaus viidestä, jolloin vahvasti eri mieltä oleminen tuottaa neljä pistettä ja vahvasti saman mielisestä vastauksesta ei lisätä pisteitä lainkaan. Saadut pistemäärät lasketaan yhteen ja summa kerrotaan luvulla 2,5. Likert-asteikon kehittäminen vaatii suuren väittämämäärän testaamista suurella vastaajamäärällä ja sopivien väittämien valitsemista tilastollisten testien perusteella.

---

<sup>1</sup> © Digital Equipment Corporation, 1986.

**Taulukko 3:** SUS-lomakkeen kysymykset (Brooke, 1996).

	Vahvasti eri mieltä				Vahvasti samaa mieltä
1. Luulen, että haluaisin käyttää tätä järjestelmää usein.	1	2	3	4	5
2. Minusta järjestelmä oli turhan monimutkainen.	1	2	3	4	5
3. Mielestäni järjestelmää oli helppo käyttää.	1	2	3	4	5
4. Luulen, että tarvitsisin teknistä tukea järjestelmän käytössä.	1	2	3	4	5
5. Mielestäni järjestelmän eri toiminnot oli yhdistetty hyväksi kokonaisuudeksi.	1	2	3	4	5
6. Mielestäni tässä järjestelmässä oli liikaa epäjohdonmukaisuutta.	1	2	3	4	5
7. Uskon että useimmat ihmiset oppivat nopeasti käyttämään tätä järjestelmää.	1	2	3	4	5
8. Minusta järjestelmän käyttäminen tuntui vaivalloiselta.	1	2	3	4	5
9. Tunsin itseni todella varmaksi käyttäessäni järjestelmää.	1	2	3	4	5
10. Minun täytyi oppia monia asioita ennen kuin pääsin alkuun järjestelmän käytössä.	1	2	3	4	5

Semanttinen differentiaali -asteikko koostuu toisilleen vastakkaisista adjektiiveista (katso esimerkiksi QUIS-lomaketta taulukossa 2). Vastaajia pyydetään merkitsemään esimerkiksi rastilla kohta, johon tietty asia heidän mielestään sijoittuu. Myös tämän asteikon tulokset saadaan luokittelemalla annetut arviot esimerkiksi yhden ja seitsemän välille ja laskemalla pistemäärät yhteen.

Vapaamuotoisten vastausten analyysi edellyttää vastausten luokittelua. Tämä luokittelu kuitenkin hävittää osan kerätystä tiedosta, ja luokittelun tarkkuus riippuu myös käytettävästä analyysimenetelmästä. Tilastollinen analyysi edellyttää usein pienempää kategorioiden määrää, mikä merkitsee suurempaa informaation hävikkiä.

Lomaketta laadittaessa voidaan avoimilla kysymyksillä hakea vastausvaihtoehtoja, joita käytetään lomakkeen myöhemmissä versioissa. Tämän menettelyn avulla lomakkeesta tulee vähemmän vaivaa vaativa niin täyttämisen kuin vastausten kirjaamisenkin kannalta..

#### 2.4.2. Kysymysten edellyttämä päättely

Hufnagel ja Conca (1991) jakavat kysymykset kolmeen kategoriaan: luokittelua, selittämistä ja ennustamista vaativiin. He käsittelevät artikkelissaan erityyppisiin kysymyksiin vastaamisen edellyttämää päättelyä, kysymysten virhelähteitä ja virheiden välttämistä.

Ensimmäisen kategorian kysymysten vaatimalla luokittelulla tarkoitetaan järjestelmän sijoittamista annetuilla ulottuvuuksilla sopiviin kohtiin. Tämä edellyttää vertaamista muihin järjestelmiin sekä yhtäläisyyksien että erojen osalta (taulukko 4). Esimerkiksi taulukossa 2 esitetyt QUIS-lomakkeen kysymykset edellyttävät muun muassa järjestelmien tyydyttävyyden ja joustavuuden vertailua.

**Taulukko 4:** Luokitteluun liittyvät tehtävät ja virhelähteet (Hufnagel & Conca, 1991).

Tehtävä	Vastausten virhelähteitä	Ehdotetut ratkaisut
henkilön tai objektin ominaisuuksien erittely	moniselitteinen kysymys tai konteksti	<ul style="list-style-type: none"><li>• Kuvaa konteksti ohjeistuksessa.</li><li>• Määrittele mahdollisesti epäselvät käsitteet.</li><li>• Käytä vertailevia kysymyksiä.</li><li>• Testaa lomaketta osalla vastaajista.</li></ul>
	sanalliset frekvenssiasteikot	<ul style="list-style-type: none"><li>• Käytä prosenttiasteikkoja.</li></ul>
	aiemman kokemuksen merkitys ja mieleenpalautuminen	<ul style="list-style-type: none"><li>• Ohjeista vastaajat arvioimaan yleisintä tilannetta.</li></ul>
	vähän aikaa sitten hankitun kokemuksen keskeisyys	<ul style="list-style-type: none"><li>• Ilmoita ajanjakso, jota vastaajien tulee tarkastella (esim. viime viikko).</li></ul>
	tietämyksen puute	<ul style="list-style-type: none"><li>• Anna mahdollisuus epävarmoin vastauksiin ("en tiedä", "en ole varma").</li></ul>
	asenteen laimeus	<ul style="list-style-type: none"><li>• Kerää tietoa asenteiden ja mielipiteiden voimakkuudesta.</li></ul>

Taulukossa 4 kuvatut luokittelukysymysten ongelmat johtuvat Hufnagelin ja Concan (1991) mukaan vastausten kontekstisidonnaisuudesta ja vihjeistä, joita aiempi kokemus, ympäristö, tutkija, kyselyn ohjeistus ja itse kysymys tarjoavat. Kysymyksen kontekstin on oltava yksikäsitteinen, sillä epämääräiset sanamuodot ja tuntematon terminologia voivat aiheuttaa sekä satunnaisia virheitä että harhoja vastauksiin. Satunnaisten virheiden vaikutuksesta vastaukset poikkeavat vastaajien todellisista mielipiteistä ja arvioista satunnaisesti, harhat aiheuttavat johdonmukaisen poikkeaman. Harhan vaikutuksesta osallistujat siis vastaavat todellista mielipidettään myönteisemmin tai kielteisemmin. Tällaiset kysymykset voidaan tunnistaa tulkitsemalla kysymysten välisiä riippuvuuksia tilastollisesti.

