

**Master's Thesis**

Masteruddannelsen i IT, Linjen i Interaktionsdesign og Multimedier

Ved Aalborg universitet i IT - vest samarbejdet

**Effekt af instruktionstyper i**

**Single vs Pair-Wise testing**

*i en opgavebaseret remote asynkron usability test*

**Aalborg Universitet**

**Vejleder:** Jan Stage

**Afleveringsdato:** 20. Oktober 2014

**Opgavens anslag:** 85648 (tegn med mellemrum)

**Studerende:** Kári Gundur Mortensen

---

([karigundur@gmail.com](mailto:karigundur@gmail.com))

Studrenr. 20110106

## Abstract

The purpose of this master thesis was to evaluate the effect of different instruction types in a remote usability test with predefined task assignments where testusers were set up as single-user and pairs using think a-loud technique and constructive interaction.

The test setup was conducted as a remote asynchronous evaluation on an online mail and communications system. 54 teachers in the Faroese public school system were placed in two major groups as 18 single users and 18 pairs and then again divided into three minor groups getting different instructions types. The pairs were furthermore grouped into two groups of males, 2 x females and 2 x combination male-female.

The results showed that single users identified more usability problems than pairs in all instruction types. According to instruction types the results showed that both single-users and pairs that recieved a conceptual definition of usability problems identified more usability problems than those who were instructed by means of examples. Furthermore, the results showed that the gender combination of pairs did also have an impact on the results recieved.

## Keywords

Remote testing; usability testing; asynchronous testing; unmoderated testing, automated testing, task assignments; instruction types; inductive instruction, deductive instruction, empirical study; pair-wise testing; single-user testing; think a-loud; constructive interaction.

## INDHOLDSOVERSIGT:

|                                                                         |           |
|-------------------------------------------------------------------------|-----------|
| <b>ABSTRACT .....</b>                                                   | <b>2</b>  |
| KEYWORDS .....                                                          | 2         |
| <b>INDLEDNING .....</b>                                                 | <b>5</b>  |
| <b>PROBLEMFORMULERING.....</b>                                          | <b>8</b>  |
| <b>HVAD ER BRUGBARHEDS TEST .....</b>                                   | <b>9</b>  |
| <b>REMOTE USABILITYTESTS .....</b>                                      | <b>12</b> |
| ASYNKRON REMOTE USABILITY TESTING:.....                                 | 12        |
| <b>RELATED WORK.....</b>                                                | <b>13</b> |
| UCI USER-REPORTED CRITICAL INCIDENTS .....                              | 13        |
| THINK ALOUD TEKNIK.....                                                 | 16        |
| CONSTRUCTIVE INTERACTION .....                                          | 17        |
| INDUKTIV/DEDUKTIV INSTRUKTION .....                                     | 18        |
| <b>EKSPERIMENT: .....</b>                                               | <b>21</b> |
| <b>DATA ANALYSE:.....</b>                                               | <b>31</b> |
| <b>RESULTAT: .....</b>                                                  | <b>33</b> |
| TESTDELTADELSE.....                                                     | 33        |
| <i>Testdeltagelse enkeltbruger.....</i>                                 | <i>34</i> |
| <i>Testdeltagelse par.....</i>                                          | <i>34</i> |
| <i>Indleveret svar, kønsfordeling.....</i>                              | <i>35</i> |
| PROBLEMLISTE .....                                                      | 37        |
| <i>Indberettede usabilityproblem, single vs pair.....</i>               | <i>39</i> |
| <i>Fordeling af indberetninger i forhold til parsammensætning:.....</i> | <i>42</i> |
| <b>DISKUSSION: .....</b>                                                | <b>43</b> |

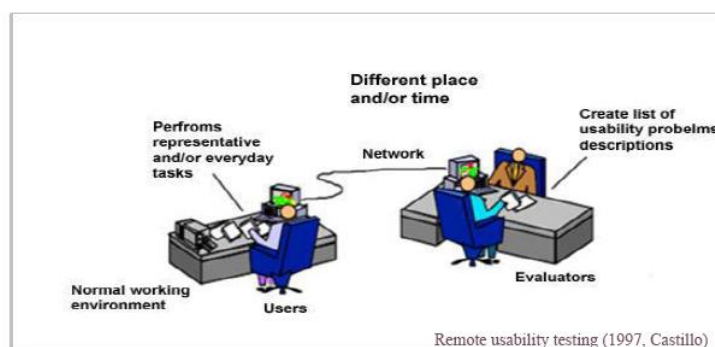
|                                                                        |           |
|------------------------------------------------------------------------|-----------|
| <i>Effekt af instruktionstype.....</i>                                 | <i>44</i> |
| <i>Effekt af datakvalitet, single vs pair-wise .....</i>               | <i>45</i> |
| <i>Forskel i testresultater i forhold til parsammensætning:.....</i>   | <i>46</i> |
| BEGRÆNSNINGER OG FUTURE WORK .....                                     | 47        |
| <i>Testbruger commitment: .....</i>                                    | <i>49</i> |
| <i>Future Work .....</i>                                               | <i>51</i> |
| <b>KONKLUSION: .....</b>                                               | <b>51</b> |
| <b>LITTERATUR: .....</b>                                               | <b>54</b> |
| <b>BILAG: ALLE BILAG LIGGER I DIGITAL FORM PÅ VEDHÆFTEDE USB-NØGLE</b> |           |
| <i>Bilag 1: E-post til skolelederne</i>                                |           |
| <i>Bilag 2: E-post til testdeltagere</i>                               |           |
| <i>Bilag 3: FirstClass usabilitytest – Data</i>                        |           |
| <i>Bilag 4: Problemliste</i>                                           |           |
| <i>Bilag 5: Mean user task</i>                                         |           |
| <i>Bilag 6: user time use</i>                                          |           |
| <i>Bilag 7: Interview Pair ind-ded instr</i>                           |           |
| <i>Bilag 7a: Interview Pair ind-ded inst, lydfil</i>                   |           |
| <i>Bilag 8: User_Prefs</i>                                             |           |

## Indledning

Usabilitytests er et værktøj, som man har benyttet sig meget af igennem tiderne. Ved at udføre usabilitytests med testpersoner, som repræsenterer de reelle brugere af produktet, får man indblik i, hvordan brugeren opfører sig i givne situationer, i forhold til det produkt man vil teste. Der er mange forskellige variationer af testsituationer, som man kan sætte op ved disse tests. En af de mest benyttede opsætninger er et usability-laboratorie, som er et specielt indrettet lokale med spejle, kameraer og udstyr til lydoptagelse.

