

Alex Liljeblad

# A/B-käyttäjätestauksen suunnittelu

Metropolia Ammattikorkeakoulu

Insinööri (AMK)

Mediatekniikan koulutusohjelma

Insinöörityö

16.4.2017

Tekijä Otsikko	Alex Liljeblad A/B-käyttäjätestauksen suunnittelu
Sivumäärä Aika	25 sivua 16.4.2017
Tutkinto	Insinööri (AMK)
Koulutusohjelma	Mediatekniikka
Suuntautumisvaihtoehto	Digitaalinen media
Ohjaajat	Esimies Tutkijaopettaja Merja Bauters
<p>Insinööriyön tarkoituksena oli tarkastella A/B-testauksen soveltuvuutta yrityksen asiakasviestintätiimin arkeen yhden viestintäkanavan, henkilökohtaisten asiakasviestien tapauksessa. Tutkimus suoritettiin päivittäisen työn ohessa tarkkailemalla ja osallistumalla usean kampanjan luonnin kuuden viikon mittaiseen työprosessiin samanaikaisesti ja haastatteleamalla asiakasviestintätiimin jäseniä. Tavoitteena oli paikantaa aikaa vievimmat ja työläimmät työosuudet ja sovittaa uuden työtavan vaatimat lisätoimet niihin minimaalisella työkuorman lisäyksellä helpoimman mahdollisen prosessimuutoksen löytämiseksi.</p> <p>Lopulta julkaistiin uudistetun prosessin mukaisesti testikampanja tiimille tulleen oikean toimeksiannon pohjalta. Prosessimuutos rajattiin teknisten valmiuksien vuoksi viestinnän tekstuaaliseen sisältöön lähestymällä samaa viestiä eri sävyillä, hypoteesina nuorille asiakkaille suunnatun tekstuaalisen viestinnän epämuodollisuuden ja helppouden korostamisen suurempi houkuttelevuus. Kontrollina toimi perinteisempää virallista sävyä käyttävä ja etuja painottava versio. Testi saatiin onnistuneesti loppuun koetilanteessa, ja tuloksia verrattiin saman kanavan keskimääräiseen suoriutumiseen testausajan loputtua.</p> <p>Tässä yksittäisessä testissä ei havaittu testiversioiden välillä huomattavia testin hypoteesia tukevia eroja. Uusitun prosessin soveltaminen olemassa oleviin kampanjoihin, eli vaihtoehtoisten uusien sisältöversioiden luominen, vaatii jatkossakin yhteydenpitoa tilaajan ja asiantajajien kanssa, mutta tässäkin on mahdollista lisätä tehokkuutta tarkistuttamalla useampi vaihtoehtoinen teksti kerralla kokeillun prosessin mukaisesti. Yhteistyö ja läpinäkyvyys sivuston ylläpitäjien ja käyttäjien tai kehittäjien välillä osoittautui byrokraattiseksi ja aiheuttaa nykyisellään kohtuuttoman työ määrän pienten ulkoasuun ja toiminnallisuuteen vaikuttavien optimointitestien potentiaaliseen hyötyyn nähden. Kattavamman sivuston koodia ja siten ulkonäköä ja toiminnallisuuksia vaihtelevan testauksen mahdollistavien työkalujen käyttöönotto tulisi tuoda esiin jo uutta järjestelmää suunniteltaessa, tai nykyiseen alustaan tehtäviin muutoksiin tulisi hakea kattavampaa hyväksyntää jo huomattavasti etukäteen. Muutosten hitaus yhdistettynä arvioon, että niiden vaikutus olisi jo ennustettavasti huomattavasti marginaalisempi kuin käyttäjäpolun kriittisimpään pudotukseen, eli viestin avaamiseen sen otsikkoa klikkaamalla, pureutuminen aiheutti niiden poisrajoituksen projektista.</p>	
Avainsanat	Prosessiuudistus, A/B-testaus, elektroninen markkinointi

Author Title	Alex Liljeblad Planning an A/B-usertest
Number of Pages Date	25 pages 16 April 2017
Degree	Bachelor of Engineering
Degree Programme	Media Technology
Specialisation option	Digital Media
Instructors	Foreman Merja Bauters, Senior Lecturer
<p>This thesis studies A/B-testing in general, and how it could be used in the daily work of the customer's Customer Communication Team with web platform's personalized private messages as the study case. The study was conducted by partaking and observing the 6 week work process of creating multiple campaigns for this medium, as well as interviewing members of the team. The goal was to pinpoint the most arduous and time consuming parts of the process and implementing the new tasks required for testing into it while causing minimal workload increase in order to find the easiest possible process change.</p> <p>A test campaign was published and executed according to an actual order from a client. Due to limited possibility of changing the technical aspects the process change was intentionally limited to renewing and testing of text content and approaching clients from the same target group with differing tones of voice with the hypothesis that younger clientele would respond better to a message conveying ease and an informal tone. A more traditional and formal message that stressed monetary gain and other benefits was used as a control. The test was conducted successfully and the results were compared against the average performance of the communication channel.</p> <p>The test results showed no significant difference that would have supported the hypotheses. Application of the renewed process for previously existing campaigns will still require contact between the team, the contractor and the legal department which was considered to be the most time consuming part of the process, but the observations can still be applied for efficiency increase by getting approval for multiple variants at the same time as per the renewed process. The cooperation and transparency between the technical personnel, the users and the development personnel turned out to be quite bureaucratic and showed that the workload caused by minor functional and appearance changes to the user interface would be greater than the benefit from minor optimization testing performed with them. To enable more fluid testing for the appearance and functionality of the user interface the introduction and requests of changes, as well as the tools needed for these changes, should be done well in advance, possibly during considerations of platform renewal plans. The slowness of changes together with the predicted minor results yielded caused them to be outscoped in order to concentrate the testing into the biggest single dropout in the user funnel – clicking on the message headline.</p>	
Keywords	Process renewal, A/B-testing, electronic marketing

## Sisällys

### Lyhenteet

1 Johdanto	1
2 A/B-testaus ja sen suhde muihin testausmenetelmiin	2
2.1 Yleistä	2
2.2 A/B-testauksien variaatioita	4
2.3 Otoksen koko, seuranta-aika ja uusintatestaus A/B-testauksessa	6
3 Testauksen suunnittelu	8
3.1 Onnistumisen määrittely ja tavoitteiden luominen	8
3.2 Vaihtoehtojen, sanavalintojen ja informaation määrän punnitseminen	10
3.3 Käyttäjän suhtautuminen informaation määrään	10
3.4 Sananvalintojen kriittisyys	11
3.5 Potentiaalisia työkaluja yleiseen A/B-testaukseen	12
4 Testauksen suunnittelu ja toteutus	13
4.1 Tarve ja nykyinen prosessi	13
4.2 Testin tavoitteet ja hypoteesi	15
4.3 Toteutus	17
4.4 Tulokset	19
5 Yhteenveto	20
Lähteet	22

## Lyhenteet

KPI	Key Performance Indicator. Tavoitteiden kannalta olennainen mitattava määre, jonka perusteella arvioidaan onnistumista.
RVP	Revenue Per Visitor. Yksikkö jolla mitataan palvelulle tai yritykselle tulevaa rahavirtaa yksittäistä käyttäjää kohden.
CTA	Call To Action. Toimintakehote, eli sisällön osa, jonka on tarkoitus saada käyttäjä suorittamaan haluttu toiminto palvelussa.
WYSIWYG	What You See Is What You Get. Käyttöliittymätyyppi, jossa näkymä päivittyy reaaliajassa edustamaan käyttäjälle esitettävää versiota esimerkiksi koodinäkömön sijaan.

## 1 Johdanto

Asiakasviestinnässä viestin ymmärrettävyys ja vastaanotto ovat avainasemassa, varsinkin jos kyseessä on aihealue, joka sisältää paljon tietyn alan teknistä sanastoa. Täysin lakivelvoitteisten, yksisuuntaisten tiedoksiantojen lisäksi insinööriyön toimeksiantaja Danske Bank Oyj tuottaa monissa eri kanavissa asiakkailleen markkinointikampanjoita, joiden tavoitteena on paitsi tarjota asiakkaan senhetkisen asiakassuhteen kannalta uusia tuotteita ja palveluja, myös informoida asiakasta edistyneemmistä, edullisemmista ja helpommista tuotteista ja toimintatavoista nykyisen asiakaskäyttäytymisen ja asiakaskokemuksen parantamiseksi.

Monikanavaviestinnän prosessit muuttuvat yrityksessä jatkuvasti, mutta painopiste yrityksen asiakasviestinnässä on ollut vahvasti uusissa ratkaisuissa säännöllisten olemassa olevien ja toistuvien asiakaskampanjoiden jatkojalostuksen sijaan. Toistuvien kampanjoiden määrän ja laajuuden vuoksi on mielekästä pyrkiä parantamaan sekä viestien tavoitavuutta että vastaanottoa A/B-testauksen avulla, jolloin olisi mahdollista hyödyntää olemassa olevia ratkaisuja ja oheismateriaaleja tavoitteena asiakaskokemuksen parantaminen lisäämättä yhteydenottojen määrää. Monikansallisessa käytössä olevan infrastruktuurin uusimisen sijaan insinööriyössä on tavoitteena selvittää kaksi asiaa:

- Millainen on tilaajayrityksen olemassa oleva kampanjanluontiprosessi?
- Kuinka prosessin eri vaiheita voidaan muokata tukemaan A/B-testausta lisäämättä työkuormaa tuntimääräisesti?

Nykyprosessi kartoitetaan haastattelemalla Danske Bankin asiakasviestintätiimin työntekijöitä ja seuraamalla heidän työntekoaan tavoitteena paikantaa aikaavievimmät ja työläimmät työosuudet. Tavoitteena on sovittaa uuden työtavan vaatimat lisätoimet työosuuksiin minimaalisella työkuorman lisäyksellä.