Jonkin tapahtuman toteutumistiheyden (frekvenssin) arviointi on myöskin virhealtista. On havaittu, että vaihtoehtojen suhteellinen sijainti vastausasteikolla vaikuttaa vastaukseen enemmän kuin vastausvaihtoehdon sisältö (Hufnagel & Conca, 1991). Parempaan tulokseen päästään käyttämällä sanallisten vaihtoehtojen sijasta numeerista asteikkoa.

Käytettävyydestä yhteydessä suoritettujen kyselyiden kannalta erityisen ongelmallista on ero juuri kohdattujen ja aiempien kokemusten muistamisessa. Viimeaikaiset ongelmat palautuvat helpommin mieleen ja vaikuttavat vastauksiin enemmän kuin pidemmältä ajanjaksolta hankitut käyttökokemukset.

Ongelmaa lisää tunnesävytyneiden, positiivisten tai negatiivisten, muistojen dominoiva vaikutus päättelyyn. Tiettyjen ongelmien esiintuontiin tähtäävä käytettävyydestä vaikuttaa sen jälkeen suoritettavaan arviointiin, vaikka käyttäjällä olisi pitkäaikainen aiempi käyttökokemus. Samat ongelmat koskevat toki myös muissa yhteyksissä tehtävää käytetyn järjestelmän arviota.

Taulukko 5 esittää selitysten muodostamista vaativiin kysymyksiin liittyvät virhelähteet ja näiden välttämiseksi ehdotetut toimenpiteet. Selitysten muodostaminen on tarpeen vastattaessa kysymyksiin, jotka edellyttävät arvioita järjestelmän käytön seurauksista. Tällaista päättelyä edellyttävä väittämä on esimerkiksi: ”[Tietyn teknologian] käyttö lisää tuottavuuttani (Hufnagel & Conca, 1991).”

Selitysten muodostaminen edellyttää Hufnagelin ja Concan mukaan ensin objektin luokittelua ja tämän jälkeen syy- ja seuraussuhteiden tunnistamista. Selitysten muodostamiseen siis liittyy samoja ongelmia kuin luokittelukysymyksiin. Lisäksi syiden ja seurausten tunnistamisen ongelmina ovat onnistumisten ja epäonnistumisten syiden vinoutunut tunnistaminen, aikaisemmat odotukset ja ryhmien väliset eroavaisuudet.

**Taulukko 5:** Selittämiseen liittyvät tehtävät ja virhelähteet (Hufnagel & Conca, 1991).

Tehtävä	Vastausten virhelähteitä	Ehdotetut ratkaisut
henkilön tai objektin ominaisuuksien erittely	luokitteluun liittyvät virhelähteet (taulukko 4)	
syy- ja seuraussuhteiden arvioiminen havaintojen perusteella	vastaajat eroavat toisistaan: <ul style="list-style-type: none"> <li>• suoriutumisessaan</li> <li>• ennakko-odotuksissaan</li> <li>• asiantuntemukseltaan</li> <li>• kokemukseltaan</li> <li>• vastuuntunteeltaan</li> <li>• vapaaehtoisuudeltaan.</li> </ul>	<ul style="list-style-type: none"> <li>• Tunnista vastaajaryhmien väliset erot lomaketta suunniteltaessa. Liitä kyselyyn kysymyksiä, joiden avulla vastaajat voidaan jakaa eri ryhmiin näiden erojen merkittävyyden arvioimiseksi.</li> </ul>

Järjestelmän ongelmiin keskittyvän lomakkeen kysymykset ohjaavat etsimään epäonnistumisen syytä ensisijaisesti järjestelmästä. Vastauksiin vaikuttaa myös osallistujien suoriutuminen ja vapaaehtoisuus. Hufnagel ja Conca (1991) viittaavat aiempiin tutkimuksiin, joiden mukaan hyvin suoriutuvat johtajat ja järjestelmää vapaaehtoisesti käyttävät arvioivat kohtaan omaansa ongelmat useammin itsestään johtuviksi kuin huonosti suoriutuvat johtajat ja järjestelmän käyttäjiksi pakotetut yksilöt, jotka näkevät ongelmien syynä useammin itse järjestelmän. Näistä vinoutuneista arvioista johtuvaa vastausten harhaisuutta voi kontrolloida tarkkailemalla eroja toisistaan poikkeavien osallistujaryhmien vastauksissa. Vastaajien jakaminen ryhmiin edellyttää kuitenkin ylimääräisen taustatiedon keräämistä.

Vaativin kysymystyyppi edellyttää, että lomakkeen täyttäjät arvioi tulevaisuuden mieltymyksiään tai toimintaansa. Hufnagel ja Conca (1991) lainaavat Mathiesonin (1991) tutkimuksesta seuraavan kysymyksen esimerkkinä ennustamisen edellyttämisestä: ”Tehtävän suorittaminen [taulukkolaskentaohjelmalla] vaatisi minulta paljon (enemmän/vähemmän) aikaa kuin saman tehtävän suorittaminen taskulaskimellani.” Ennustaminen edellyttää selitystä edellyttävien kysymysten vaatiman syy- ja seuraussuhteiden tunnistamisen lisäksi näiden suhteiden pysyvyyden arvioimista.

Viimeaikaisen kokemuksen merkitys korostuu vastauksissa enemmän kuin aikaisempi kokemustieto. Lisäksi erityisesti päätelmien tekeminen rajoitetusta tiedon määrästä on vaikeaa. Tämä vaikeus koskee erityisesti survey-tyyppistä markkinakartoitusta, joka kerää jonkin järjestelmän potentiaalisten käyttäjien mielipiteitä. Taulukko 6 esittää tiivistelmän ennustamisen vaatimuksista.