Før i tiden var det en stor og økonomisk udfordring at gennemføre en usabilitytest. Med den teknologiske udvikling, er det blevet muligt for store og små virksomheder at teste deres produkter, uden at det økonomisk og resourcemæssigt skal vælte et projekt. For 12-15 år siden begyndte man derudover at udføre remote usability tests. Det gav en helt ny dimension i testsituationen. Den konventionelle laboratorietest er defineret som en situation, hvor testevaluator sidder sammen med testdeltager og gennemfører testen med spørgsmål og svar. Remote usability tests bliver defineret som en situation, hvor testevaluators er separerede fra testbrugerne i rum og /eller tid. (1998,Castillo et al).

I de senere år, har den teknologiske udvikling har gjort det meget lettere at udføre remote usability tests. Disse remote tests bliver sammenlignet med konventionel laboratorietests, hvor testevaluator sidder sammen med testpersonen og udfører nogle opgaver (2008,Tullis). Mange internetbaserede kommunikationsværktøjer kan bruges til at optage video, mousetrack og lyd, så lab situationen i princippet kan være hvor som helst. Denne metode bliver kaldt Remote Synkron Usabilitytest (2010,Soucy). Ordet “remote” forklarer at der er en fysisk distance mellem testperson og evaluator (1997,Castillo)



Ved denne metode bliver det meget lettere at afvikle tests, da man slipper for den logistiske udfordring at transportere testdeltagere til og fra et usabilitylaboratorium. Tidsmæssigt sparer man megen tid på gennemførslen af testen, men efterarbejdet af de indsamlede data kræver den samme analyseindsats, som ved en konventionel laboratorietest. Der skal bruges lige så megen tid på gennemgang og analyse af videomateriale, som ved lab tests.

En anden metode, som der stadig bliver videre udviklet på, er Asynkron Remote Usabilitytest. Der bliver brugt andre ord på samme testes rundt om i verden. Kyle Soucy, fra UX Matters, bruger ordet Unmoderated Remote User Testing i en artikel fra 2010 (2010,Sousy), og Tom Tullis bruger terminologien Automated Usability Testing for det samme(2008,Tullis). Det, der karakterisere Asynkron Remote Usability testing, er at testdeltagere og evaluatore er separerede både i tid og rum (2009, Bruun, Stage)(2007, Andreasen et al),(2008, Tullis),(2010,Sousy). Dvs. at når testen afvikles, udfører testpersonerne nogle pre-definerede opgaver på den lokation de befinder sig på, og på det tidspunkt det passer dem bedst. De har den frihed at betemme selv, når de udfører testen. Andreasen et al viser i deres forskning, at asynkrone metoder er betydelig mere tidskrævende for testpersonerne og giver betydeligt færre problemidentifikationer end ved de konventionelle metoder (2007, Andreasen et al). På den anden side viser forskningsresultater fra Bruun og Stage at i forhold til indsatsen i konstruktionen af remote tests, er det ret interessante at kigge nærmere på.

På trods af at testresultaterne er betydelig færre end ved konventionelle lab tests, er de billigere at konstruere, udføre og analysere. Man kan lave flere remote tests for den samme sum penge, og dermed får man også mere ud af det at teste (2009,Bruun et al). Den primære styrke ved asynkron remote usabilitytest er potentialet at indsamle data fra ret mange deltagere på et kort tidsrum. Da alle testbrugere i princippet kan teste på samme tid på internettet, er antallet af testbrugere kun begrænset af ressourcer fra testleders side (2008, Tullis). Martin et al påviser i deres artikel, at asynkron remote testing er et godt alternativ til laboratorietests, men det ene udelukker ikke det andet (2014,Martin et al). De har lavet en asynkron remote usabilitytest på et udvalgt website og sammenlignet det med en laboratorie test af samme website. De kommer frem til at

remote testing er lige godt eller bedre end laboratorietests, ud fra et resourcemæssigt og økonomisk perspektiv. De konkluderer i deres artikel, at remote asynkron usabilitytesting af websites er en mere cost-effektiv metode end laboratorie baserede “tænke højt” tests (2014, Martin et al).

Da vi arbejder med asynkrone testmetoder, vil det være interessant at undersøge, hvilke metoder virker bedre end andre. Kan man, uden modifikationer, overføre de konventionelle synkrone metoder over på asynkrone? I usability labs og ved remote synkrone test, har testleder direkte kontakt til testdeltagere, og kan give instrukser og oplæring i, hvad man forstår ved usability problemer. Der er blevet forsket i, hvordan kommunikationen mellem testleder og testperson kan foregå i en asynkron test (2009, Bruun et al). De har undersøgt effekten af forskellige metoder og sammenlignet resultaterne med hinanden og med en konventionell test. Alle tre asynkrone tests blev udført hjemme hos testbrugerne på deres egne computere. De fik introduktionmateriale tilsendt, som de skulle sætte sig ind i før testen og derefter udføre den.

Als et al sammenligner “Tænke-højt” metoden med “Constructive Interaction” i artiklen “Comparison of Think-Aloud and Constructive Interaction in Usability Testing with Children”. Der sætter de børn sammen i par på forskellige måder for at undersøge hvordan de responderer på pre-definerede opgaver i en lab test. (2005, Als et al) Deres eksperiment involverede 60 børn, delt op i enkelt brugere og nogle par, som kender hinanden og andre par, som ikke kender hinanden. Deres resultater viser, at parrene, som kendte hinanden, identificerede flere usabilityproblemer end de andre par og enkeltbrugerne. Endvidere kunne de påvise, at parrene, som kendte hinanden, brugte langt mindre energi til at løse opgaverne end enkeltbrugerne og parrene, som ikke kendte hinanden inden testen (2005, Als et al).

Min plan er at lave en asynkron remote usabilitytest bygget på de metoder der allerede er blevet afprøvet Anders Bruun og Jan Stage (2012, Bruun, A; Stage, J ). I deres forskning, præsenterer de tre instruktionstyper, som testpersonerne får præsenteret inden testen. Deres resultater viser, at kombinationen af induktiv og deduktiv instruktion giver bedst forståelse for hvordan testpersoner opfatter et brugbarhedsproblem. I mit

akademiske spørgsmål vil jeg undersøge om kvaliteten og mængden af testresultater ændres, ved at sætte nogle testpersoner sammen parvis, i forhold til de resultater, som man får fra testpersonerne, der løser opgaverne enkeltvis. Min hensigt er ikke at gentage den undersøgelse, der allerede er blevet lavet af Bruun og Stage, men lave en kombination af deres eksperiment og det eksperiment med single vs pair-wise testing, som Als et al har arbejdet med.