## 2 A/B-testaus ja sen suhde muihin testausmenetelmiin

### 2.1 Yleistä

Yleisimpien käyttäjätestausmenetelmien joukossa A/B-testaus sijoittuu käyttäytymistä tutkivien ja kvantitatiivisten metodien joukkoon, eli sen peruskysymykset ovat ”Mitä käyttäjä tekee?” ja ”Kuinka moni käyttäjä toimii näin, ja kuinka paljon?” Samantyyppisiä, tuotetta tai alustaa käyttäviä metodeja ovat esimerkiksi klikkivirta-analyysi (clickstream analysis) ja lämpökarttojen käyttö, joissa käyttäjän näkymää ja toimintaa testattavalla alustalla tallennetaan analysointia ja myöhempiä toimenpiteitä varten. Molemmat soveltuvat myös passiivisen luonteensa vuoksi käytettäväksi A/B-testauksen yhteydessä joko hypoteesia luotaessa tai testausvaiheessa lisätiedon keräämiseen. Tämän tyyppisten menetelmien vastakohtana ovat käyttäjien subjektiivista asennetta tutkivat ja laadulliset menetelmät, jotka vastaavat kysymyksiin ”Mitä ihmiset sanovat?” ja ”Miksi ja miten asia korjataan?” Näissä menetelmissä, kuten asiakaspalautteiden kerääminen ja haastattelut, ei käyttäjän havainnoinnissa toimita epäsuorasti hyödyntämällä analytiikkatyökaluja tai mittauksia, vaan havainnot ja näkemykset tutkittavasta aiheesta kerätään suullisesti ja/tai kirjallisesti, jolloin alustan tai tuotteen käyttö ei ole kaikissa menetelmissä välttämätöntä. Myös näitä menetelmiä voidaan käyttää tukena A/B-testausta suunniteltaessa ja testin tuloksia arvioitaessa. [10; 11; 12.]

Ääripäiden välille mahtuu lukuisia metodeja, joissa tuotetta tai alustaa (tai sen prototyyppiä tai konseptuaalista mallia) käytetään omaehtoisesti tai valvotuissa olosuhteissa annettujen tehtävien suorittamiseen tai luontaisesti käyttäjän haluamalla tavalla ilman käsikirjoitusta. Monia erityyppisiä menetelmiä voidaan käyttää yhdessä tai erikseen tuotteen kehityksen eri vaiheissa mahdollisuuksista ja tavoitteista riippuen. Tuotteen *suunnitteluvaiheessa* hyödynnetään usein sekä laadullisia että määrällisiä menetelmiä, kuten kenttätutkimuksia, kamera- ja päiväkirjatutkimuksia ja analytiikkaa, kehityksen suunnan määrittämiseksi kysynnän, tarjonnan ja potentiaalisen kohdeyleisön päivittäisen toiminnan perusteella. *Toteutusvaiheessa* keskitytään pääosin laadulliseen tutkimukseen, jolloin valitun kehityssuunnan mukaisesti tuotetta jatkokehitetään paperiprototyypeillä, osallistavilla askartelupajamaisilla tutkimuksilla, korttien lajittelulla ja houkuttelevuustutkimuksella. Näissä menetelmissä koehenkilöille tarjotaan mahdollisuus tutustua tuotteen konseptiin ja hiomattomiin näkymäluonnoksiin, luoda oma näkemyksensä optimaalisesta

tuotteesta, vaikuttaa tulevan palvelun arkkitehtuuriin ryhmittelemällä korteilla olevaa sisältöä ja arvioida eri vaihtoehtoisten näkymien visuaalista houkuttelevuutta. Ottamalla käyttäjien mieltymykset huomioon minimoidaan riski jo tuotannossa olevan tuotteen heikosta suoriutumisesta markkinoilla. Kun tuotteen kehitys on riittävän pitkällä ja se saavuttaa riittävän määrän käyttäjiä, voidaan siirtyä *arviointi- ja jatkokehitysvaiheeseen* eli keräämään pääasiassa määrällisin tutkimusmenetelmin kuten suorituskykytestein tai A/B-testauksella dataa tuotteen suoriutumisesta ja vertaamaan saatuja tuloksia joko aikaisempiin tuloksiin tai kilpaileviin tuotteisiin.

[10; 11.]

A/B-testaus on vaihtoehtoisten toteutustapojen käyttäjätestausta alkuperäisen rinnalla ennalta määritellyn tavoitteen saavuttamiseksi. Palvelusta luodaan kontrollikappaleen eli alkuperäisen sisällön lisäksi uusia rinnakkaisia variaatioita, joihin käyttäjät ohjataan. Käyttäjien toiminta kussakin versiossa tallennetaan koneellisesti mittaamalla esimerkiksi käyttösession kestoa, painikkeiden klikkausten määrää, mainoksien esityskertoja, rekisteröitymisten määrää tai vaikka lahjoitusten summaa ja prosentuaalista todennäköisyyttä asetetuista tavoitteista riippuen. Näin saatua dataa vertaillaan, ja tavoitteet parhaiten saavuttanut versio otetaan käyttöön. Nimen antamasta mielikuvasta ja alkuperäisestä käyttötavastaan poiketen A/B-testaus voi nykymuodossaan sisältää useita rinnakkaisia vaihtoehtoisia versioita. Käsitteen laventuminen ajan saatossa selittyy testaustyylin historialla, joka ulottuu Internetiä edeltävään aikaan. TV- ja lehtimainonnan valtakaudella markkinoijat painivat samantyyppisten kysymysten kanssa halutessaan selvittää tehokaimman markkinakanavansa, mutta paino- ja TV-tuotannossa potentiaalisten asiakkaiden käyttäytymisestä jäi jälki vain heidän ottaessaan yhteyttä. Mitatakseen eri kanavien keskinäistä suoriutumisen eroa mainostajat julkaisivat kullekin esimerkiksi oman puhelinnumeronsa käytettäväksi markkinointimateriaalissa, ja kuhunkin numeroon tullutta puheluvirtaa verrattiin vastakkain. Nykyään rinnakkaisten versioiden julkaisun helppouden ja halpuuden sekä datankeruun välittömyyden ja tarjolla olevien suurien kohderyhmien myötä yritykset pyrkivät usein tehostamaan testausta kohderyhmien ja versioiden määrää kasvattamalla. [1; 7; 10.]

Useissa tapauksissa testaus suoritetaan jo olemassa olevalle palvelulle, jolloin ensimmäistä versiota kutsutaan kontrollikappaleeksi. Vaikka palvelusta ja sen käyttäjistä olisi käytössä jo valmiiksi dataa, on olennaista suorittaa kaikkien vaihtoehtojen testaus sa-



manaikaisesti koska loputon määrä alustasta ja testattavista ominaisuuksista riippumattomia muuttujia saattaa vaikuttaa palvelun suosioon; hellejaksolle tai kansainvälisille pyhäpäiville sijoittuva internetsivuston testaus saattaisi vääristää joitakin testattavia suoriutumisarvoja kuten kävijä- ja myyntimääriä huomattavasti, jos uusien versioiden suoriutumista verrattaisiin kontrolliversioon sesonkiajan suoriutumiseen. Esimerkiksi leluja myyvän sivuston asetteluvaihtoehdon B:n tammikuisen suoriutuminen vertaaminen kontrolliversio A:n joulukuun dataan ei antaisi tarkkaa kuvaa joululahjaostosten, veronpalautusten ja muiden mahdollisten ennalta-arvaamattomien muuttujien vuoksi. [10]

Testauksen tavoitteiden tulisi olla selkeät oikeiden mittareiden valitsemiseksi ja koko testin suunnittelun ohjauksen helpottamiseksi. Onnistuneen testauksen takaamiseksi kävijämäärien tulisi olla kohtuulliset ja testausajan riittävän pitkä, jotta koetuloksia, erityisesti konversioiden määrää ja konversiosuhdetta voidaan pitää tilastollisesti luotettavina. Myös käyttäjien laatua on olennaista seurata; palvelun vakiokäyttäjät ovat luoneet sivuston käytölle rutiinin, joten elementtien tai niiden sijainnin muuttaminen saattaa vaatia totuttelua. Toisaalta nämä toistuvasti sivustolla vierailevat käyttäjät myös sietävät huomattavasti enemmän pieniä virheitä, sillä heillä on kokemusta ja luottoa palveluun ja he muistavat sen sisältävän ominaisuuksia, joita he hakevat. A/B-testauksen tilastollisen luonteen vuoksi on olennaista, että testattavalla palvelulla on riittävästi käyttäjiä. Tämän vuoksi se onkin tehokkaimmillaan jo olemassa olevien ja tehokkaasti markkinoitujen palveluiden testaamisessa. [2; 9.]

## 2.2 A/B-testauksien variaatioita

### Multivariaatiotestaus

Vaikka A/B-testauksen ja multivariaatiotestauksen välillä on yhtäläisyyksiä, ovat niiden käyttötarkoitukset ja testimäärät aivan erilaisia. A/B-testauksessa testattavien versioiden määrä on puhtaasti eri sivuvariaatioiden mukainen, eli muutamia kappaleita, kun taas multivariaatiotestauksessa (multivariate testing) kutakin yksittäisen elementin vaihtoehdoista muotoa testataan ristiin kaikkien muiden tutkinnan alla olevien elementtien variaatioiden kanssa. Tällöin kolmen eri elementin, joilla kullakin on kolme vaihtoehdoista esitystapaa, mahdollisten erilaisten yhdistelmien määrä nousee kolmen kolmanteen potenssiin, eli 27 erilaiseen yhdistelmään. Vaihtoehdoisten sivunäkymien kasvaessa myös vaadittu kävijä- ja konversiomäärä luotettavan ja merkittävän testituloksen saamiseksi

kasvaa. Vaikka multivariointissakin suositellaan pysymään 25 vaihtoehdon alapuolella, on ero A/B-testauksen muutamaan rinnakkaiseen vaihtoehtoon ja niiden vaatimiin liikennemääriin valtava. [3; 7.]

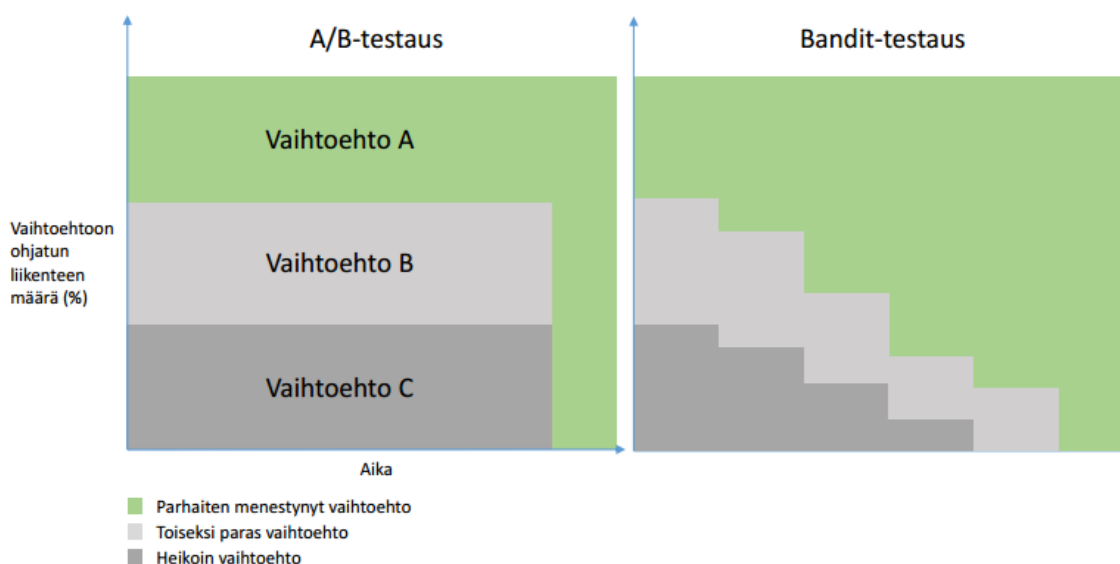
Koska multivariointi perustuu usein yksittäisten pienempien elementtien muunteluun, sitä käytetään usein olemassa olevan sivunäkymän hienosäätöön ja optimointiin, kun taas A/B-testaus on useimmiten käytössä mittavammassa sivuston muotoiluun vaikuttavissa testeissä. Hienovaraisuutensa vuoksi multivariaatiotestaus antaaakin yksityiskohtaisempaa dataa yksittäisten elementtien vaikutuksesta; monessa multivariaatiotestauksen raportointityökaluissa on erillinen ”impact factor” -niminen tai vastaava määre, joka kertoo testaajalle suoraan, millä yksittäisellä elementillä oli suurin vaikutus konversioasteeseen. [3; 7.]