**Taulukko 6:** Ennustamiseen liittyvät tehtävät ja virhelähteet (Hufnagel & Conca, 1991).

Tehtävä	Vastausten virhelähteitä	Ehdotetut ratkaisut
henkilön tai -objektin ominaisuuksien erittely	luokitteluun liittyvät virhelähteet (taulukko 4)	
syy- ja seuraussuhteiden arvioiminen havaintojen perusteella	selittämiseen liittyvät virhelähteet (taulukko 5)	
havaintujen suhteiden pysyvyyden arvioiminen	vastaajalle annettu puolueellinen ennakkotieto, esimerkiksi arvioinnin kohteen esittely tai mahdollisuus kokeilla sitä	<ul style="list-style-type: none"> <li>• Ohjeista vastaajaa ottamaan vastaamista edeltäneen esityksen tms. lisäksi myös vastakkainen näkökulma huomioon.</li> </ul>
	oman tulevaisuutta koskevan tärkeysjärjestyksen päättäminen	<ul style="list-style-type: none"> <li>• Käytä luokittelukysymyksiä, jotka pakottavat harkitsemaan vaihtoehtojen keskinäistä edullisuutta.</li> </ul>
	tulevaisuudessa vallitsevan oman tärkeysjärjestyksen epävarmuus	<ul style="list-style-type: none"> <li>• Pyydä arvio vastauksen luotettavuudesta.</li> </ul>

Vaikuttaisi siis siltä, että luokittelukysymyksiä kannattaa suosia. Tähän ryhmään kuuluviin kysymyksiin liittyy vähiten potentiaalisia ongelmia. Kysymysten asettelu riippuu kuitenkin tutkimusasetelmasta ja tutkimuksen aiheesta, eikä muiden kysymystyyppien käyttöä voida kokonaan välttää.

Ensi silmäyksellä voi vaikuttaa siltä, että Hufnagelin ja Concan käsittelystä on unohtunut tulevaisuuden ennustamisen vastakohta: menneisyyden kuvaaminen. Tämä on kuitenkin kahden ensimmäisen kategorian taustalla, sillä kaikkiin edellä kuvattuihin kysymystyypeihin vastattaessa on turvaututtava muistiin ja aiempia kokemuksia koskeviin muistoihin. Arvatenkin luokittelukysymyksiä koskevana ongelmana mainittu tunneperäisen kokemuksen vaikutus on sitä korostuneempi, mitä kauemmas ajallisesti vastaajan on muisteltava.

### 2.4.3. Kysymysjärjestys

Hufnagelin ja Concan (1991) mainitsema kontekstin merkitys kyselyiden tuloksissa näkyy myös kysymysjärjestyksen vaikutuksessa vastauksiin. Babbien (1973) esimerkin mukaan kyselyssä ensiksi pyydetty arvio osallistujan yleisestä uskonnollisuudesta vaikuttaa seuraaviin tietyistä uskonnollisuuden piirteistä annettuihin vastauksiin, sillä vastaajat pyrkivät olemaan johdonmukaisia.

Joissakin tutkimuksissa käytetty kysymysten satunnaistaminen ei Babbien mukaan ole suositeltavaa, sillä sekava joukko kysymyksiä vaikuttaa vastaajista hyödyttömältä ja vaatii ajatusten jatkuvaa siirtämistä aiheesta toiseen. Satunnaistaminen ei myöskään poista kontekstin merkityksen ongelmaa, ainoastaan piilottaa sen, jolloin sitä ei myöskään voida kontrolloida. Jos kysymysjärjestyksen vaikutus koetaan erityisen tärkeäksi, on syytä laatia useita kysymysjärjestykseltään vaihtelevia lomakkeita ja testata kysymysjärjestyksen vaikutus pienellä osallistujista muodostetulla ryhmällä.

Mikään ei tietenkään takaa, että kyselyyn osallistuja lukee lomakkeen kysymykset järjestyksessä. Vastaaaja voi esimerkiksi ohittaa vaikeat kysymykset palatakseen niihin myöhemmin. Jos järjestys on kriittisen tärkeä, on syytä toteuttaa kyselyn sijasta strukturoitu haastattelu, jossa kysymykset voidaan esittää juuri toivotussa järjestyksessä. Haastattelusta ja siihen liittyvistä ongelmista kertoo tarkemmin raportin luku 3 (Vuorela).

Robsonin (1994) mukaan ainoa tutkimuksen tuottama suositus kysymysjärjestyksestä on yleisempien kysymysten sijoittaminen ennen tarkkoja kysymyksiä. Babbie (1973) suosittelee tämän lisäksi kuitenkin sijoittamaan mielenkiinnottomat kysymykset, joilla vastaajat luokittelevat (demografisiin) ryhmiin, kyselyn loppuun. Tällaisia kysymyksiä ovat esimerkiksi sukupuoli ja ikä. Kiinnostavat kysymykset lomakkeen alussa luultavasti herättävätkin vastaajan mielenkiinnon ja kannustavat aloittamaan lomakkeen täytön.

### 2.4.4. Kyselylomakkeen arviointi

Kyselylomakkeella saatujen tulosten luotettavuus voidaan jakaa *reliabiliteettiin* ja *validiteettiin*. Heikko reliabiliteetti merkitsee, että satunnainen vaihtelu vaikuttaa suuresti saatuihin vastauksiin, minkä vuoksi tutkimuksen tuloksetkin ovat satunnaisia. Tutkimus ei siis ole toistettavissa, eikä se tämän vuoksi ole myöskään luotettava (Hufnagel & Conca, 1991).

Reliabiliteettia voidaan arvioida teettämällä kysely samassa muodossa ja samoilla vastaajilla useita kertoja ja vertaamalla tuloksia. Tulosten vastaavuus yleisesti merkitsee hyvää reliabiliteettia, mutta osallistujat voivat myös tietoisesti pyrkiä vastaustensa johdonmukaisuuteen. Toinen menetelmä reliabiliteetin arvioimiseen on kyselyn toistaminen lomakkeella, josta esimerkiksi kysymysten järjestystä tai sanamuotoja on muutettu. Tällöinkin tavoitteena on annettujen vastausten johdonmukaisuus. Lomakkeen sisäistä johdonmukaisuutta voidaan arvioida vertaamalla vastauksia, joita on annettu saman käsitteen eri osa-alueita mittaaviin kysymyksiin. (Kitchenham & Pflieger, 2002c)

Reliabiliteettia voidaan parantaa kasvattamalla otoksen kokoa. Tämä on kuitenkin Hufnagelin ja Concan (1991) mukaan usein kalliimpaa ja vaikeampaa kuin lomakkeiden huolellinen suunnittelu, jolla virheitä myös voidaan välttää.