Det vil sige at mit setup bliver en remote asynkron usability test, hvor testbrugerne vil få nogle pre-definerede opgaver, som de skal gennemgå på et bestemt website. Testpersonerne vil blive delt op i enkeltbrugere og par, hvor de igen vil blive delt op i tre grupper og vil blive præsenteret for tre forskellige instruktionstyper. Parrene bliver derudover delt op efter køn.

## Problemformulering

**Jeg vil i denne opgave undersøge effekten af instruktionstyperne deduktiv, induktiv og kombineret deduktiv + induktiv instruktion i en opgavebaseret remote asynkron usability test, hvor testbrugerne er delt op som enkeltbrugere og par. Kan man påvise forskel i testresultaterne i forhold til sammensætning af par, som er opdelt efter køn?**

Effekten af remote opgavebaseret usabilitytest kan måles på indkomne anvendelige testsvar indsamlet fra enkeltpersoner og fra par. Instruktionen vil blive forskellig til både enkelt personer og par i henhold til problemformuleringen. Derudover vil testdeltagerne blive instrueret i “Tænke højt” metoden (2008, Rubin), hvor de bliver bedt om at tænke aktivt over hvad det er de gør under selve testen og “Constructive Interaction” (1986, Miyake, N), så de bedre kan formulere deres svar, når de støder på noget brugerproblem. Parrene snakker med hinanden, og deres dialog er en variation af “Tænke højt”, som bliver kaldt “Constructive Interaction” (2005, Als et al)(1986, Miyake, N.) Vil der kunne



ses forskel på testresultaterne mellem disse to grupper, og vil instruktionstypen har en indvirkning på resultaterne.

FirstClass er et kanadisk produceret mail og dokument håndterings software, som bliver brugt som primær kommunikations form på alle folkeskoler på Færøerne. Udover at sende og modtage mails, kan man oprette konferencer og grupper, hvor f.ex. lærere tilføjer sine elever i bestemte fag. Skoleadministrationen har sit pædagogiske arkiv på systemet og de forskellige skoler kan være sammen om diverse konferencer.

Det er det canadiske firma OpenText FirstClass, som udvikler systemet. Alle skoleelever og lærere i folkeskolen og gymnasium er registrerede brugere. Derudover er der flere højere læreinstitutioner også med i systemet. Antallet af brugere ligger omkring 20000, hvor der er ca. 3000 unike indlogninger dagligt. For ca. et år siden lancerede OpenTEXT et nyt web interface, baseret på html5 standarden. Det nye interface blev aktiveret på den færøeske FirstClass server i dec. 2013, og bliver brugt dagligt af en del lærere på folkeskolen. Jeg vil i denne opgave lave en usability test på det nye web interface for at se, hvordan brugerne får løst deres opgaver ud fra generelle usability kriterier (2008, Rubin). Da systemet ikke er udviklet inhouse, er det begrænset hvad man kan gøre med de problemer og bugs, som måske dukker op. De problemer, som kommer frem ved testen, vil blive indrapporteret til udviklingsfirmaet i Canada via vores konsulent i Danmark. Men primært er planen med denne test at få afdækket, hvor brugerne har problemer og derfra sætte ind med introduktions videoer og undervisning, så vi får udnyttet systemet fuldt ud og samtidigt får mange glade og tilfredse brugere.

## Hvad er brugbarheds test

Begrebet brugbarhedstest bliver brugt mere om alle de teknikker, som bliver brugt når man evaluerer et system. Jack Rubin bruger begrebet brugbarhedtesting som en process, *”that employs people as testing participants who are representative of the target audience to evaluate the degree to which a product meets specific usability criteria.”* (Rubin, 2008)

Der er mange metoder til at måle brugbarhed. Den mest enkle og brugbare metode er at lave brugertest. (2012,Nielsen) Brugbarhedstest er et værktøj, som man metodisk bruger at teste forskellige designs og funktionaliteter. Nielsen introducerer tre komponenter, som skal være med i en brugertest: 1. Brugere, som repræsenterer de reelle brugere, som systemet er beregnet til. 2. Testbrugerne skal til at udføre nogle repræsentative opgaver med dit design. 3. Observer brugernes adfærd, hvor de lykkes, og hvor de har problemer med brugerfladen. Lad brugerne få lov til at tale frit og bland dig ikke i deres snak. (2012, Nielsen) Det er vigtigt at brugerne bliver testet individuelt og at de får lov til at udføre opgaverne uden nogenform for indblanding. Hvis testleder hjælper eller vejleder testpersonen i at løse opgaverne, kan man ikke bruge testresultaterne til noget(2012, Nielsen)

Man kan spørge hvorfor det er nødvendigt at teste et design, produkt eller system. Fra et forretnings perspektiv, er det ikke et spørgsmål om man inden for produktfirmaet kan finde ud af at bruge et produkt, men om brugeren, som produktet er tilegnet, kan forstå at bruge det. Profitabiliteten ved et produkt øges ved større forståelse fra brugeren. Har brugeren success med et produkt, bliver han automatisk ambassadør for produktet, da han taler positivt om det i sine omgivelser. Og har man et produkt, som brugeren kan lide, er det stor sandsynlighed at samme bruger køber næste version af produktet eller han køber også et andet produkt fra firmaet.(Rubin, 2008) På markedet er der mange produkter, der ligner hinanden og udfører næsten de samme funktionaliteter for brugeren. Har man et produkt, som har en høj brugeroplevelse, er sandsynligheden for at de vælger netop det fremfor konkurrentens produkt. Man opbygger også et loyalitetsforhold til brugeren, som gør at brugeren føler sig sikker og godt behandlet af firmaet.