#### Bandit- eli yksikätkäinen rosvo -testaus

Klassisesta ajatusleikistä tai ongelmasta nimensä saanut Bandit-testaus eroaa A/B-testauksesta rinnakkaisten versioiden esityksen laajuuden, testiajan pituuden ja siten myös algoritmin monimutkaisuuden osalta: Yhden identtisillä ja staattisilla kävijämäärillä ja ajanjaksoilla suoritettun testin sijaan käyttäjiä ohjataan tasaisen lähtöasetelman jälkeen dynaamisesti enenevässä määrin käyttämään kullakin hetkellä parhaiten suoriutuvaa versiota. Siirtyminen testin voittajaan on siis vaihteellinen. Vahvimpia tämän testausmuodon puolesta esitettyjä argumentteja ovatkin testituloksista hyötyminen jo testin suoritushetkellä koneoppimista ja automatisaatiota hyväksikäyttäen, jolloin alemman konversion vaihtoehtoihin ohjattujen potentiaalisten asiakkaiden menetetty rahavirta minimoidaan. Vastavuoroisesti datan kertymisen jatkuminen heikommista vaihtoehdoista mahdollistaa niiden kelpoisuuden uudelleenevaluoinnin ennen lopullista päätöstä. [13; 14; 15.]

Bandit-testauksen yleisiä vahvuuksia ovat optimointi ja ansainta joko lyhyissä, aikataulukriittisissä ja ketteryyttä vaativissa tilanteissa, kuten promootiokampanjoissa, tai pitkän aikavälin automatisoiduissa testeissä, kuten sivustojen sisältösuosituksissa. A/B-testauksella tutkimuskysymykseen ja hypoteesiin perustuvaan muutokseen saadaan ymmärrettävämpi ja tilastollisesti pätevä vastaus lyhyemmässä ajassa. Monimutkaisuutensa vuoksi Bandit-testauksen soveltuvuus varsinkin pienille yrityksille on rajallinen, pit-

källe viety automaatio ja monimutkaisten tulosten tulkinta vaativat huomattavasti enemmän osaavaa työvoimaa ja aiheuttavat isompia riskejä kuin yksinkertaiset testimuodot. [13; 14; 15; Kuva 1.]



Kuva 1: Asiakasvirran ohjautuminen A/B- ja Bandit-testauksessa. [13.]

### 2.3 Otoksen koko, seuranta-aika ja uusintatestaus A/B-testauksessa

Yleinen virhe on lopettaa A/B-testi ennen riittävän otoksen kertymistä. Käyttäjien tai vierailijoiden määrää tärkeämpää testauksessa on konversiosuhde eli se, kuinka moni käyttäjä toimi asetetun tavoitteen tai toiveen mukaisesti vierailtuaan sivulla. Tavoiteltu konversiosuhde on hyvä suhteuttaa yleiseen käyttäjämäärään, mutta lähtökohtaisesti testin tuloksia tulisi tarkastella vasta, kun konversiotavoite on täytynyt eli se on riittävän suuri. Tällöin tilastollisen vääristymän todennäköisyys laskee ja tulokset ovat varmempia. Esimerkiksi Peep Laja [2] suosittelee 15 vuoden kokemuksellaan keräämään vähintään 250 konversiota kutakin vaihtoehtoa kohden ennen kokeen lopettamista. Mikäli riittävän suuria otoskokoja on vaikeaa tai mahdotonta saada, tulisi harkita A/B-testauksen mielekkyyttä ja keskittyä kokonaiskävijämäärän kasvattamiseen markkinoinnin ja mainonnan kautta. SuperAnalyticsin Jarno Oksasen [5] mukaan luotettavan A/B-testin rajana voi pitää 10 000:ta kuukausikäyttäjää ja 100:aa konversiota. [2; 3; 4; 5.]

Internetissäkin laajalti tarjolla olevat laskurit antavat kaavasta testin prosentuaalisen tilastollisen luotettavuuden, jonka saa selville jo kymmenten käyttäjien osallistujaryhmillä. Testing Agencyn perustaja Tom Wesseling [42] muistuttaa, että mikäli merkittävyyttä, eli todennäköisyyttä toteutuneen nollahypoteesin (vaihtoehdoilla ei huomattavaa eroa) hylkäämättä jättämiselle sen ollessa tosi, käytetään avainperusteena testin lopettamiselle, 77 % samaa versiota itseään vastaan vertaavista testeistä tarjoaa selkeän voittajan. Toisaalta, mikäli rinnakkaisena tilastollisena mittarina päätetään käyttää myös testin tehoa, eli todennäköisyyttä virheellisen nollahypoteesin hylkäämiselle, ja tavoiteltu teho asetetaan todella korkealle, saattavat vaaditun kohderyhmän koon laskelmat antaa kohtuuttoman suuria lukuja alustan kävijämääriin nähden. Strategisesti tärkeisiin testeihin suositellaankin käytettäväksi tehon ja merkittävyyden yhdistelmää. Mikäli riittävän suuria otoskokoja on vaikeaa tai mahdotonta saada, tulisi harkita A/B-testauksen mielekkyyttä ja keskittyä ensin kokonaiskävijämäärän kasvattamiseen markkinoinnin ja mainonnan kautta. [4; 2; 40; 41; 42.]

Päivittäinen vaihtelu sivuston liikenteessä ja kävijöiden käyttäytymisessä saattaa vaikuttaa huomattavasti testin tuloksiin. Testattaessa tulisi siis aina ottaa huomioon, minä päivänä testi on aloitettu, ja testata kokonaisia viikkoja tilastollisen vääristymän välttämiseksi. Myös vuodenajoilla, juhlapyhillä ja muilla ulkoisilla tekijöillä, kuten vaikka jalkapallon maailmanmestaruuskisoilla, saattaa olla vaikutusta erityisesti samaa kohdeyleisöä tavoittelevan urheiluaiheisen alustan testituloksiin, vaikka kaikki vaihtoehdot ovatkin yhtäaikaaisesti rinnakkain. Tämän vuoksi riittävän pitkät ja konversiosuhteellisesti riittävät testit tulisi olla valmis uusimaan ulkoisten tekijöiden poissulkemiseksi ja tulosten varmistamiseksi. [2.]

Seuranta-ajan pituuden tulisi määrittäytyä testin valmistumisen, ei ulkoisten aikataulupaineiden perusteella. Kuten otoksen kokoa määriteltäessä, myös sen pituudessa tulisi konversioita olla mieluummin vähintään satoja ja tilastollisen merkittävyyden ainakin 95 %. Konversiot eivät kuitenkaan yksin riitä, sillä erityisesti korkean liikenteen sivustoilla toivottuja konversioita voi kertyä todella nopeasti, jolloin asetettu minimitaloite saattaa täytyä yksittäisen kävijäpiikin seurauksena. Testin tulisi mieluiten olla voimassa vähintään yksi kokonainen tunnistettava palvelu-, käyttäjä- tai ostosykli (konversiotavoitteesta ja testattavan alustan tyypistä riippuen), joiden tunnistamisessa auttaa huomattavasti ennalta kerätty käyttäjädata. Mikäli aikataulupaineita ei ole, yhdeksi testin pituuden rajoittajaksi

muodostuu datan saastuminen esimerkiksi toistuvien kävijöiden poistaessa välissä evästeet alustalla, jossa käyttäjiä ei tunnisteta muilla keinoin. Vuonna 2007 tehdyssä tutkimuksessa kolme kymmenestä internetin käyttäjästä poisti evästeet kuukauden sisällä, keskimäärin viikon välein, ja vuonna 2013 tehdyssä kokeessa vain 26 % evästeistä selvisi neljä viikkoa ilman poistamista. Tästä huolimatta, mikäli asetettu minimikonversiotavoite ei täyty, on parempi jatkaa testiä viikolla kuin tehdä johtopäätöksiä vähäisellä datalla. [40; 43; 44.]

### 3 Testauksen suunnittelu

#### 3.1 Onnistumisen määrittely ja tavoitteiden luominen

Onnistuneen testin perusta on selkeä tavoite. Nykymallisessa testauksessa tarjolla on työkalusta riippuen vaihteleva määrä mahdollisia mitattavia arvoja, joten on tärkeää valita mitattava data sekä yrityksen että testattavan alustan tavoitteen mukaan ennen muutosten tekemistä. Mikäli kyseessä on markkinointitoimisto ja yhden sivuston osuuden konversioksi lasketaan esimerkiksi uutiskirjeen tilaaminen, on muutosten tavoitteena saada useamman kävijän asiakaspolku kestämään aloitussivulta tilauslomakkeelle ja tilausnapin painamiseen saakka. Tällöin testattavan alustan osalta osatavoitteiksi voitaisiin määritellä mitä tahansa, minkä voidaan olettaa vaikuttavan tilauksiin, kuten sivuston ulkoasun parantaminen, tilausprosessin keskeyttävien asiakkaiden määrän vähentäminen ja tilausten määrän kasvattaminen. [8; 16.]