Validiteetista voidaan edelleen erottaa useita eri tyyppisiä (Kitchenham & Pfleeger, 2002c; Robson, 1994). *Rakennevaliditeetti* (*construct validity*) kuvaa, kuinka hyvin menetelmät mittaavat sitä ominaisuutta mitä niiden tulisi mitata. Lomakkeen kysymysten muotoilu liittykin suoraan rakenteelliseen validiteettiin. Rakenteellista validiteettia voidaan arvioida vain teorian ja siihen liittyvien aiempien tutkimusten pohjalta. Tilastollisia tai muitakaan testejä ei voida käyttää arvioinnissa (Hufnagel & Conca, 1991; Kitchenham & Pfleeger, 2002c).

*Sisäinen validiteetti* (Robson, 1994) ei suoranaisesti liity kyselylomakkeisiin mutta vaikuttaa toki tutkimusasetelmaan yleisesti. Sisäinen validiteetti kuvaa erityisesti löydettyjen syy- ja seuraussuhteiden luotettavuutta. Toisin sanoen tutkimuksen sisäinen validiteetti on hyvä, jos muutokset riippuvassa muuttujassa ovat aiheutuneet riippumattoman muuttujan manipuloinnista. Sisäinen validiteetti onkin erityisen keskeinen kokeellisessa tutkimuksessa.

*Ulkoinen validiteetti* (Robson, 1994) kertoo, kuinka luotettavia tuloksista tehtävät yleistykset ovat. Yleistäminen merkitsee tässä yhteydessä tulosten tulkitsemista laajempaa joukkoa koskeviksi, esimerkiksi tietyn opiskelijamäärän suorittaman kokeen tulosten yleistämistä kaikkia oppilaita koskeviksi. Yleistämisen kannalta tärkeintä on kyselyyn vastaavan joukon onnistunut valinta.

*Sisältövaliditeetti* on subjektiivinen arvio lomakkeen soveltuvuudesta tutkittavan asian mittaamiseen. Arvion muodostavat aihealueen asiantuntijat, jotka voivat haluta käyttää kehitettävää lomaketta tutkimuksissaan. Sisältövaliditeetin arviointi on erityisen hyödyllistä kehitettäessä lomaketta arvioimaan sellaista ominaisuutta, jota mittaavia aiemmin kehitettyjä lomakkeita ei ole olemassa. (Kitchenham & Pfleeger, 2002c)

*Kriteerivaliditeetti* kertoo, kuinka hyvin lomake vastaa toisella tiedonkeruumenetelmällä saatuja tuloksia tai noudattaa jotakin muuta kriteeriä. Parhaiten se on arvioitavissa, kun lomakkeella saadut tulokset voidaan liittää tulevaisuuden ennustamiseen. Esimerkiksi koulumenestystä ennustavan lomakkeen tuloksia voidaan suoraan verrata todelliseen koulumenestykseen. (Kitchenham & Pfleeger, 2002c)

Muista validiteetin tyypeistä kyselylomakkeisiin liittyvät myös *konvergoiva validiteetti* ja *erotteleva validiteetti*. (Kitchenham & Pfleeger, 2002c) Näiden arviointi vaatii useimmissa tapauksissa monien vuosien kokemuksen kyselylomakkeiden kehittämisestä, eikä niitä tarkastella tässä enempää.

## 2.5. Kyselyn toteuttaminen

Kyselyn varsinaiseen toteutukseen liittyy monia huomioon otettavia seikkoja. Vastaajien valinta, lomakkeiden toimittamiseen valittu menetelmä ja kyselyn suorittaminen erilaisissa konteksteissa voivat kaikki vaikuttaa tuloksiin ja tulosten analyysiin.

### 2.5.1. Otoksen valitseminen

Yleensä kaikki tutkimuksen kannalta merkittävät henkilöt eivät voi osallistua tutkimukseen. Kohderyhmän koko on useimmiten hyvin suuri, vaikkakin poikkeuksia löytyy. Suuresta ryhmästä on valittava ne osallistujat eli *otos*, joka vastaa kyselyyn.

Otoksen valinta eli *otanta* on tarpeen tietojen keräämisen ja analysoinnin vaatimien resurssien pienentämiseksi. Hyvin suuren otoksen vaatima aika saattaisi olla niin suuri, että esimerkiksi markkinakartoituksen tulokset olisivat saatavilla vasta toteutusvaiheen jälkeen. Otanta mahdollistaa myös vastaamatta jättäneiden henkilöiden kartoittamisen, mikä edesauttaa otoksen edustavuuden määrittelyä ja auttaa arvioimaan vastaajien valikoitumisen merkitystä saatujen tulosten kannalta. (Kitchenham & Pfleeger, 2002d)

*Otos* valitaan *populaatiosta*. Populaatio merkitsee kaikkia tarkasteltavia tapauksia, kuten kaikkia potentiaalisia tuotteen tai ohjelmiston käyttäjiä. *Otantamenetelmät* voidaan karkeasti jakaa satunnaisiin ja ei-satunnaisiin.

Satunnaiset menetelmät, kuten yksinkertainen satunnaisotanta, systemaattinen otanta ja ryväotanta, tähtäävät edustavan otoksen muodostamiseen populaatiosta. *Edustava otos* mahdollistaa koko populaatiota koskevien päätelmien muodostamisen. Satunnaisen otoksen muodostaminen onkin tärkeää erityisesti kokeellisessa tutkimuksessa, jossa tavoitellaan ulkoista validiteettia.

Ei-satunnainen otanta on usein helpommin järjestettävissä ja vaatii vähemmän resursseja kuin satunnaisen otoksen muodostaminen. Ei-satunnainen otos onkin käytännöllisempi, jos tuloksia ei ole tarkoitus arvioida tilastollisesti. Yleistettävät tulokset kuitenkin edellyttävät jonkinlaista analyysia, minkä vuoksi menetelmä sopii lähinnä tapaustutkimuksiin.

Kysely mahdollistaa laajan otoksen käytön, mikä edesauttaa tulosten yleistettävyyttä. Otoksen edustavuus on kuitenkin usein kyseenalainen. Esimerkiksi Helsingin Sanomien Nyt-liitteen verkkosivulla (Nyt.fi, 2004) esitetyn kyselyn tuloksia ei voi yleistää kaikkia suomalaisia koskeviksi, sillä kyselyyn vastanneet eivät luultavasti muodosta kaikkia suomalaisia edustavaa joukkoa.