I vores dagligdag tænker man nok ikke så meget over brugbarhed som begreb. På engelsk hedder det "User Experience". I bogen hos David Benoy, "Designing Interactive Systems" citerer han McCarthy og Wright's begreb "user experience". *"Experiences are about how we bring artefacts and services into our lives and accept them"* (citeret fra 2010, Benoy). Alle brugere møder begrebet uden at være klar over det, fordi hver gang de betjener et eller andet produkt, det være sig en tv fjernstyring, opvaskemaskine, kaffemaskine, computer etc., så oplever de at produktet udfører det de forventer af

det.(2010,Benyon) Rubin berører også emnet, hvor han siger, at hvis produktet derimod ikke fungerer efter hensigten, eller at man ikke kan finde ud af at betjene det, bliver man gal eller frustreret over det. (2008,Rubin) Har man har investeret mange penge i produktet f.eks. en computer eller tv/stereo, vil man gerne afsætte noget tid til at lære at bruge produktet og i fleste tilfælde går det godt. Men man kan også opleve det besværligt at betjene produktet efter længere tids pause. Et sådant produkt bliver sat bagerst på hylden for aldrig at blive brugt igen, eller de avancerede funktioner på f.eks. en tv kommer aldrig til sin ret, fordi man gider ikke finde ud af det alligevel.(2008, Rubin)

Rubin definerer brugbarhed som fravær af frustration, når man bruger det. Et produkt eller en tjeneste er rigtig brugbar, da *“the user can do what he or she wants to do the way he or she expects to be able to do it, without hindrance, hesitation, or questions”* (Rubin,2008). Han sætter op nogle kriterier, som retningslinier på at et produkt er brugbart. Anvendelighed (Usefulness), Ydeevne (Efficiency), Effektivitet (Effectiveness), Learnability og Satisfaction

Rolf Molich bruger nogle andre om om det samme, når han definerer brugervenlighed. Et produkt er brugervenligt, når det er: (2007, Molich),(ISO 9241)

#### **Nyttigt**

Egnet til at løse de opgaver, som **brugere** ønsker at løse.

#### **Let at lære**

Indlæringstiden er den tid det tager brugere at lære at løse bestemte opgaver ved hjælp af produktet.

#### **Let at huske**

Genindlæringstiden er den tid det tager brugere, som har været væk fra produktet eller som kun anvender det en gang imellem, at løse bestemte opgaver.

#### **Effektivt at bruge**

Effektiviteten er den hastighed hvormed brugere løser bestemte opgaver i forening med produktet. Effektiviteten afhænger bl.a. af dialogens udformning, svartid, fejlfrekvens og kvaliteten af fejlmeddelelser.

#### **Rart at bruge**

Den subjektive tilfredshed er den tilfredshed med produktet som brugere udtrykker i professionelle spørgeskemaer eller interviews.

I mit eksperiment vil ovenstående begreber blive præsenteret i den deduktive instruktion, hvor testbrugerne får denne konceptuel definition af usability-begreberne, som er defineret af Molich.(2007, Molich) Men for at gøre det lettere for testbrugerne, vil definitionerne blive præsenteret som en inverteret definition på samme måde, som Bruun og Stage gør i deres eksperiment (2012, Bruun og Stage). Ved at sætte en negation foran definitionen, får brugerne lettere ved at afgøre, hvad der forstås ved et usabilityproblem. Mere om dette i punktet "Eksperiment".

## Remote Usabilitytests

### Asynkron remote Usability testing:

I forhold til *synchron* Remote Usability testing, som er at brugere og evaluatore udfører testen på samme tid, men er ikke placerede på samme location, bliver asynkron Remote Usabilitytest defineret som en situation, hvor testdeltagerne er separerede fra testevaluator i både tid og rum. Test-evaluator definerer nogle testspørgsmål og laver et remote testsetup, som deltagerne logger sig ind på via f.eks. en internet browser. Testdeltagerne udfører selve testen på et tidspunkt, som passer dem bedst. Testdeltagerne er uafhængige af testevaluator, og skal ikke tilrettelægge deres testafvikling i samarbejde med testevaluator. De får dog et forudbestemt tidsrum, hvor testen skal være afviklet i. Testresultaterne lagres på en database, hvor de bliver eksporteret ud på et excel regneark og systematiseret og analyseret af testevaluator.

Sammenlignet med konventionel lab test, har forskningen vist at asynkron remote usability tests har nogle ulemper i form af manglende feed back fra testbrugere, at det er tidskrævende for testbrugerne og at det giver færre kvalitative resultater (2005, Olmsted, Gill). Bruun og Stage kommer til et resultat (2012, Bruun, A; Stage, J.), hvor de delvis er enige med Olmsted og Gill om at det giver færre resultater, men også finder frem til at arbejdsindsatsen til en opgavebaseret remote asynkron usability test er langt mindre og lettere end ved konventionel lab-test, samt at den giver ret gode testresultater i forhold til indsats.

Ved asynkron remote usabilitytests er kommunikationen mellem evaluators og testdeltager meget vigtig, men også meget svær. Da kommunikationen er asynkron, kan det hurtigt ske, at testdeltager ikke forstår eller misforstår opgaven. Det kan resultere i at man ikke får nok af anvendelige testsvar ind. Forskingresultater har vist at måden testdeltagerne bliver instrueret på har stor betydning på testresultaterne (2012, Bruun, A; Stage, J.)

## Related Work

Den sværeste del i en asynkron remote usabilitytest, er hvordan testleder og testpersoner kommunikerer med hinanden. Bruun et al redegør i deres artikel *“Let Your Users Do the Testing: A Comparison of Three Remote Asynchronous Usability Testing Methods”*, hvordan kommunikationen mellem testleder og testperson kan foregå (2009, Bruun et al). De har undersøgt effekten af forskellige metoder og sammenlignet resultaterne med hinanden og med en konventionell laboratorie test. De definerer tre slags asynkrone tests:

- User-reported critical incident (UCI)
- Forum-based online reporting and discussion (Forum)
- Diary-based longitudinal user reporting (Diary)

## UCI User-reported Critical Incidents

UCI står for User-reported Critical Incidents. José Castillo (1997, Castillo) nævner at der findes enighed blandt forskere på en definition på en kritisk hændelse, trods der er flere variationer på hvordan man indsamler og analyserer kritiske hændelser:

*“A critical incident is an event observed within task performance that is a significant indicator of some factor defining the objective of the study”* (1997, Castillo) .