Tavoitteita asettaessa tulee niille kyetä määrittelemään suorituskykymittarit (KPI, Key Performance Indicator) eli mitattavat arvot, jotka mahdollistavat eri tavoitteiden täyttymisen seurannan. Tämä helpottaa selkeiden osatavoitteiden valintaa ja auttaa rakentamaan selkeän kuvan testin tehokkuudesta. Kullekin suorituskykymittarille tulisi valinnan yhteydessä asettaa tavoitearvo, joka tukee testin ja yrityksen tavoitteita, jolloin testit pysyvät varmasti niiden suhteen olennaisina. Osatavoitteiden suorituskykymittareiden tulosten tulkinnan kannalta on tärkeää, että ylätasen tavoitteet ovat selkeät; yksittäisen osatavoitteen osoittama paras vaihtoehto ei välttämättä täytä yrityksen päällimmäisiä tavoitteita parhaalla tavalla. Jos esimerkin markkinointitoimisto asettaisi päätavoitteekseen mainostulojen kasvattamisen, saattaisi mainospaikkojen lisääminen uutiskirjeen tilauslo-

makkeelle johtavan asiakaspolun varrelle vähentää tilausten määrää kasvaneen katkenneiden istuntojen määrän myötä ja siten pienentää uutiskirjeessä olevien mainosten leikkikettä, mutta näytettävien mainosten kokonaismäärä saattaisi silti kasvaa. Tällöin yksittäisen osatavoitteen osalta heikommin suoriutuva versio olisi yritykselle edullisempi vaihtoehto, mikäli ylätasen tavoite on valittu oikein. Yksittäisistä mittareista A/B-testauksessa tärkeimmäksi nousevat usein konversiot, eli palvelun tai alustan toivottujen funktioiden täyttyminen yksittäisen kävijän osalta. Erityisesti kaupallisissa palveluissa konversion rinnalla kuitenkin myös toinen, käytön laatua arvioiva mittari. Kävijäkohtainen tuotto (RPV, Revenue Per Visitor) mittaa yksittäisen käyttäjän tai kävijän tuomaa rahavirtaa ja voi tutkittaessa osoittaa esimerkiksi edellä mainitun mainospaikkojen lisäämisen asiakkaan kannalta paremmaksi strategiaksi. Kehittyneemmät testaus- ja analytiikkatyökalut, kuten Optimizely, tarjoavat oletuksena kävijäkohtaisen tuoton laskennan.

[8; 16; 17; 18.]

Jotta muutoksia testattavaan alustaan voidaan tehdä, tulee muutosten perustua hypoteesiin, eli nykyisestä asiakasvirrasta tarjolla olevaan tietoon pohjautuvaan olettamukseen asiakkaiden käytöksen motiiveista. Mikäli edellisen esimerkin markkinointitoimiston alustalla huomattava osa käyttäjistä saavuttaa lomakkeen mutta ei kirjoita kenttään sähköpostiosoitettaan tai paina tilausnappia, on oletettavaa, että lomakkeen näkyvyys tai sen välittömässä läheisyydessä oleva tekstisisältö ja tilausnapin teksti ovat osallisia heikkoon konversioon. Mitä enemmän tietoa asiakkaista on tarjolla, sitä pidemmälle oletuksen toivotusta tekstisisällöstä voi viedä: kaivataanko lomakkeessa ehkä epämuodollisempaa kielenkäyttöä nuorille, suurempaa kirjasinkokoa iäkkäille kävijöille tai mahdollisesti vain ytimekkäämpiä perusteluja uutiskirjeen hyödyistä? Vahvan hypoteesin luomiseksi tulisi myös mahdollisuuksien mukaan kerätä ja hyödyntää asiakaspalautetta esimerkiksi palautelomakkeen avulla. Suoraan käyttäjiltä tullut palaute saattaa vahvistaa hypoteesia entisestään ja tuoda uusia näkökulmia. [2; 8.]

### 3.2 Vaihtoehtojen, sanavalintojen ja informaation määrän punnitseminen

Testiä suunniteltaessa on hyvä käydä läpi useita mahdollisia ratkaisuja ennen pidemmälle vietyä jalostamista. Verkkosivun nykyinen tila saattaa olla testaustuloksiltaan huomattavasti heikompi kuin reilusti eri suuntaan viety toinen sisältö- ja asetteluvaihtoehto. Jos keskitytään hiomaan olemassa olevaa sivustoa pienin muutoksin, saatetaan menettää arvokasta dataa kävijöiden mieltymyksistä ja tämä muuntuu menetetyiksi potentiaalisiksi asiakkaiksi. Tässä vaiheessa on erittäin hyödyllistä, jos käytettävissä on dataa jo julkaistun sivuston kävijöiden toiminnasta, kuten yleisimmistä hakusanoista ja asiakaspoluista. Muokkaamalla koko valikkorakennetta ja etusivun asettelua vastaamaan yleisimpiä haun kohteita koko sivuston tehokkuus, kuten sivulatausten määrän lasku toivotua sisältöä etsiessä tai käyttäjäpolun lyheneminen, paranee ja asiakkaat todennäköisesti myös käyttävät sivua pidempään ja säilyttävät sen priorisoituna tietolähteenä vaihtoehtoisiin listaussivustoihin tai hakukoneisiin verrattuna. Asiakaspolkuja tutkimalla kyetään osoittamaan nykyisen alustan tehokkaat osuudet ja pullonkaulat, joissa käyttäjä syystä tai toisesta lopettaa käytön. Tämän tiedon avulla voidaan joko pyrkiä tehostamaan nykyistä asiakaspolkua korjaamalla ongelmakohtat saatavilla olevan tiedon pohjalta tehdyn hypoteesin perusteella tai pitää vain hyvin suoriutuvat osuudet ja suunnitella asiakaspolkua uudestaan niiden ympärille. [19; 20.]

### 3.3 Käyttäjän suhtautuminen informaation määrään

Informaatiota täyteen pakatut sivut ja mahdolliset turhat kentät lomakkeissa saattavat ajaa käyttäjiä pois palvelusta ja vaikeuttaa hakukoneiden toimintaa. Yksinkertaistaminen ja vaihtoehtojen, toivotun käyttäjän toiminnan kannalta epäolennaisiin lopputuloksiin johtavien reittien poistaminen ja toiminnan helpottaminen on tärkeää. Jos käyttäjä lukee keskimäärin vain viidenneksen esitetyn näkymän tekstisisällöstä ja sijoittaa vain 4,4 sekuntia enemmän aikaa näkymän läpikäyntiin jokaista 100 lisättyä sanaa kohden (jonka aikana keskivertokäyttäjä ehtii lukea 14,7 sanaa 200 sanan minuuttinopeudella), on tärkeää tiivistää sisältö käyttäjän kannalta olennaiseen, helppo- ja nopealukaiseen muotoon, jotta mahdollisimman moni käyttäjä saa konversion kannalta olennaisen tiedon ja kykenee jatkamaan toimintaa haluttua lopputulosta kohti. Hyvänä suuntaa antavana muistisääntönä voi pitää NN-Groupin tutkimuksessa [21.] saatua lukuparia, jossa käyttäjät lukivat puolet tai enemmän tekstuaalisesta informaatiosta vain sivuilla joilla oli 111 sanaa tai vähemmän. [21; 22; 23.]

Myös käyttöliittymän asettelua suunniteltaessa selkeys ja väljyys on tärkeää. Silmän liikkeitä seuraten on esimerkiksi päätelty 58 %:n käyttäjistä etsivän katseellaan ostoskorin näytön oikeasta yläneljänneksestä ja 56 %:n käyttäjistä etsivän hakukenttää samasta paikasta. Mikäli tämä alue käyttöliittymästä on täynnä erilaisia ärsykeitä tai muutoin epäselvä, kasvaa riski käyttäjäpolun katkeamiseen ja konversioiden vähenemiseen. Toisaalta 88 % käyttäjistä käytti aikaansa katsellakseen sivuston tyyliin sopivia mainoksia. Tämä on potentiaalisten mainostulojen kannalta huomattava luku, joka palvelun tai sovelluksen tavoitteista riippuen saattaa hyvinkin ansaita oman painotuksensa sisällössä. Kaiken olennaisen, mainokset mukaan lukien, arvottaminen ja valitseminen helpottuu analysoimalla näkymän huomiosuhdetta eli tarjolla olevien linkkien tai muiden toimintojen määrää suhteessa sivuston konversiotavoitteiden määrään: esimerkiksi yhden painikkeen sisältävällä sivulla suhde on 1:1, mikäli sivulla ei ole muita tavoitteita. Keskimäärin verkkosivuilla tämä suhde on 20:1, eli toivotun käyttäjän suorittaman toiminnon lisäksi tarjolla on 19 vaihtoehtoa, käyttäjän huomiosta kilpailevaa toimintoa. Esimerkiksi hollantilaisen lahjakortteja myyvän verkkosivun konversiot ja tuotto kasvoivat 10,9 % ja 19,7 %, kun etusivulta poistettiin toinen rinnakkain asetelluista suurista painikkeista, ja käyttäjät ohjattiin ainoastaan ”tutustumaan kokemuksiin” ilman huomiosta kilpailevaa ”Osta nyt” -vaihtoehtoa ja niiden ympäriltä poistettiin toistakymmentä runsasväristä ja kuvallista linkkiä. [24; 25; 26.]

### 3.4 Sanavalintojen kriittisyys

Sanavalintojen ja erityisesti toimintakehotteiden (Call To Action, CTA) tärkeä asema käytettävyydestänsä perustuu niiden sijaintiin käyttäjäpolulla; huomattava osa käyttöliittymien osa- tai lopullisista mitattavista konversioista tapahtuu juuri tekstipainikkeen tai linkin avulla. Käyttäjäpolun ja käyttöistunnon ylläpitämiseksi erityisesti välivaiheiden toimintakehotteiden tulee olla selkeästi ja houkuttelevasti sanoitettuja, jotta käyttäjän informaatiiovainu (information scent) pysyy vahvana. Jokaisen käyttäjän toimintaa vaativan painikkeen tulee pitää yllä tai vahvistaa käyttäjän näkemystä siitä, että hän edistyy tavoitteessaan. Mikäli käyttäjä kokee tullessa johdetuksi harhaan tai edistyminen tuntuu liian hitaalta, hän olettaa helposti, ettei hänen hakemaansa asiaa löydy palvelusta tai että tavoitteen saavuttaminen ei ole vaivan arvoista. [23; 27; 28; 29; 30.]



Käyttäjä pyrkii käyttöliittymää käyttäessään suurimpaan mahdolliseen tehokkuuteen. Otsikoista, joita linkitkin usein ovat, luetaan ensimmäiset ja viimeiset kolme sanaa, joista ensimmäiset kaksi tarkasti, erityisesti jos linkit ovat listassa. Tällöin otsikolla ja linkillä on noin kuusi sanaa, joilla vakuuttaa käyttäjä siirtymään eteenpäin. Tunnelatauksen luomiseksi tehokkaita ovat aktiiviset verbit, kiireellisyyttä, lupauksia arvon tai edun saavuttamisesta sekä henkilökohtaisuutta kuvaavat sanat. ”Osta heti” kertoo asiakkaalle tuotteen saatavuudesta ja ostoprosessin, ja siten asiakaspolun, lopun lähenemisestä. ”Liity nyt!” antaa sekin lupauksen helppoudesta ja välittömyydestä, ja mikäli saatesanoina on jossain muodossa esimerkiksi ”tarjous päättyy”, on käyttäjälle asetettu ehto ja aikaraja. Sellaiset sanat kuin ”säästä” ja ”ilmainen” ovat nekin syystä yleisiä mainosbannereissa, sillä riskitön tai edullisempi toiminto vetoaa käyttäjiin. Asiakkaan tarpeeseen ja henkilökohtaiseen puhutteluun auttavat omistusliitteet, esimerkiksi ”Varaa omasi” tai ”Lisää varauksiini”. Oikean sanayhdistelmän tekemiseksi on myös tärkeää tunnistaa palvelun kohdeyleisö. Värikäs ja aggressiivinen kielenkäyttö saattaa toimia nuorelle tai uraorientoituneelle yleisölle, mutta vauvanvaatteita tai eläkevakuutuksia kauppaavan on todennäköisesti turvallisempaa käyttää pehmeämpää kieltä. [31; 32; 33; 34.]