Laajalle ihmisjoukolle postitetun kyselylomakkeen palauttaneet ja kyselyyn vastaamatta jättäneet eroavat luultavasti muussakin kuin viitseliäisyydessä. Vastaajien valikoituminen ei ole näin suuri ongelma, kun lomake on henkilökohtaisesti toimitettu jokaiselle osallistujalle, esimerkiksi laboratoriotutkimuksen yhteydessä.

Kitchenham ja Pfleeger (2002a) huomauttavat, että otoksen koon merkitys kyselyn onnistumisen kannalta usein yliarvioidaan. Käyttämällä pienempää otosta osa kyselylle varatuista resursseista voidaan käyttää vastaamattomuuden syiden selvittämiseen. Näiden syiden ymmärtäminen on tärkeää paitsi aineiston analyysissä, myös kyselystrategian, kyselylomakkeen ja kyselyn suorittamisen kehittämiseksi.

Sopivan otoksen muodostamista ja määrittelyä ei tässä esityksessä käsitellä tarkemmin. Etenkin kokeellista tutkimusta tai kyselyä harkittaessa kannattaa perehtyä tämän raportin luvun 18 (Anttonen) lisäksi tilastollisiin menetelmiin liittyvään kirjallisuuteen.

## 2.5.2. Lomakkeiden toimittaminen vastaajille

Tutkimuksessa käytettävän otoksen koko määrää pitkälti, millaista menetelmää kyselyn suorittamisessa kannattaa käyttää. Useimmiten lomaketta ei voida toimittaa jokaiselle vastaajalle henkilökohtaisesti, poikkeuksena laboratoriotutkimuksen yhteydessä jaettava taustatietolomake. Kun otoskoko on suuri, kyselylomakkeet toimitetaan usein postittamalla tai sähköisesti. Postitettujen vastausten määrää voi kasvattaa kyselyn mukaan liitettyllä kirjekuorella, johon palautusosoite on valmiiksi merkitty ja jonka postimaksu on maksettu (Nielsen, 1997).

Sähköisen kyselyn etuna postitettuun verrattuna on kustannustehokkuus, sillä vastausten syöttäminen analysoinnissa käytettävään sovellukseen voidaan automatisoida ja kyselystä ei aiheudu postikuluja (Punter et al., 2003). Sähköinen kysely voidaan toimittaa sähköpostin välityksellä tai verkkosivulla. Sähköpostin liitetiedostoihin ohjeistetaan kuitenkin suhtautumaan varauksella, mikä saattaa vähentää vastaajien määrää. Sähköpostilla annettujen vastausten muotoa on myös vaikeampi kontrolloida, mikä vaikeuttaa automaattista käsittelyä.

Jos kyselyyn vastaajat päätetään populaatiosta muodostetun otoksen perusteella, mahdollistaa kyselyn sähköinen muoto helpomman vastanneiden ja vastaamattomien seurannan. Huonosti laaditut kysymykset voidaan tunnistaa tarkkailemalla, mitkä kysymykset aiheuttavat vastaamisesta luopumisen (Punter et al., 2003). Muistutusten lähettäminen sähköisessä muodossa on myös halvempaa ja helpompaa. Jos kyselyyn vastaajat valikoituvat oman kiinnostuksensa perusteella eli hakeutuvat kyselyn sisältävälle verkkosivulle tutkijoiden ilmoitusten houkuttelemina, muodostuvat kyselyn tulokset todennäköisesti innokkaampien osallistujien vastauksista (Nielsen, 1997).

GVU:n järjestämään kyselyyn (GVU's 10th WWW user survey, 1998) toivottiin vastauksia Internetin kaikilta käyttäjäryhmiltä. Kyselystä tiedotettiin Internetin uutisryhmissä, tunnetuilla verkkosivustoilla, monilla verkkosivuilla esiintyvissä mainoksissa ja julkisissa viestimissä,

kuten päivälehdissä ja ammattilehdissä. Tämän toivottiin takaavan vastaajajoukon heterogeenisyys ja edustavuus. Muihin satunnaista otantaa käyttäneisiin internetin käyttöä koskeviin tutkimuksiin verrattuna kyselyn vastaajat olivat kuitenkin kokeneempia, taitavampia ja intensiivisempiä Internetin käyttäjiä. Tutkijoiden mukaan tämä oli joidenkin tutkimuksen tavoitteiden kannalta jopa toivottavaa.

Verkkosivustoilla olevat sivuston arviointilomakkeet voivat antaa tarkempia tuloksia kuin erilliset paperilomakkeet. Rootin ja Draperin (1983) tutkimuksen mukaan käyttäjät antavat ongelmistaan tarkemman kuvauksen, jos he käyttävät arvioitavaa järjestelmää juuri ennen lomakkeen täyttämistä. Nielsenin (1997) mukaan tämä tukee sähköisten lomakkeiden käyttöä verkkosivujen arvioinnissa. Verkkosivustolla on myös mahdollista automaattisesti kerätä tietoa käyttäjän toiminnasta sivustolla ja yhdistää kerätyt tiedot lomakkeen vastauksiin.

Verkkolomakkeen käyttö säästää tutkijalle koituvien kustannusten lisäksi myös vastaajien aikaa (Punter et al., 2003). Erityisesti motorisesti vammaiset tai näkövammaiset voivat käyttää olemassa olevia apuvälineitä, kuten ruudunlukijoita, verkkolomakkeen täyttämiseksi. Heille paperisen lomakkeen käsittely saattaisi olla vaikeaa, jopa mahdotonta. Toki verkkokysely edellyttää käyttäjiltä jonkin verran kokemusta tietokoneiden ja Internetin käytöstä, minkä vuoksi se ei aina sovellu ainoaksi tiedonkeruumenetelmäksi. Ongelma ei ole niin suuri tutkittaessa verkkosivun varsinaisia käyttäjiä, koska kohderyhmän voidaan useimmiten olettaa hallitsevan tarvittavat perustaidot. Suurempi ongelma on vastaajien kokemaa anonymiteetin puute (Shannon et al., 2002). Sähköiset lomakkeet sisältävät useimmiten jonkin tunnustetiedon, jolla vastaamattomuutta voidaan tarkkailla.