UCI teknikken bliver også omtalt i artiklen *“Investigating the Effectiveness of Applying the Critical Incident Technique to Remote Usability Evaluation”*. Jennifer A. Thompson nævner RECITE (the REMote Critical Incident TEchnique), at være en tilpasning af den

samme teknik, som Castillo (1997) udviklede. Critical incidents er interaktion med et system, som både kan være let og svært at udføre, og som både kan føre til ret gode eller ret dårlige resultater. (1999, Thompson)

Oprindelsen til denne metode kommer fra flyvetekniske studier under anden verdenskrig gennem optagelser og interviews med piloter, som oplevede kritiske hændelser under flyvning. Gennem tiderne er metoden blevet udviklet og i halvfemserne definerede Carroll, Koenemann-Belliveau, Rosson, and Singley (1993) og Koenemann-Belliveau et al. (1994) en kritisk hændelse som:

*“a critical incident in the context of human-computer interaction as an event that stands out during usability evaluation (e.g., a major breakdown during task performance). In that context, empirical usability evaluation, particularly formative evaluation, depends heavily on observing and interpreting critical incidents during user task performance”.*  
(1997, Castillo)

Generelt kan man sige at user-reported critical incident teknikken er en brugbarheds evaluering til kritiske hændelser, som opfylder følgende kriterier:

- Data er centreret omkring kritiske hændelse, som opstår under opgaveløsning
- Rigtige brugere løser opgaverne.
- Brugere er placeret i deres normale arbejdsmiljø.
- Data bliver indsamlet dagligt på en billig måde.
- Direkte interaktion mellem bruger og evaluator er ikke nødvendig.
- Data er af høj kvalitet og deraf er det let at konvertere dem til usability problemer.

(1999, Thompson)

Castillo udviklede videre et on-line critical incident reporting værktøj, som bestod af et sæt nøje strukturerede spørgsmål, som skulle indfange forskellige faktorer, som typisk var associeret med den kritiske hændelse.

| Question                                                                                                   | Contextual Factor(s)                                          | Format               |
|------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|----------------------|
| Explain what you were trying to do when the critical incident occurred                                     | Specific beginning of the task                                | Text box             |
| Describe what you expected the system to do just before the critical incident occurred.                    | User expectations                                             | Text box             |
| In as much detail as possible, describe the critical incident that occurred and why you think it happened. | Detailed description of the critical incident                 | Text box             |
| Describe what you did to get out of the critical incident.                                                 | Ability to recover and specific ending of a critical incident | Text box             |
| Where you able to recover from the critical incident?                                                      | Ability to recover                                            | Yes/No               |
| Are you able to reproduce the critical incident and make it happen again?                                  | Critical incident reproducibility                             | Yes/No               |
| Indicate in your opinion the severity of this critical incident.                                           | Critical Incident Severity/Task Frequency                     | 5-point Rating Scale |

I overensstemmelse med UCI metoden instruerer Bruun og Stage deres testbrugere i at rapportere alle usability problemer, som de finder under test af et website. Selve testen er en asynkron remote usabilitytest. (2012, Bruun, A; Stage, J,)

*“In the context of formative usability evaluation, a critical incident is an occurrence during user task performance that indicates something (positive or negative) about usability”.* (1997, Castillo)

Testdeltagerne bliver instrueret i at indberette alle negative kritiske tilfælde, så snart man støder på dem. Når man indberetter et tilfælde, skal testbrugerne svare på følgende spørgsmål for hvert enkelt tilfælde (2012, Bruun, A; Stage, J,) :

- Hvilken opgave var du i gang med, da det kritiske problem opstod?
- Hvad er navnet på vinduet, hvor det kritiske problem opstod?
- Forklar hvad du prøvede at udføre, da det kritiske problem opstod.
- Forklar hvad du forventede af systemet lige inden det kritiske problem opstod.
- Prøv at beskrive så detaljeret som muligt det kritiske problem som opstod og giv et bud på hvorfor det opstod
- Forklar hvad du gjorde for at komme forbi det kritiske problem.
- Kunne du komme videre efter at det kritiske problem opstod?
- Kan du gentage det kritiske problem og få det til at ske igen?
- Gi en bedømmelse af hvor alvorligt du karakteriserer dette kritiske problem.

Disse testspørgsmål vil jeg genbruge i mit eksperiment, som bliver beskrevet senere i denne opgave under punktet “Eksperiment”.

### **Forum-baseret online rapportering og diskussion** (2012, Bruun, A; Stage, J.)

Forum-baseret online rapportering og diskussion er en metode, hvor testdeltagerne skal installere et program, udføre nogle forudbestemte opgaver og notere de usabilityproblemer de støder på, og derefter afinstallere programmet igen. Derefter skal de sende deres noter til et online forum og diskutere det med de andre testdeltagere. Selve forummet er bliver hermed til data indsamlingsværktøjet.

### **Diary-based longitudinal user reporting (Diary)** (2012, Bruun, A;Stage,J)

Testdeltagerne skal løse opgaver og skrive om deres oplevede usabilityproblemer. Der er ingen formel struktur over svarudformningen. Testdeltagerne får opgaver tilsendt via mail, som skal svares igen via mail. Over fem dage får de opgaver at løse, som skal svares via mail efter den første og sidste dag. Data indsamlingen bestod af indsendte emails.

I mit eksperiment, vil jeg kun koncentrere mig om UCI, User-reported Critical Incident, metoden. Mit eksperiment går ud på at undersøge effekten af de forskellige instruktionstyper og ikke sammenligning af dataindsamlingsmetoder, som nævnt ovenfor.

Testdeltagerne vil få et sæt opgaver, som de skal gå igennem og skrive ind i en online formular, som er designet til formålet, alle deres usability problemer de kommer på ved gennemgang af opgaverne.

### **Think aloud teknik**

I min testopsætning bliver testdeltagerne sat sammen enkeltvis og i par. De testepersoner, som går igennem testen som enkeltpersoner, bliver introduceret for “Tænke højt” teknikken. Thinking-aloud technique bliver meget anvendt i laboratorie test setup (2008,Rubin). I den tilsendte mail til testbrugerne, beder jeg dem hele tiden om at være meget opmærksomme på, hvad det er de gør, mens de løser opgaverne. Da de så støder på et problem, bedes de indberette problemet. Til det formål , er der designet en



svarformular, hvor ovenstående spørgsmål (på side 16) skal besvares. Tænke højt metoden betyder, at testpersonerne skal sige det de tænker, mens testleder lytter med. Det kan føles underligt og intimiderende at bede remote testdeltagere om at snakke med sig selv, mens de udfører testen, så det bliver de ikke direkte bedt om, men de bliver alligevel bedt om at være opmærksomme på, at tænke igennem hvad det er de laver, så de bedre kan formulere sine usability indberetninger.