### 3.5 Potentiaalisia työkaluja yleiseen A/B-testaukseen

#### Optimizely

Optimizely on Googlen entisten työntekijöiden luoma monipuolinen ja helppokäyttöinen optimointialusta. Käyttöliittymänä toimii selainpohjainen, lopputuloksen reaaliajassa näkyviin tulostava (WYSIWYG, What You See Is What You Get) editori, jolla verkkosivua ja sen kävijäliikenteeseen liitettäviä sääntöjä, tavoitteita ja muita toimintoja voidaan muokata dynaamisesti reaaliajassa. Kuten suurin osa kilpailevista alustoistakin, Optimizelyn seuranta perustuu evästeiden tarkkailuun JavaScriptin avulla. Optimizely aktivoidaan verkkosivulla kopioimalla JavaScript-koodileike testattavan sivun lähdekoodiin. Optimizely tallentaa sisällön omille palvelimilleen ja kirjaa sivuston koko käyttöhistorian käyttöönottohetkestä eteenpäin myöhempää tarkastelua varten. Optimizelystä on tarjolla ilmaisversio, joka mahdollistaa yhden verkkosivun, iOS-ohjelman ja Android-ohjelman, testaamisen kolmen käyttäjän toimesta. Ilmaisversio tarjoaa tilastojen lisäksi kävijäsegmentoinnin saapumisreitit, selaimen ja mobiilistatuksen perusteella. Ilmaisversio mahdollistaa myös yksinkertaisten integroinnin moniin lämpökartta- ja analytiikkatyökaluihin, kuten Google Analyticsiin, ClickTaleen ja KissMetricsiin. [35; 36.]

## Google Analytics ja Content Experiments

Google Analytics on Google-tunnuksilla käytettävä ilmainen, maksullisilla laajennuksilla varustettu verkkopohjainen verkkoanalytiikkatyökalu. Analytics tallentaa evästeitä hyödyntäen tietoa verkkosivuston käyttäjistä ja kykenee vastaamaan useimpiin yksinkertaisen A/B-testien tarpeisiin. Sen avulla on helppoa määrittää liikenteen määrä, asiakaspolut ja niiden katkeamispisteet, suosituimmat sivut, käytetyt laitteet ja asetettujen tavoitteiden täytyminen ja paljon muuta. Tavoitteita voi muokata vastaamaan omia suoriutumismittareitaan, kuten konversioita, keskimääräistä tilauksen arvoa tai kävijäkohtaista tuottoa. Muun Analyticsin tarjoaman liikennöintidatan tarjolla on myös Experiments-työkalu, joka mahdollistaa yksinkertaisen A/B-testauksen helposti. Tarjolla on tapa jakaa liikennettä prosentuaalisesti (tai dynaamisesti vaihtoehtojen suoriutumisen mukaan) vaihtoehtoisille sivuille ja kerätä ennalta määritettyjä tietoja, kuten session kestoa.

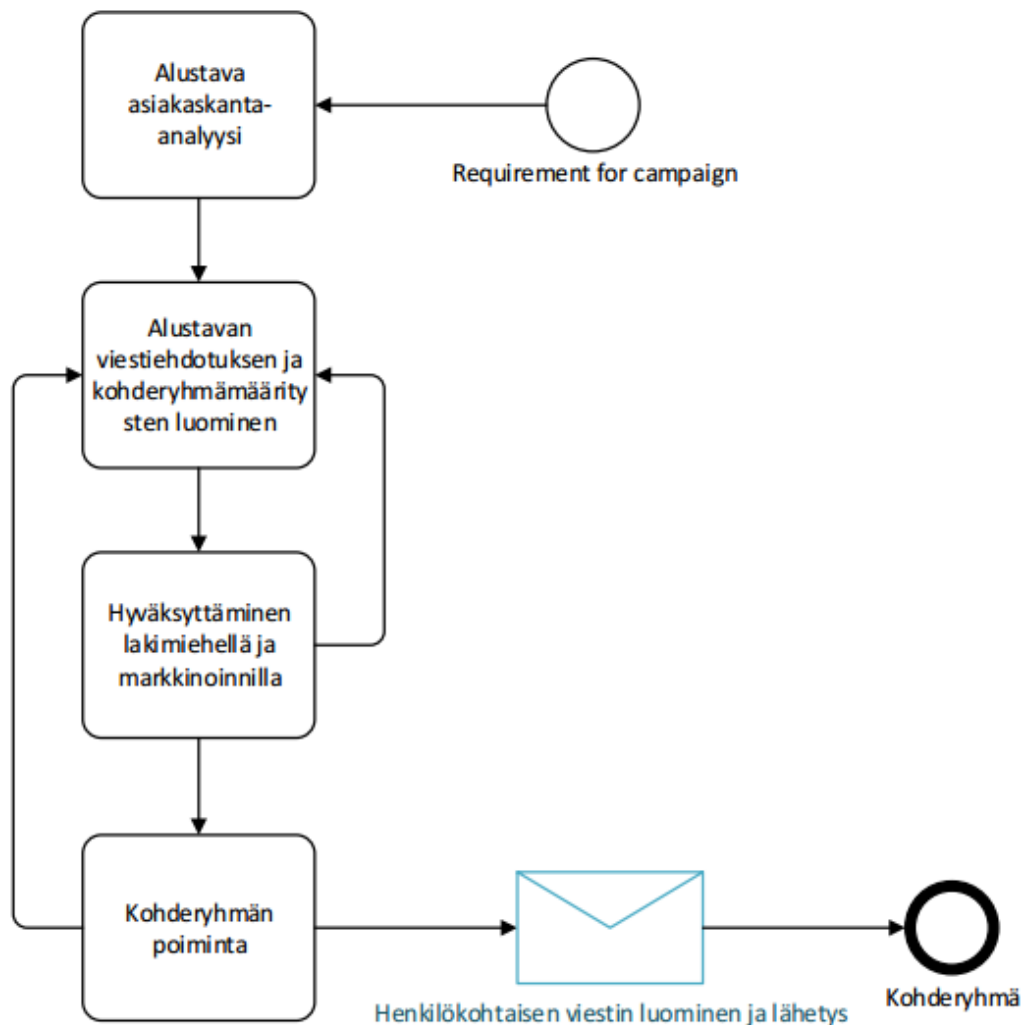
[37; 38.]

## 4 Testauksen suunnittelu ja toteutus

### 4.1 Tarve ja nykyinen prosessi

A/B-testauksen tarve syntyi olemassa olevien, toistuvien asiakaskontaktointien optimoinnin puutteesta. Olemassa olevien, tietyt kriteerit täyttävillä asiakkaille lähetettävien viestien kohderyhmien koot ja laatu sekä katselukerrat (hitrate) pysyivät tasaisesti samansuuntaisina, ja koska asiakaskanta oli rajallinen, oli luontevaa pyrkiä parantamaan viestinnän laatua rasittamatta asiakaskantaa enempää viestintämäärällisesti.

Tavoitteeksi muodostui kehittää suoraviivaisempi toimintamalli sisällön uusimiselle sekä hyödyntää käytössä olevia teknisiä alustoja tämän mahdollistamiseksi: tavallisessa tapauksessa täysin uuden viestin luominen jo olemassa olevaan kampanjaan vei noin kuusi viikkoa toimeksiannosta julkaisuun. Pyrkimyksenä oli karsia työvaiheita ja hyödyntää mahdollisimman paljon olemassa olevia datahakuja ja julkaisuväyliä, jolloin toistuvaksi uusiutuvaksi työmääräksi jäisi uuden sisällön tuotanto (viestin päivittäminen) ja kevyt tarkistuskierros sekä viestin kohderyhmäksi määräytyneen asiakaskannan jakaminen kahteen satunnaispoimittuun yhtä suureen ryhmään SAS Customer Intelligence Studiassa. [Kuva 2.]



Kuva 2. Kampanjan kehitysprosessi lähtötilanteessa.

Henkilökohtainen viesti lähtee tilaajalta, eli osastolta tai projektilta, jonka on määrä saada kontakti asiakkaaseen joko tiedoksiannon, kontaktipyynnön tai muun viestin muodossa. Tämän tarpeen pohjalta tehdään kannattavuusarvio, jossa huomioidaan myös asiakaskannan analyysi tavoitettavien ja tavoiteltavien asiakkaiden määrästä. Jos valo näyttää vihreää, eli mikäli kustannukset ja viestin hyöty ovat tasapainossa, viestiä lähdetään toteuttamaan. Tällöin tilaaja toimittaa suunnitteluryhmälle alustavat sisältövaatimukset ja aikataulun, minkä pohjalta luodaan viestiluonnos, mikä hyväksytetään moniportaisesti tilaajalla, lakimiehillä, markkinoinnilla ja mahdollisesti johtoryhmällä. Kohderyhmän tekninen toteutus eli asiakaspoiminta ja analyysit sekä julkaisun valmistelu tapahtuvat edellä kuvatun prosessin rinnalla painottuen ajallisesti projektin loppupäähän. Prosessi on raskas, sillä se vaatii osallistumista useilta tahoilta ja on kestoaltaan noin kuusi viikkoa.

Tämän jälkeen viesti julkaistaan, sen toimivuus varmistetaan, ja viestin oltua näkyvissä määritetyn ajan voidaan tuloksia tarvittaessa tarkastella.

Mikäli kyseessä on toistuva kampanja, voi analyytikko ajastaa sen, jolloin asiakkaita poimitaan samoilla kriteereillä viestin vastaanottajaksi ja viesti pysyy lähtökohtaisesti samaa, ellei sen muuttamiseen ole nähty tarvetta. Ajankäytöllisesti pisimmäksi osakokonaisuudeksi osoittautuivat sekä haastattelujen että käytännön tekemisen perusteella tekstisisällön tarkistuskierrokset: lakiperusteisten muutosten palautuminen lakimiehiltä odotutti muuta työryhmää, ja näiden muutosten sovittaminen haluttuun viestin sävyyn ja muihin tavoitteisiin vaati lisää yhteistyötä kampanjan tilaajan ja asiakasviestintävastavain kesken.