Tutkijalta verkkokyselyn järjestäminen edellyttää jonkin verran teknistä osaamista, joskin joitakin valmiita työkaluja kyselyn laatimiseen ja aineiston keräämiseen on olemassa (Globalpark, 2004; Logisoft, 2004) ja näitä esitellään erityisesti tämän raportin luvussa 12 (Heimonen). Verkkokyselyssä olisi hyvä olla myös mahdollisuus vastaamisen keskeyttämiseen ja keskeytetyn kyselyn jatkamiseen. Muita toivottuja ominaisuuksia ovat kyselyssä edistymisen osoittaminen vastaajalle, jotta tämä ei turhaudu, ja vastausten automaattinen seuranta. (Punter et al., 2003)

Vastausten määrän kasvattamiseksi Robson (1994) antaa 26 erilaista suositusta. Robsonin suosituksista osa soveltuu sekä postitettujen että sähköisten kyselyjen suorittamiseen (taulukko 7).

**Taulukko 7:** Kymmenen suositusta kyselyyn vastaajien määrän kasvattamiseksi Robsonia (1994) mukaillen.

- 
1. Lomakkeen täytyy näyttää helposti täytettävältä.
  2. Sanamuotojen tulee olla selkeitä.
  3. Järjestä lomakkeen sisältö huolellisesti. Esimerkiksi vältä lomakkeen alussa kysymyksiä, jotka vihjaavat, ettei kysely kosketa osaa vastaajista.
  4. Väritys voi selkeyttää lomakkeen rakennetta.
  5. Toista kyselyn täyttöohjeet, jos sekaannus on mahdollista.
  6. Ensimmäisten kysymysten pitää olla kiinnostavia, seuraavat voivat käsitellä vaikeampia asioita.
  7. Kyselyn sanamuodot ovat tärkeitä. Lomakkeen testaaminen ennen kyselyä on ehdottoman tärkeää.
  8. Tutkimuksen tarkoitus ja tarkoitus pitää tuoda esiin. Vastaajat on vakuutettava kyselyn luottamuksellisuudesta, vastausten seurantaan tarkoitettujen menetelmien (kuten tunnusmerkkijonojen) käyttö on perusteltava.
  9. Muistuttaessasi kyselystä älä vihjaa vastaamattomuuden olevan yleistä.
  10. Kolmen muistutuksen lähettämistä suositellaan yleisesti. Ne voivat kasvattaa vastausten lukumäärää (postikyselyssä) kolmanneksella.
-

### 2.5.3. Lomakkeen täyttöhetken merkitys

Teague ja muut (2001) tutkivat käyttäjien antamia arvioita tehtävän helppoudesta ja miellyttävyydestä käytettävyydestessään. Tutkimuksessa käyttäjien tuli arvioida tehtävän helppoutta ja miellyttävyyttä joko vastaamalla tehtävän suorituksen aikana säännöllisin väliajoin esitettyihin kysymyksiin tai vastaamalla kerran välittömästi tehtävän suorituksen jälkeen. Heidän tulostensa mukaan tehtävien jälkeen annetut arviot olivat positiivisempia kuin niiden aikana muodostetut.

Tulos selittyy Hufnagelin ja Concan (1991) esittämällä vastausten kontekstisidonnaisuudella. Tehtävän suorituksen aikana kohdatut ongelmat ovat tuoreemmassa muistissa, jos arvio annetaan tehtävän aikana. Kysymysten esittäminen tehtävän aikana myös ohjaa käyttäjän tarkkaavuuden kysymyksen aiheen kannalta merkittäviin asioihin, esimerkiksi kohdattuihin helppouteen ja miellyttävyyteen liittyviin ongelmiin. Tätä tukee myös Teaguen ja muiden (2001) analyysi, jonka mukaan käyttäjät antoivat yksityiskohtaisemman kuvauksen ongelmasta, jos he arvioivat tehtävän sitä suorittaessaan.

Myös muussa tutkimuksessa kuin käytettävyydestä arvioinnin suorittaminen ja kyselylomakkeen täyttäminen erilaisissa konteksteissa on tuloksiin merkittävästi vaikuttava tekijä. Esimerkiksi ohjelmiston käyttökokemuksia mitattaessa osa vastaajista saattaa vastata pelkän muistin varassa, kun taas osa saattaa täyttää lomakkeen käyttäessään järjestelmää muistinsa virkistämiseksi.

Rootin ja Draperin (1983) tutkimuksessa osallistujia pyydettiin arvioimaan editorin toimintojen ongelmallisuutta. Vastaajat, jotka pystyivät kokeilemaan jonkin toiminnon tekemistä editorilla lomakkeen täytön aikana, raportoivat useampia ongelmia. Kuitenkaan esiin tulleiden ongelmien suhteellinen järjestys ei muuttunut. Toisin sanoen samat ongelmat tulivat esille, mutta ongelmat esiintyivät useammissa arvioissa, jos editori oli saatavilla.

Kuvattujen tutkimusten perusteella näyttäisi siis siltä, että jos tavoitteena on mahdollisimman monien ongelmien löytäminen ja yksityiskohtainen kuvaaminen, on kyselyn suorittaminen käytettävyydestin tehtävien tai vapaan käytön aikana kannattavaa. Tämä kuitenkin vaikuttaa tehtävistä suoriutumiseen ja tuotteen käyttöön, mikä voi haitata kokonaisvaikutelman muodostamista.

Jälleen on todettava, että tutkimuksen tarkoitus on merkittävin tekijä arvioinnin ajankohtaa määritettäessä. Tapaustutkimus, joka tähtää esimerkiksi tietyn ohjelmiston kaikkien ongelmien löytämiseen, hyötyy luultavasti käytön kanssa rinnakkain tehtävistä arvioista. Käytön aikana esitetyt kysymykset saattavat aiheuttaa vinoutuneita vastauksia, joiden objektiivinen vertailu muista järjestelmistä saatujen tulosten kanssa ei onnistu. Luotettavuus ja yleistettävyydet eivät silloin voi olla kyselyn tavoitteena, vaan tavoite on lähempänä: kyseisen ohjelmiston arvioinnissa.

### 2.5.4. Vastausten analysointi

Vastausten analysoinnin vaatima työmäärä riippuu sekä valitun otoksen koosta että aineiston laadusta. Laadullisen aineiston analysointi on työläämpää kuin määrällisen aineiston. Kyselylomakkeilla kerätyn aineiston tyyppi liittyy suoraan kysymystyyppiin: vaihtoehtokysymykset synnyttävät määrällistä aineistoa ja avoimet kysymykset laadullista aineistoa.