Clayton Lewis og John Rieman beskriver metoden i deres bog “Task Centered User interface Design, A Practical Introduction”: (1994, C. Lewis, J. Rieman )

*The basic idea of thinking aloud is very simple. You ask your users to perform a test task, but you also ask them to talk to you while they work on it. Ask them to tell you what they are thinking: what they are trying to do, questions that arise as they work, things they read. You can make a recording of their comments or you can just take notes. You'll do this in such a way that you can tell what they were doing and where their comments fit into the sequence.*

Teknikken er meget simpel og er meget brugt i konventionel lab tests, hvor den hjælper testevaluator at “læse” tanker hos testdeltagerne. I modsætning til lab test, er det ikke muligt for testevaluator at følge op på denne interne dialog hos testdeltageren, da evaluator ikke er til stede ved en asynkron remote test. Jeg vælger at introducere denne tænke-højt teknik for at give testdeltagerne et værktøj, så de lettere kan koncentrere sig og holde fokus under testen, og lettere kan formulere sig, når de støder på brugerproblemer og skal svare på gennem nogle spørgsmål vedrørende det.

## Constructive Interaction

Konstruktiv interaktion er en metode, som bliver anbefalet, når man sætter testdeltagere sammen i par (2005, Als et al ; 1993, Nielsen, ). Metoden er en variation af think-aloud teknikken, som involverer to testpersoner i et samarbejde om opgaveløsning. Personerne snakker sammen under testen, de har en normal dialog, hvor de overfor hinanden ikke kun forklarer, hvad de funderer over i forbindelse med opgaven, men også hvorfor de

tænker som de gør. Situationen gør det usynlige synligt, de interagerer konstruktivt med hinanden. (1986, Miyake, N.)

I arbejdet hos Als et al, hvor testpersoerne er børn, er der sat sammen tre sessioner. I 1. session er børnene enkeltbrugere, 2. session er de klassekammerater og kender hinanden gennem min. 5 år og 3. session kender de ikke hinanden og går på forskellige skoler. Testen går ud på at de skal teste en ny model af en mobil telefon. Deres resultater viser, at konstruktiv interaktion er en anvendelig metode i brugbarhedstests med børn. Resultaterne viser også at børn, der kender hinanden, finder flere brugbarhedsproblemer end hvis de ikke kender hinanden. De kommer også frem til at tænke-højt metoden, selv om den kan føles kunstig for testdeltagerne, udmærket kan bruges af 13-14 årige børn, da de godt forstår at bruge den i udførslen af testen.

Mine testdeltagere, som er sat sammen i par, bliver bedt om at snakke sammen under testen og forklare for hinanden, hvad de gør og hvorfor de gør det. Også bliver de bedt om at skiftes om at sidde ved skærmen og udføre opgaverne, så det ikke kun er den ene af parrene, som er aktiv tester, mens den anden noterer. Når parrene bliver gjort opmærksomme på det, undgår man at den ”dygtigere” testperson udfører testen ”alene”, og den anden kun følger med som observatør og ikke aktivt deltager i testen. Da testen er asynkron remote, er det umuligt at måle på om de nu faktisk gør som instrueret. Men det vil jeg gå ud fra at de gør.

### **Induktiv/deduktiv instruktion**

Ved induktiv og deduktiv instruktion forstås måden man præsenterer information for elever/studerende. (Beck, 1986.)(Prince et al,2006) Ved *deduktiv* instruktion går man fra det almene til det specielle. Dvs. at eleverne først skal sætte sig ind i det princip, der skal læres og derefter anvende princippet i konkrete tilfælde. Informationen vil blive overleveret fra lærer til elev i form af definitioner og principper, som eleverne ud fra disse slutter sig til regler og almene begreber. Lærerens opgave er at formidle stoffet, træne eleven med anvendelsen af det og kontrollere rigtigheden.

Den *induktive* metode går den anden vej, hvor læreren. I stedet for at begynde med generelle principper, definitioner og regler, tilrettelægger han forskellige undervisningssituationer og eksperimenter, som eleverne så ud fra egen opfattelsesevne tilegner sig en forståelse af, hvordan det hænger sammen, og får klarlagt og afdækket de generelle principper, regler og definitioner. Princippet er, at man går ud fra enkelte fakta og iagttagelser, og ud fra disse lader man eleverne slutte sig til regler og almene begreber.

I læringssituationen er det forskelligt fra individ til individ, hvilken instruktionstype fungerer bedst. Nogle lærer bedst ved induktiv instruktion, mens andre får mere ud af deduktiv instruktion. I alle læringssituationer bruger læreren/instruktøren begge typer instruktion og meget gerne en kombination af begge. (1986, Beck)(2006, Prince et al)(2012, Bruun, A;Stage,J)

I artiklen “The Effect of Task Assignments and Instruction Types on Remote Asynchronous Usability Testing” sætter Bruun og Stage op denne hypotese:

*H3: A) Inductive instructions cause users to identify more problems than by deduction. B) Instructions based on a combination of deduction and induction cause users to identify more problems than those receiving pure inductive or deductive instructions.*(2012, Bruun, A;Stage, J)

I deres undersøgelse bliver testdeltagerne opdelt i forskellige grupper med forskellige slags instruktioner og opgaver. En gruppe får induktiv instruktion i hvad usability problemer er. Dvs. de får præsenteret et par eksempler fra andre sites, og ud fra disse eksempler forventes det af testdeltagerne, at de har en forståelse for hvordan de opfatter noget som usability problem, og i hvilken kategori man kan klassificere det. En anden gruppe får en deduktiv instruktion, idet de fik præsenteret definitionen (2007, Molich) af hvad usabilityproblem er og noget mere introduktion for bedre forståelse. En tredje gruppe får en kombination af begge typer instruktion i form af en tekstsider. Disse tre grupper bliver yderligere opdelt efter om de får opgaver at løse eller om de bare skal indrapportere usabilityproblemer i deres daglige brug af systemet. Som kontrolgruppe bliver der lavet en klassisk lab test. Måden, som testdeltagerne indberetter deres usabilityproblemer bliver gjort efter UCI metoden.