Asiakasyrityksen käyttämä sovellus on integroitu monikanavamarkkinointityökalu (SAS Customer Intelligence Studio), joka mahdollistaa monimutkaiset asiakastietokantojen haut graafisessa käyttöliittymässä sekä poimintojen julkaisun kaikkiin yrityksen käyttämiin markkinointikanaviin. Asiakaspoiminnoissa hyödynnetään yrityksen sisällä ennalta luotuja datahakuja (Data Item), niistä saatavaa tietoa hyödyntäviä laskelmia (Calculated Item) sekä Base SAS- ja/tai SQL-kieltä hyödyntäviä koodielementtejä (Code Node), jotka toimivat syötetyn koodin mukaisesti datan mistä tahansa käytettävissä olevasta tietokannasta. Näitä, ja esimerkiksi Ja- ja Tai-elementtejä yhdistelemällä asiakaskannasta kytetään luomaan helposti ymmärrettävä virtauskaavio, jolla poimitaan monimutkaiset ehdot täyttäviä asiakasjoukkoja kohdennettua markkinointia varten joko yksittäisajoina tai ajastettuina toistuvina kampanjoina. [39.]

## 4.2 Testin tavoitteet ja hypoteesi

Testin tavoitteet jakautuivat kahteen osaan: kokonaisvaltaiseen laadulliseen kehitykseen ja pilottitestinä luotavan yksittäisen kampanjan kannalta oleellisen ja siinä tarjolla olevan datan perusteella arvioitaviin tekijöihin. Laadullisen kehityksen kannalta olennaista tulisi olemaan parantunut asiakastytyväisyys, kasvanut viestin avausprosentti sekä kanavan kaikkien viestien vuotuisen keskiarvoon että toistuvissa kampanjoissa aiempaan versioon verrattuna kasvanut konversiosuhde ja viestien ymmärrettävyyden ja relevanttiuden lisääminen.

Asiakastytyväisyyden mittarina toimivat vuotuiset eri yritysten tuottamat kyselyt ja asiakasyrityksen omat asiakastytyväisyyskyselyt sekä verkkoalustalla tarjolla olevien palautekanavien kautta saatu asiakaspalaute. Viestien avausprosentin keskiarvoa seurataan sähköisesti sekä kokonaisuutena että kampanjakohtaisesti mittaamalla lähetettyjen viestien määrän ja auki klikattujen viestien suhdelukua. Konversiosuhde on kampanjakohtainen, sillä kullakin julkaistulla viestillä on oma tiedotuksellinen tai myynnillinen tavoitteensa. Suuressa osassa henkilökohtaisina viesteinä lähetettävästä yrityksen asiakasviestinnästä tavoitteena on vain tiedoksianto, jolloin konversioksi lasketaan viestin avaaminen, mutta toisissa kampanjoissa asiakas pyritään ohjaamaan tiettyyn verkkoalustan toimintoon tai kampanjasivustolle, jolloin mitattava toiminto muuttuu linkin tai napin painallukseksi. Henkilökohtaisten viestien ymmärrettävyyttä ja relevanttiutta mitataan paitsi avausprosentilla, myös viestipohjaan sisäänrakennetulla arviointilomakkeella, jossa asiakas voi pisteyttää subjektiivisesti lukemansa viestin relevanttiuden tähtiluokituksella asteikolla yhdestä viiteen.

Pilottikampanjan kannalta oleellisia, lyhyellä aikavälillä arvioitavia mittareita olivat viestin avaamisprosentti, joka suhteutettiin vaihtoehtoisten esitystapojen keskinäisen suoriutumisen lisäksi myös kyseisen kanavan pitkän ajan keskiarvoon eli 30 %:iin ja viestin arviointilomakkeelta saadut käyttäjäarviot eli tähtiluokitukset.

Rinnakkain julkaistut sisällöt oli kohdennettu nuorelle yleisölle, joten luonteva sisällön muokaus oli viestin kieliasu: kohderyhmää A lähestyttiin puhekielisesti ”Kohti ekaa omaa kotia” -otsikolla, kun taas kohderyhmälle B esitettiin myös leipätekstisisällöltään aikuismaisempi ”Huomattavia asuntoja ensiasunnon ostajalle” -otsikolla varustettu viesti. Oletuksena oli, että vaikka molemmissa viesteissä asiasisältö oli sama, saattaisi otsikon puhekielisempi muotoilu houkutella useamman nuoren lukijan avaamaan viestin.

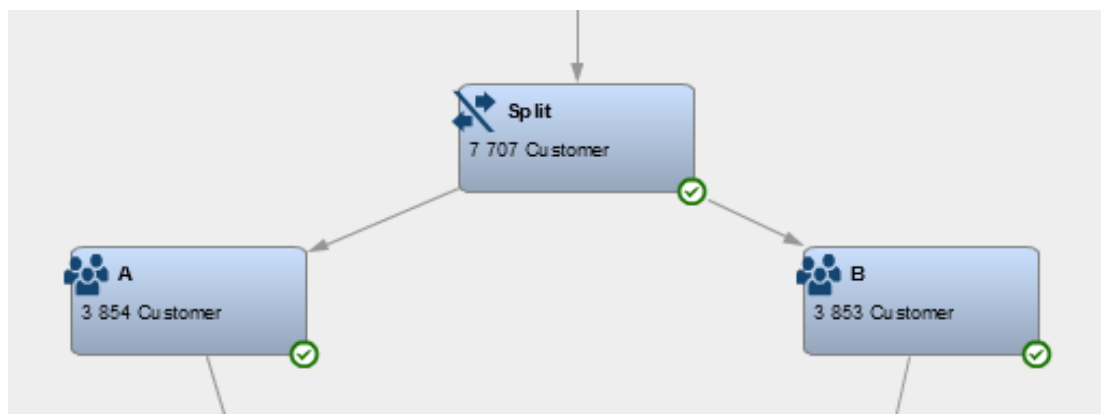
Prosessin uusimisessa A/B-testauksen kannalta tärkeintä oli rinnakkaisten tekstisisältöjen kerääminen ja hypoteesien määrittely mahdollisimman aikaisessa vaiheessa, jotta vaihtoehtoisia viestejä olisi testausta varten jo varastossa hyväksyttynä sekä tilaajan että asianajajien toimesta, jolloin jokaista testikierrosta varten ei tarvitsisi käydä prosessin pisintä vaihetta uudelleen läpi.

### 4.3 Toteutus

Olemassa olevan julkaisualustan, SAS Customer Intelligence Studion, ja jo pitkälle optimoidun työnkulun vuoksi suuria muokkauksia ulkonäköön ei edes pyritty tekemään, jotta uusi testausyksi olisi toteutettavissa mahdollisimman kevyesti mahdollisimman monelle toistuvasti julkaistavalle kampanjalle. Täten testattavaksi valikoitui puhtaasti viestin tekstuaalinen sisältö. Muokattava sisältö jakautui kolmeen osaan: viestin otsikko, leipäteksti ja mahdolliset painiketekstit. Asiakaspolun kannalta kriittisimmäksi pisteeksi Omniture Site Catalystillä ja SAS Visual Analyticstillä osoittautuivat asiakkaalle kirjautumisen jälkeen latautuvassa etusivunäkymässä esiintyvät otsikot [kuva 3, ”Avaa viesti”]. Nämä olivat ja tulisivat pysymään tyylittelyn suhteen keskenään homogeenisina viestin prioriteetin perusteella määrittyvää listausjärjestystä lukuun ottamatta, joten otsikon sisältö oli hypoteesissa avainasemassa. Kohderyhmän jakaminen kahteen oli lähtökohtaisesti erillainen tyypilliseen web-pohjaiseen A/B-testaukseen nähden; liikenteen ohjaamisen sijaan jouduttiin jako tekemään jo etukäteen kampanjapöimintää tehtäessä. Kaikki mukaan poimittuvien asiakkaiden pöimintakriteerit määriteltiin pöimintakaaviossa jo etukäteen, ja ne jaettiin Split-moduulilla [kuva 4] satunnaisesti kahteen yhtä suureen ryhmään. Kohderyhmän koko pöimintakriteereillä oli 7 707 henkilöä, joten odotettavissa ei ollut joidenkin asiantuntijoiden miniminäkin pitämiä katselukertamääriä. [2; Kuva 3; Kuva 4.]

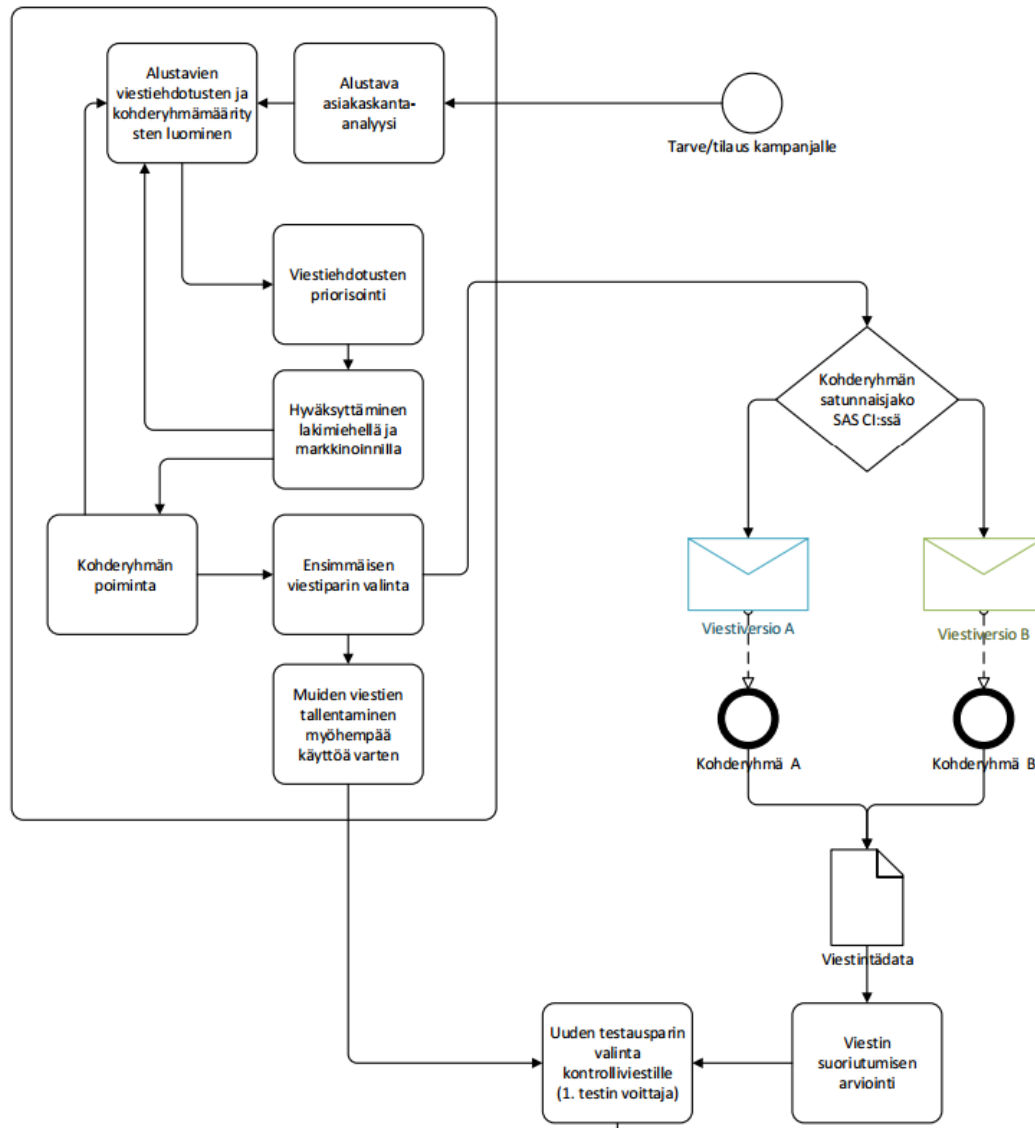


Kuva 3. Asiakasmassan vuotuinen keskimääräinen eroosio asiakaspöimintällä Guide-viestinnässä.