Kerätyn määrällisen aineiston analyysin vaatima työmäärä riippuu tutkimuksessa käytetyistä lomakkeista. Valmiiksi sähköisessä muodossa tai koneellisesti luettavissa oleva informaatio voidaan käsitellä suoraan. Kirjallisesti annettujen vastausten siirtäminen sähköiseen muotoon vaatii luonnollisesti enemmän resursseja.

Sopivien tilastollisten menetelmien soveltaminen kerättyyn aineistoon ei ole työlästä saatavilla olevien ohjelmistojen ansiosta, mutta menetelmien valinta ja tulosten arviointi vaatii asiaan perehtymistä. Mitä laajempi aineisto on kerätty, sitä helpompaa on löytää merkitseviä tuloksia, joiden pohjalla onkin sattuma oikean ilmiön sijasta (Kitchenham, 2003). Esimerkiksi 14 tilastollista testiä, joista jokaisen virhemarginaali on viisi prosenttia, tuottavat todennäköisesti (yli 51 % varmuudella) ainakin yhden tuloksen, jonka tilastollinen

merkitsevyys on täysin sattuman aiheuttama. Tulosten tilastollinen merkitsevyys ei myöskään takaa niiden käytännöllisyyttä; laaja otos ja suuri kysymysten määrä voi tuoda esiin ilmiön, jota ei voida hyödyntää.

Laadullisen aineiston analyysissä päästään suhteellisen pienellä työmäärällä, jos tutkimusta ohjannut viitekehys (teoria) on selkeä. Jos kysymysten avoimuus johtuu vain siitä, ettei kaikkia mahdollisia vaihtoehtoja tunneta (kuten vastaajien mahdollisia ammatteja), on luokittelu eri luokkiin (esimerkiksi palvelu-, tuotanto- ja muihin ammatteihin) melko helppoa, joskaan ei yksikäsitteistä. Sen sijaan uutta teoriaa luovan tutkimuksen analyysivaihe voi olla hyvinkin työläs.

## 2.6. Päätelmät

Kyselylomake on tiedonkeruun työkalu, joka ei pakota minkään tietyn tutkimusstrategian käyttöön eikä myöskään määrää tutkimuksen tavoitteita. Kyselyn soveltuvuutta tutkimukseen on kuitenkin syytä arvioida strategian ja tavoitteiden perusteella. Strategia ja tavoitteet ohjaavat myös lomakkeen ja kysymystyyppien valintaa.

Käytettävyyden arviointiin on olemassa lukuisia valmiita lomakkeita. SUMI-lomakkeen etuna ovat etenkin standardoidut tulokset, joiden avulla voidaan verrata arvioidun järjestelmän käytettävyyttä muiden markkinoilla olevien järjestelmien käytettävyyteen. Valmiiden lomakkeiden haittapuolena on niillä kerätyn tiedon yleinen luonne. Jonkin yksittäisen ominaisuuden tai piirteen arviointi edellyttää juuri tähän tarkoitukseen laadittavaa lomaketta tai valmiin lomakkeen muokkaamista. Lomakkeen muokkaaminen voi kuitenkin olla haitaksi tulosten luotettavuudelle, joka on eräs valmiiden lomakkeiden suurimmista eduista.

Oman lomakkeen suunnitteluun liittyy monia mahdollisia ongelmia ja huomioitavia seikkoja. Tutkimuskysymykset ohjaavat lomakkeen sisällön laatimista, mutta lomakkeen sanamuotojen ja kysymysjärjestyksen vaikutusta on usein vaikea ennakoida. Oman lomakkeen kehittäminen vaatiikin testauksen, luotettavuuden arvioinnin ja lomakkeen muokkaamisen toistamista useita kertoja. Jotta lomakkeen suunnittelussa ja arvioinnissa päästäisiin vähäisemmällä työmäärällä, oman lomakkeen pohjaksi kannattaa aina ottaa jokin valmis lomake tai ainakin tutustua useisiin erilaisiin lomakkeisiin.

Tutkimusstrategian ja tutkimuksen tavoitteiden vaikutus näkyy selvimmin kyselyn varsinaisen toteutustavan valinnassa. Tarkasti suunniteltu vastaajien valinta voi tutkimuksesta riippuen olla hyvin tärkeää tai suhteellisen merkityksetöntä. Lomakkeiden toimitustapa voi vaihdella, mikä vaikuttaa vastaajien valikoitumisen lisäksi myös kyselyn suorittamisessa ja vastausten analysoinnissa vaadittujen resurssien määrään.

Kysely soveltuu parhaiten tutkimuksiin, joissa ollaan kiinnostuneita suuren osallistujajoukon mielipiteistä ja muista subjektiivisista arvioista. Kyselylomakkeet eivät kuitenkaan sovellu kaikkiin tarkoituksiin; käytettävyytutkimuksessa niillä ei välttämättä saada yksityiskohtaista tietoa käyttöliittymän ongelmista, mutta niillä kerätty aineisto tukee hyvin muita käytettävyyden arviointimenetelmiä ja voi tarjota tukea erityisesti tulosten validointiin, sillä lomakkeiden avulla saadaan määrällistä aineistoa ja tilastollisia arvioita.



**Toni Vanhala**, Fil. yo. Opintoni alkoivat Tampereen yliopiston tietojenkäsittelytieteiden laitoksella vuonna 2000. Kesällä 2002 aloitin työskentelyn laitoksen TAUCHI-yksikössä, mistä lähtien olen toiminut tutkimusassistenttina Research Group for Emotions, Sociality, and Computing -tutkimusryhmässä. Tutkimukseni aihealueena ovat fysiologiset tunnevasteet ihminen-tietokone –vuorovaikutuksessa. Valmistun maisteriksi vuoden 2005 keväällä, minkä jälkeen jatkan opinnoissani kohti väitöskirjaa.