Deres resultater af disse comparative tests hos Bruun og Stage, viser at testbrugere, som modtager predefinerede opgaver, løser signifikant flere opgaver og identificerer signifikant flere usability problemer og større problemforståelse end testdeltagerne, der ikke skulle løse nogen opgave, men kun indrapportere usability problemer, når de stødte på dem i deres daglige brug af systemet.

Ydermere viser deres undersøgelse at typen af instruktion har stor påvirkning af resultaterne. De brugere, der fik induktiv instruktion, fandt signifikant flere usabilityproblemer end de testbrugere, der fik en deduktiv konceptionel definition af usabilityproblemidentifikation.

Det interessante ved asynkrone tests er, at selv om disse resultater synes at være utilfredsstillende sammenlignet med den klassiske lab test, så blev disse resultater frembragt med en arbejds-indsats, som svarer til ca. 13% af, hvad den klassiske lab test kræver (2009, Bruun et al). Dette giver metoden en attraktiv vinkel, da man kan få flere billige og hurtige tests for færre penge. I et udviklingsforløb er det ofte meget relevant at få lavet billige brugbarhedstest, selv om de ikke kan siges at være helt fuldkommen, da man indsamler værdifulde oplysninger på det stadie man er på udviklingsmæssigt. Og disse remote tests, som Bruun og Stage har testet, har tendens til at komplimentere hinanden i den forstand, at kombinerer man dem, giver de en cost-effektiv løsning. Martin et al (2014) har lavet lignende undersøgelser, hvor de sammenlignet kvaliteten mellem laboratorie teting og asynkron remote teting i forbindelse med website usability testing. Deres resultater viser, at remote testing kan sammenlignes med lab testing i kvalitet. Endvidere kommer de frem til, at når man sammenligner de to metoder, remote vs lab, i forhold til tidsforbrug, indsats og økonomi, så er remote testing en mere effektiv løsning. Remote testen var registreret at bruge 15 dage mindre med direkte evaluator involvering, hvilket resulterer i mindre udgifter per problem i forhold til lab test. (2014, Martin et al) Tullis (2008, Tullis) kommer også frem til, at i modsætning til traditionel usability testing, er der ingen signifikant øgning af udgifter eller ressourcer per bruger i asynkron remote testing.

Men asynkron usabilitytesting er ikke altid hensigtsmæssigt at bruge, fordi der er nogle begrænsninger i den. Tidlige explorative undersøgelser, hvor en gentagende dialog med testdeltagerne foregår, er ikke egnet til en asynkron remote test (Tullis,2008).

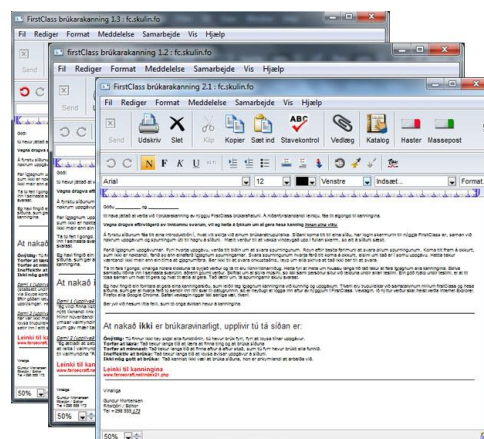
## Eksperiment:

Mit eksperiment går ud på at teste effekten af tre forskellige instruktionstyper, deduktiv, induktiv og deduktiv + induktiv instruktionstyper i en testopsætning, hvor testdeltagerne er delt op i enkeltbrugere og par. Herefter Single vs Pair Wise.

I mit eksperiment vælger jeg at benytte de samme instuktioner, som Bruun og Stage benytter i deres ekperiment (2012,Bruun og Stage).

Den deduktive instruktion bygger på at testbrugerne bliver præsenteret for en konceptuel definition af Usabilitybegreberne. Begreberne, som Molich definerer dem(2007, Molich), bliver præsenteret i instruktionen, som testbrugerne får tilsendt med mail. Bruun og Stage bruger disse usability begreber med den variation, at de inverterer begreberne og tilføjer et par linier til hver definition, så begreberne kommer til at være følgende:

- **Unyttigt:** Du kan ikke finde de dokumenter eller funktioner, du har brug for til at løse dine opgaver.
- **Svært at lære:** Det tager lang tid at lære at finde ting og bruge siden.
- **Svært at huske:** Det tager lang tid at finde ting, som du tidligere har brugt eller fundet.
- **Ineffektivt at bruge:** Det tager lang tid at løse bestemte opgaver med sitet.
- **Utilfredsstillende at bruge:** Sitet føles ikke behageligt at bruge, sa det er ikke en fornøjelse at arbejde med det.



Den induktive instruktion bygger på at testdeltagerne bliver præsenteret for to eksempler på usabilityproblem. De får ikke nogen definition af dem, men det forventes at de, ud fra eksemplerne, selv skal kunne komme frem til en generel opfattelse af, hvad et

usabilityproblem er. Nedenstående vises de to eksemplerne, som jeg venligst har lånt fra Bruun og Stage (2012, Bruun og Stage):

**Eksempel 1 (oplevet på Facebook):** "Jeg ville finde en liste over mine venner på Facebook. Jeg gik ind på websitet og via forsiden (kaldet 'Nyheder') valgte jeg 'Venner' (placeret under profilbilledet). Jeg havde før benyttet et lignende link til at få vist listen over mine venner, men i stedet fik jeg vist en side, hvor jeg kunne tilføje nye venner via Skype kontakter, Messenger osv., men mine nuværende venner fremgik ikke. Jeg trykkede flere gange på 'Venner' linket, men med samme resultat, hvilket var temmelig frustrerende. Efter lang tids trykken frem og tilbage mellem forskellige menuer trykkede jeg på mit profilbillede, hvilket sendte mig væk fra siden 'Nyheder' og hen til mine profiloplysninger. Her trykkede jeg igen på 'Venner' linket, som viste den ønskede liste. Det var meget forvirrende at to links, der ser ens ud, gør noget helt forskelligt."

**Eksempel 2 (oplevet under brug af Microsoft Word 2007):** "Jeg ville indtætte en fodnote i mit dokument. Jeg forventede at finde funktionen under menuen 'Indsæt', men det var ikke muligt at finde denne funktion her. Jeg forsøgte også at lede i menuerne 'Sidelayout' og 'Gennemse', dog uden held. Det var lidt irriterende. Jeg forsøgte at løse problemet ved at bruge hjælpefunktionen og fandt frem til menuen 'Referencer', hvor funktionaliteten til at indsætte en fodnote var. Det var ulogisk at, at noget man indsætter i et dokument ikke er placeret under 'Indsæt'."