Kuva 4. Kohderyhmän jakomoduli SAS CI:ssä, jonka kautta asiakasryhmä jaettiin A- ja B-ryhmään.

Vaihtoehtoiset tekstisisällöt syntyivät asiakasviestintäkonsultin kanssa yhteistyössä luontevasti, ja rinnakkaiset viestit saatiin hyväksytyttyä samalla kertaa. Ylimääräistä aikaa ei siis kulunut itse sisällöntuotannossa juurikaan. Uuden prosessin [kuva 5] mukaisesti näitä vaihtoehtoisia viestejä voisi tulevaisuudessa myös ketjuttaa, jolloin tuloksia saataisiin nopeasti, joskin tilastotieteellisesti vähemmän pätevästi, eikä mahdollisesti pienemmissä kohderyhmissäkään riittävän käyttäjäjoukon kerääntyminen muodostuisi ongelmaksi. [5; Kuva 5.]



Kuva 5. Uusittu kampanjanluontiprosessi.

#### 4.4 Tulokset

Testikampanjan viestejä luettiin yleiseen henkilökohtaisten viestien verkkoliikenteen selvitysraportissa ilmoitettuun keskiarvoon (30 %, kuva 3, "Avaa viesti") verrattuna suuri määrä asiakkaita, 45 % kohderyhmässä olleista. Viesti A avattiin ja luettiin marginaalisesti useammin, 1 755 kertaa eli se generoi vain 20 lukukertaa enemmän kuin viesti B 1 735 mitatulla lukukerralla. Tilastollisesti tämä 1,13 %:n ero ei ole riittävän suuri osoittamaan sananvalintojen merkitystä asiakkaiden toiminnalle. Hypoteesi siis osoittautui vä-



räksi; testsignificance.com:n mukaan 95 % luotettavuuden saavuttamiseksi 45 %:n konversiolla ja poimitulla asiakasmassalla (toteutunut asiakasmassa 7 700 verrattuna taulukon 7 886) prosentuaalisen eron tulisi olla vähintään 2,2 % [Kuva 6].

"A" Conversion Rate	"B" Conversion Rate	Difference	Size Per Group	Total Participants	Confidence Level
45.00%	47.20%	2.20%	3,943	7,886	95%

Kuva 6. Testsignificance.com:n tavoitetaulukko. [4.]

Hieman yllättäen asiallisempi sävy viestissä B generoi positiivisemmän palautteen 3,5/5 tähden keskiarvolla joka oli 0,3 tähteä korkeampi verrattuna viesti A:n 3,2/5:n keskiarvoon. Tämä ero tosin selittynee todella pienellä arviointien määrällä, sillä yhteensä arvioita annettiin seitsemän kappaletta. Näin pienellä erolla ja vähäisellä arviointimäärällä on hedelmätöntä tehdä johtopäätöksiä, joten jatkotestaukselle on tarvetta. Myös viestin esitysaikaa, 30 päivää, tulisi koemuotoisesti pidentää, jotta kattavampi otanta saataisiin aikaiseksi. [Kuva 7.]

Communication name ▲	Communication code ▲	Contact date time ▲	Days ▲	Target group	%Read/TG	Rating avg	Total Ratings	Missing	Advice listed	Advice is read
D161_FirstOwnHomeA	1606070732203270	9. kesäkuuta 2016 07:37:41	30	3 859	45 %	3,20	5	494	3 195	1 735
D161_FirstOwnHomeB	1606070736187224	9. kesäkuuta 2016 07:37:41	30	3 854	45 %	3,50	2	486	3 229	1 735

Kuva 7. Testikampanjoiden suoriutumisasi data SAS VA:sta.

## 5 Yhteenveto

Kampanjan luonnin nykyprosessin kartoitus tehtiin haastattelemalla yrityksen asiakasviestintätiimin työntekijöitä ja seuraamalla heidän työntekeään tavoitteena paikantaa aikaa vievimmat ja työläimmät työosuudet ja sovittamalla uuden työtavan vaatimat lisätoimet niihin minimaalisella työkuorman lisäyksellä. Prosessi oli oletuksena kuuden viikon mittainen toimeksiannosta julkaisuun ja sisälsi ajallisesti huomattavan määrän eri osastojen välistä sisällön tarkastelua ja hyväksyttämistä ilman suunnitelmia mahdollisista päivityksistä, sisältövaihtoehtoista tai jatkotestauksesta. Ongelmakohtat kävivät ilmi melko nopeasti jo muutaman rinnakkaisen kampanjan kehitystä seuraamalla, ja niiden muokkaamisen suunnittelu oli prosessin visualisoinnin myötä helposti toteutettavissa. Rinnakkaisten tekstisisältöjen yhtäaikainen luominen ja hyväksyttäminen samalla kertaa vaikutti parhaalta vaihtoehdolta tulevia testejä ajatellen.

Prosessin ei-tekni­sen osuuden osalta tämä oletus piti paikkansa, ja uutta lähestymistä­pää eli vaihtoehtoisten viestien luomista jo alkuvaiheessa on mahdollista soveltaa tois­tu­vaksi aiottuihin kampanjoihin. Olemassa oleviin kampanjoihin vaihtoehtoisten sisältö­versioiden luominen tulee vaatimaan yhteydenpitoa tilaajan ja asianajajien kanssa, mutta tässäkin on mahdollista lisätä tehokkuutta tarkistuttamalla useampi variantti ker­ralla.

Yrityksen olemassa olevan järjestelmän infrastruktuurin vuoksi prosessin tekni­sen osuu­den rakenteelliset muutokset viestinnässä ovat haastavia toteuttaa, erityisesti lyhyellä aikavälillä. Yhteistyö ja läpinäkyvyys sivuston ylläpitäjien ja käyttäjien tai kehittäjien vä­lillä on byrokraattista ja aiheuttaa nykyisellään kohtuuttoman työmäärän pienten opti­mointitesti­en potentiaaliseen hyötyyn nähden. Esimerkiksi testauksessa ennalta tie­dossa olevan pullonkaulan ulkonäköön vaikuttaminen, eli henkilökohtaisen vinkin verk­koalustan etusivulle generoituvan linkin tyylittelyn uusiminen tai vinkkien yhdenmukai­suuden rikkominen, vaatisi huomattavan määrän neuvottelua graafisen ohjeiston laati­neiden ja niiden käytöstä huolehtivien osastojen kanssa. Näin raskaaseen ja monipuoli­seen järjestelmään olisi nykyiseltään työlästä rakentaa kapeammalle asiakasjouk­kolle nä­kymää muuttava liikenteen uudelleenohjaus tai käyttäjäkohtaisiin muuttujiin pohjautuva tyylitiedostojen käyttö.

Turvallisuustekijöiden määrä, turvallisuuden verifiointi ja verkkokehityksen jakautuminen useisiin osastoihin ja maihin pitkittäisi yksinkertaisenkin testin prosessia kohtuuttomasti. Siksi­pä, jos mahdollisuudet antaisivat myöten, toisin kattavamman sivuston koodia va­rioivan testauksen mahdollistavien työkalujen käyttöönoton esiin jo uutta järjestelmää suunniteltaessa, tai hakisin jo ajoissa kattavampaa hyväksyntää yksinkertaisten viestin sisäisten muutosten tekemiseen (vaikka niiden vaikutus olisikin jo ennustettavasti hu­omattavasti marginaalisempi kuin käyttäjäpolun kriittisimpään pudotukseen pureutumi­nen). Uskon kuitenkin, että tulevien tietotekni­sten ratkaisujen myötä kynnys moderniin testaukseen ja alustakehitykseen laskee ja suunnittelemani sisällöllinen testaus ja asi­akkaalle näkymättömän osuuden alustaminen toimivat hyvänä pohjana tulevaisuutta aja­tellon.