**Ohjaaja: Saila Ovaska**

**Opponentit: Raija Perälä ja Kimmo Vuorinen**

## Lähteet

- Babbie, E.R. (1973) *Survey Research Methods*. Wadsworth publishing.
- Bailey, J.E. & Pearson, S.W. (1983) Development of a tool for measuring and analyzing computer user satisfaction. *Management Science*, 29 (5), 530-545.
- Bernard, M., Lida, B., Riley, S., Hackler, T. & Janzen, K.A. (2002) Comparison of popular online fonts: Which size and type is best? *Usability News*, 4(1). Software usability research laboratory. Wichita state university.
- Brooke, J. (1996) SUS: A quick and dirty usability scale. In Jordan, P.W., Thomas, B., Weerdmeester, B.A. & McClelland, I.L. (Eds.) *Usability Evaluation in Industry*, 189–194. UK: Taylor and Francis. <http://www.usability.serco.com/trump/documents/Suschapt.doc>. (3.3.2004)
- Chin, J.P., Diehl, V.A. & Norman, K.L. (1988) Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI 88)*, 213–218.
- Churchill G.A. (1979) A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research* Vol. XVI, 64–73.
- Davis, F.D. (1993) User acceptance of information technology: system characteristics, user performance and behavioral impacts. *International Journal of Man-Machine Studies*, 38, 475–487.
- Doll, W.J. & Torkzadeh, G. (1988). The measurement of end-user computing satisfaction. *MIS Quarterly*, 12, 259–274.
- Gediga, G., Hamorg, K.-C. & Düntch, I. (1999) The IsoMetrics usability inventory: An operationalisation of ISO 9241/10. *Behaviour and Information Technology*, 18, 151–164. <http://people.freenet.de/gediga/bit99.pdf>. (3.3.2004)
- Globalpark. <http://www.globalpark.de/home/index.htm>. (4.3.2004)
- GVU's 10th WWW user survey. (1998)  
[http://www.cc.gatech.edu/gvu/user\\_surveys/survey-1998-10/](http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/) (4.3.2004)
- Hufnagel, E.M. & Conca, C. (1994) User response data: The potential for errors and biases. *Information Systems Research*, 5(1), 48–73.
- Ives, B., Olson, M.H. & Baroudi J.J. (1983) The measurement of user information satisfaction. *Communications of the ACM*, 26 (10), 785–793.
- Järvinen, P. & Järvinen, A. (2000) *Tutkimustyön metodeista*. Tampere: Opinpajan kirja.
- Kalbach, J. & Bosenick, T. (2003) Web page layout: a comparison between left- and right-justified site navigation menus. *Journal of Digital Information*, 4(1). <http://jodi.ecs.soton.ac.uk/Articles/v04/i01/Kalbach/> (19.8.2004).
- Keinonen, T. (1998) Usability of artifacts. One-dimensional usability—influence of usability on consumers' product preference. *UIAH publication A21*. <http://www2.uiah.fi/projects/metodi/158.htm>. (3.3.2004)
- Kirakowski, J. (1994) The use of questionnaire methods for usability assessment. Background notes on the SUMI questionnaire. <http://www.ucc.ie/hfrg/questionnaires/sumi/sumipapp.html>.
- Kitchenham, B.A. & Pfleeger, S.L. (2002a) Principles of survey research part 2: Designing a survey. *Software Engineering Notes*, 27(1), 18–20.
- Kitchenham, B.A. & Pfleeger, S.L. (2002b) Principles of survey research part 3: Constructing a survey instrument. *Software Engineering Notes*, 27(2), 20–24.

- Kitchenham, B.A. & Pfleeger, S.L. (2002c) Principles of survey research part 4: Questionnaire evaluation. *Software Engineering Notes*, 27(3), 20–23.
- Kitchenham, B.A. & Pfleeger, S.L. (2002d) Principles of survey research part 5: Populations and samples. *Software Engineering Notes*, 27(5), 17–20.
- Kitchenham, B.A. & Pfleeger, S.L. (2003) Principles of survey research part 6: Data analysis. *Software Engineering Notes*, 28(2), 24–27.
- Lewis, J.R. (1995) IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7, 57–78.
- Logisoft. <http://www.logisoft.com/>. (4.3.2004)
- Lucey, N.M. (1991) More than meets the I: User-Satisfaction of Computer Systems. Unpublished thesis for Dipl. Applied Psychol, University College Cork, Ireland.
- Lund, A.M. (2001). Measuring usability with the USE questionnaire. *Usability Interface*, 8 (2). [http://www.stcsig.org/usability/newsletter/0110\\_measuring\\_with\\_use.html](http://www.stcsig.org/usability/newsletter/0110_measuring_with_use.html). (3.3.2004)
- Mathieson, K. (1991) Predicting user intentions: Comparing the Technology Acceptance Model with the Theory of Planned Behavior. *Information Systems Research* 2(3), 173–191.
- Neale, J.M. & Liebert, J. (1986) *Science and Behavior: Introduction to Methods of Research*. Prentice-Hall.
- Nielsen, J. (1994) Enhancing the explanatory power of usability heuristics. *Proceedings of CHI'94*, 152–158. ACM Press.
- Nielsen, J. (1997) Let's ask the users. *IEEE Software*, 14 (3), 110–111.
- Nyt.fi. <http://www.nyt.fi> (26.1.2004)
- Punter, T., Ciolkowski, M., Freimut, B. & John, I. (2003) Conducting on-line surveys in software engineering. *2003 International Symposium on Empirical Software Engineering (ISESE'03)*, 80–88.
- Robson, C. (1994) *Real World Research: A Resource for Social Scientists and Practitioner-researchers*. Blackwell publishers.
- Root, R.W. & Draper, S. (1983) Questionnaires as a software evaluation tool. *Proceedings of CHI'83*, 83–87.
- Shannon, D.M., Johnson, T.E., Searcy, S. & Lott, A. (2002) Using electronic surveys: advice from survey professionals. *Practical Assessment, Research & Evaluation*, 8(1) <http://pareonline.net/>. (4.3.2004)
- SUMI (2004) Human Factors Research Group. SUMI: Pricing and order information. <http://www.ucc.ie/hfrg/questionnaires/sumi/pricing.html>. (19.8.2004)
- Teague, R., De Jesus, K. & Nunes-Ueno, M. (2001) Concurrent vs. post-task usability test ratings. *CHI '01 extended abstracts on Human factors in computing systems*, 289–290.
- WAMMI consortium. (2002) WAMMI – Web Usability Questionnaire. <http://www.wammi.com/about.html>. (21.6.2004)
- Wong, G.K. & Rengger, R. (1990) The validity of questionnaires designed to measure user-satisfaction of computer systems. *National Physical Laboratory report DITC 169/90*, Teddington, Middx., UK.