## Lähteet

1. Brian, Christian. 2012. Test Everything: Notes on the A/V Revolution. Verkkodokumentti. <[www.wired.com/2012/05/test-everything](http://www.wired.com/2012/05/test-everything)>. Luettu 25.01.2017.
2. Laja, Peep. 2014. 12 A/B Split Testing Mistakes I See Businesses Make All The Time. CXL Institute. <[conversionxl.com/12-ab-split-testing-mistakes-i-see-businesses-make-all-the-time](http://conversionxl.com/12-ab-split-testing-mistakes-i-see-businesses-make-all-the-time)>. Luettu 25.02.2017.
3. Paras, Chopra. 2011. Multivariate Testing 101: A Scientific Method Of Optimizing Design. Verkkodokumentti. <[www.smashingmagazine.com/2011/04/multivariate-testing-101-a-scientific-method-of-optimizing-design](http://www.smashingmagazine.com/2011/04/multivariate-testing-101-a-scientific-method-of-optimizing-design)>. Luettu 15.12.2016.
4. RJMetrics. A/B Test Significance – How Many Participants Do You Need? Verkkodokumentti. <<http://www.testsignificance.com/>>. Luettu 15.12.2016.
5. Oksanen, Jarno. 2015. 4 ohjetta matalan liikenteen sivuston A/B-testaukselle. Verkkodokumentti. <[www.superanalytics.fi/blogi/4-ohjetta-matalan-liikenteen-sivuston-ab-testaukselle](http://www.superanalytics.fi/blogi/4-ohjetta-matalan-liikenteen-sivuston-ab-testaukselle)>. Luettu 24.11.2016.
6. Brian, Christian. 2012. The A/B test: Inside the technology that's changing the rules of business. Verkkodokumentti. <[www.wired.com/2012/04/ff\\_abtesting/](http://www.wired.com/2012/04/ff_abtesting/)>. Luettu 11.11.2016.
7. Hanington, Jenna. 2012. The ABCs of A/B Testing. Verkkodokumentti. <<http://www.pardot.com/blog/abcs-ab-testing/>>. Luettu 15.1.2017
8. Sonberg, Jaan M. 2013. How to Build a Strong A/B Testing Plan That Gets Results. Verkkodokumentti. <[conversionxl.com/how-to-build-a-strong-ab-testing-plan-that-gets-results/](http://conversionxl.com/how-to-build-a-strong-ab-testing-plan-that-gets-results/)>. Luettu 20.9.2016
9. Rohrer, Christian. 2014. When to Use Which User Experience Research Methods. Verkkodokumentti. <[www.nngroup.com/articles/which-ux-research-methods/](http://www.nngroup.com/articles/which-ux-research-methods/)>. Luettu 17.12.2017.
10. Nielsen, Jakob. 2012. A/B Testing, Usability Engineering, Radical Innovation: What Pays Best?. Verkkodokumentti. <[www.nngroup.com/articles/ab-testing-usability-engineering/](http://www.nngroup.com/articles/ab-testing-usability-engineering/)>. Luettu 2.1.2017.
11. McLeod, Saul. 2008. Qualitative vs. Quantitative. Verkkodokumentti. <<http://www.simplypsychology.org/qualitative-quantitative.html>>. Luettu 10.1.2017.
12. Cardello, Jennifer. 2013. Three Uses for Analytics in User-Experience Practice. Verkkodokumentti. <[www.nngroup.com/articles/analytics-user-experience/](http://www.nngroup.com/articles/analytics-user-experience/)>. Luettu 10.1.2017

13. Birkett, Alex. 2016. When to Run Bandit Tests Instead of A/B/n Tests. Verkkodokumentti. <<http://conversionxl.com/bandit-tests/>>. Luettu 18.12.2016.
14. Gershoff, Matt. 2012. Balancing Earning with Learning: Bandits and Adaptive Optimization. Verkkodokumentti. <[conductrics.com/balancing-earning-with-learning-bandits-and-adaptive-optimization/](http://conductrics.com/balancing-earning-with-learning-bandits-and-adaptive-optimization/)>. Luettu 18.12.2016.
15. White, John Myles. 2012. Bandit Algorithms for Website Optimization: Developing, Deploying, and Debugging. Sebastopol: O'Reilly Media.
16. Wainwright, Kayla. 2016. A/B Testing: Cut through your KPIs by knowing your ultimate goal. Verkkodokumentti. <[www.marketingexperiments.com/blog/general/cut-through-kpis.html](http://www.marketingexperiments.com/blog/general/cut-through-kpis.html)>. Luettu 14.1.2017.
17. Wright, Lynne. 2014. What are the best KPIs to measure a split test? Verkkodokumentti. <[www.awa-digital.com/blog/kpis-to-measure-a-split-test](http://www.awa-digital.com/blog/kpis-to-measure-a-split-test)>. Luettu 14.1.2017.
18. Kirkpatrick, David. 2012. Marketing 101: What is conversion? Verkkodokumentti. <<http://sherpablog.marketingsherpa.com/marketing/conversion-defined/>>. Luettu 11.11.2016.
19. Loranger, Hoa. 2014. Redesign Your Website? Don't Ditch Your Old Design So Soon. Verkkodokumentti. <[www.nngroup.com/articles/redesign-competitive-testing/](http://www.nngroup.com/articles/redesign-competitive-testing/)>. Luettu 15.12.2016
20. Nielsen, Jakob. 2012. Usability 101: Introduction to Usability. Verkkodokumentti. <[www.nngroup.com/articles/usability-101-introduction-to-usability/](http://www.nngroup.com/articles/usability-101-introduction-to-usability/)>. Luettu 2.1.2017
21. Nielsen, Jakob. 2008. How Little Do Users Read? Verkkodokumentti. <[www.nngroup.com/articles/how-little-do-users-read/](http://www.nngroup.com/articles/how-little-do-users-read/)>. Luettu 28.12.2016.
22. Held, Stefan. 2015. SEO: How Much Text Do You Need On Your Website Pages? Verkkodokumentti. <[www.linkedin.com/pulse/seo-how-much-text-do-you-need-your-website-pages-stefan-held](http://www.linkedin.com/pulse/seo-how-much-text-do-you-need-your-website-pages-stefan-held)>. Luettu 28.12.2016.
23. Lee, Kevan. 2016. The Ideal Length of Everything Online, Backed by Research. Verkkodokumentti. <[blog.bufferapp.com/the-ideal-length-of-everything-online-according-to-science](http://blog.bufferapp.com/the-ideal-length-of-everything-online-according-to-science)>. Luettu 28.12.2016.
24. Nagpal, Mohita. 2014. Reducing Choices on Landing Page Increased Revenue by 19%. Verkkodokumentti. <[vwo.com/blog/reducing-choices-landing-page-increased-revenue-19/](http://vwo.com/blog/reducing-choices-landing-page-increased-revenue-19/)>. Luettu 13.1.2017.
25. Gardner, Oli. Conversion-Centered Design. Verkkodokumentti. <<http://unbounce.com/conversion-centered-design/attention.html>>. Luettu 13.1.2017.

26. Nielsen, Jakob & Pernice, Kara. 2010. Eyetracking Web Usability. New Riders.
27. Nielsen, Jakob. 2003. Information Foraging: Why Google Makes People Leave Your Site Faster. Verkkodokumentti. <[www.nngroup.com/articles/information-scent/](http://www.nngroup.com/articles/information-scent/)>. Luettu 10.1.2017.
28. Nielsen, Jakob. 2004. Deceivingly Strong Information Scent Costs Sales. Verkkodokumentti. <[www.nngroup.com/articles/wrong-information-scent-costs-sales/](http://www.nngroup.com/articles/wrong-information-scent-costs-sales/)>. Luettu 10.1.2017.
29. Pernice, Kara. 2014. A Link is a Promise. Verkkodokumentti. <[www.nngroup.com/articles/link-promise/](http://www.nngroup.com/articles/link-promise/)>. Luettu 10.1.2017
30. Sherwin, Katie. 2015. "Learn More" Links: You Can Do Better. Verkkodokumentti. <[www.nngroup.com/articles/learn-more-links/](http://www.nngroup.com/articles/learn-more-links/)>. Luettu 14.1.2017.
31. Creative Bloq Staff. 2015. Why words are a design issue. Verkkodokumentti. <[www.creativebloq.com/ux/why-words-are-design-issue-91516852](http://www.creativebloq.com/ux/why-words-are-design-issue-91516852)>. Luettu 14.1.2017.
32. Nielsen, Jakob. 2009. First 2 Words: A Signal for the Scanning Eye. Verkkodokumentti. <[www.nngroup.com/articles/first-2-words-a-signal-for-scanning/](http://www.nngroup.com/articles/first-2-words-a-signal-for-scanning/)>. Luettu 14.1.2017.
33. Olyslager, Paul. Call to Action Buttons Part 4: Message. Verkkodokumentti. <[www.paulolyslager.com/call-to-action-buttons-message/](http://www.paulolyslager.com/call-to-action-buttons-message/)>. Luettu 15.1.2017.
34. Mykyte, Samantha. 2015. The 25 Best Words to Use in Your Call-To-Action Buttons. Verkkodokumentti. <[blog.wishpond.com/post/103290853633/the-25-best-words-to-use-in-your-call-to-action](http://blog.wishpond.com/post/103290853633/the-25-best-words-to-use-in-your-call-to-action)>. Luettu 15.1.2017.
35. Optimizely. Verkkodokumentti. <[en.wikipedia.org/wiki/Optimizely](http://en.wikipedia.org/wiki/Optimizely)>. Wikipedia. Luettu 3.1.2017
36. Optimizely. Verkkodokumentti. <[www.crunchbase.com/organization/optimizely#/entity](http://www.crunchbase.com/organization/optimizely#/entity)>. Crunchbase. Luettu 3.1.2017.
37. Google Analytics. Verkkodokumentti. <[support.google.com/analytics/answer/6014873?hl=en](http://support.google.com/analytics/answer/6014873?hl=en)>. Google Support. Luettu 3.1.2017.
38. Google Analytics. Verkkodokumentti. <[searchbusinessanalytics.techtarget.com/definition/Google-Analytics](http://searchbusinessanalytics.techtarget.com/definition/Google-Analytics)>. Techtarger. Luettu 3.1.2017.
39. SAS CI. Verkkodokumentti. <[www.softwareadvice.com/marketing/sas-customer-intelligence-profile/](http://www.softwareadvice.com/marketing/sas-customer-intelligence-profile/)>. Softwareadvice. Luettu 3.1.2017.

40. Laja, Peep. 2015. Statistical Significance Does Not Equal Validity (or Why You Get Imaginary Lifts). Verkkodokumentti. <[conversionxl.com/statistical-significance-does-not-equal-validity/](http://conversionxl.com/statistical-significance-does-not-equal-validity/)>. Luettu 15.1.2017.
41. Schippers, Kees. 2012. How large should your A/B test sample size be?. Verkkodokumentti. <[vwo.com/blog/how-to-calculate-ab-test-sample-size/](http://vwo.com/blog/how-to-calculate-ab-test-sample-size/)>. Luettu 1.2.2017.
42. Mount, John. 2014. A clear picture of power and significance in A/B tests. Verkkodokumentti. <[www.win-vector.com/blog/2014/05/a-clear-picture-of-power-and-significance-in-ab-tests/](http://www.win-vector.com/blog/2014/05/a-clear-picture-of-power-and-significance-in-ab-tests/)>. Luettu 15.1.2017.
43. Lipsman, Andrew. 2007. Cookie-Based Counting Overstates Size of Web Site Audiences. Verkkodokumentti. <[www.comscore.com/chi/Insights/Press-Releases/2007/04/comScore-Cookie-Deletion-Report](http://www.comscore.com/chi/Insights/Press-Releases/2007/04/comScore-Cookie-Deletion-Report)>. Luettu 16.1.2017.
44. Woodman, Nathan. 2012. Cookie Deletion and Upper-Funnel Targeting. Verkkodokumentti. <[www.clickz.com/cookie-deletion-and-upper-funnel-targeting/44068](http://www.clickz.com/cookie-deletion-and-upper-funnel-targeting/44068)>. Luettu 16.1.2017.