Til den kombinerede instruktionstype får testdeltagerne begge typer instruktion i et dokument. Først den konceptuelle definition og derefter de to eksempler.

I eksperimentet er der 18 single-users og 18 par. Seks Single-users bliver præsenteret for deduktiv instruktion, seks får induktiv instruktion og seks får kombineret instruktion. Parrene vil også blive delt op efter instruktionstyper. Derudover har jeg delt parrene op i køn og fordelt instruktionstyperne jævnt, således at alle sammensætningerne af køn får en instruktionstype af hver art.

|          |    |    |    |
|----------|----|----|----|
| Deduktiv | MM | MK | KK |
| Induktiv | MM | MK | KK |
| Ded+ind  | MM | MK | KK |

*Parsammensætning: MM = mand-mand, MK=mand-kvinde, KK= kvinde-kvinde*

Testdeltagerne skal teste brugergrænsefladen på den nye FirstClass grænseflade. De bliver præsenteret for syv enkle opgaver. Da et kriterie er, at de allerede skal være registrerede brugere af FirstClass systemet og forventes at have styr på de basale funktionaliteter, som ligger heri, er opgaverne konstrueret ud fra det. Testen bliver primært lavet for at se, hvordan brugerne forstår at bruge den nye grænseflade.

Opgaverne, som testdeltagerne skal udføre, er følgende:

**1. Profil**

På det nye web interface skal du redigere i dine profil oplysninger. Find frem til din profil og indtast eller opdater dine oplysninger. Indsæt også et profilbillede.

**2. Signatur**

Du skal lave en signatur til dine mails. Find frem til stedet, hvor du gør det og sæt op din signatur, så den slår igennem som standard på alle nye mails.

**3. E-mail**

Opret en mail. Vedhæft et dokument til mailen og indsæt et billede, så det ses sammen med teksten. Send mailen til [usetest@skulin.fo](mailto:usetest@skulin.fo).

**4. Kalender**

På venstre side er der en kalender. Gå ind i den og opret en ny aftale til i morgen kl. 14 til 16. Giv den overskriften "usetest" og inviter brugeren [usetest@skulin.fo](mailto:usetest@skulin.fo) med i aftalen.

**5. Arbejdsområde**

På vestre side er der et menupunkt med navnet "Workspace" eller "Arbejdsområde". Der inde finder du File Storage, hvor du skal uploade et dokument fra din computer ind i denne mappe.

**6. Konference**

I Konferencen "FL-limir" er en mappe med titlen 'Overenskomster etc.' Find mappen og hent dokumentet "Løn- og indplaceringsoversigt - okt.2007-0kt.2009.xls". Åbn dokumentet og luk det igen. Da du har gjort dette, er opgaven fædig.

**7. Oprette konference**

På forsiden kan man oprette egne konferencer. Find frem til, hvor det kan gøres, opret en konference og giv den et tilfældigt navn. Åbn derefter den ny-oprettede konference og tilknyt to personer til den. De to brugere skal være "gundur umsitari" og "royndarbrúkarin hjá Gunduri", som du finder i brugeroversigten. Giv dem tilladelser som moderator.

Brugerne bliver instrueret i at løse opgaverne efter User - reported Critical Incident, UCI metoden(1997,Castillo). Når de gennemgår opgaverne og så støder på et problem, skal de besvare et sæt spørgsmål, som er tilgængeligt på samme websted som opgaverne. Støder

de på i flere problemer i samme opgave, bedes de om at besvare spørgsmålene om igen, for hvert tilfælde. Hvis de ikke støder på noget i opgaverne, bedes de om at skrive nr. og navn på opgave og teksten “*Intet problem*” ind i det første svarfelt. Svarene registreres anonymt i en database.

Eksperimentet bliver designet som en 2x3 matrix. En gruppe er med individuelle brugere, som i test situationen bliver instrueret i at “snakke” med sig selv efter tænke-højt metoden (2008, Rubin), og en anden gruppe med par, som bliver instrueret at snakke sammen i test situationen efter Constructive Interaction metoden(2005, Als et al) .

Grupperne af enkeltbrugere og par bliver fordelt til intruktionstype, så 1/3 får deduktiv instruktion, 1/3 induktiv instruktion og 1/3 kombination af deduktiv og induktiv instruktion.

Tabel 1

|                                 | Enkelt bruger | Par |
|---------------------------------|---------------|-----|
| Deduktiv instruktion            | 6             | 6   |
| Induktiv instruktion            | 6             | 6   |
| Deduktiv + induktiv instruktion | 6             | 6   |
|                                 | 18            | 18  |

I forbindelse med Constructive Interaction, er parrene sat sammen i kombinationen mand-mand, kvinde-kvinde og mand-kvinde, for at se om det har nogen indvirkning på testresultaterne i denne gruppe testbrugere. Når de går igennem testen, bliver de derudover bedt om at skiftes til at sidde ved skærmen og løse opgaverne, så man undgår den situation, at det kun er den ene af testdeltagerne, der løser opgaver mens den anden kun følger med.



Tabel 2

|                                 | MM | MK | KK |
|---------------------------------|----|----|----|
| Deduktiv instruktion            | 2  | 2  | 2  |
| Induktiv instruktion            | 2  | 2  | 2  |
| Deduktiv + induktiv instruktion | 2  | 2  | 2  |

54 personer (27 kvinder og 27 mænd) i alderen 25 - 65 år deltog i testen. Alle deltagerne er aktive lærere i folkeskolen. De blev udvalgt tilfældigt på 9 forskellige folkeskoler jævnt fordelt på Færøerne. Skoleinspektørerne på de 9 udvalgte skoler, fik tilsendt en mail fra mig, hvor selve projektet og hensigten med testen blev præsenteret (*Bilag 1*).

I mailen bad jeg skoleinspektørerne om at finde 9 lærere, som frivilligt ville deltage i testen. Der blev understreget, at de frivillige ikke skulle repræsentere selve skolen, men at de kun repræsenterede sig selv.

Tre dage efter at mailen blev sendt ud, ringede til personligt til alle skoleinspektørerne, og fik en snak med dem om projektet. De var alle villige at hjælpe mig med at finde testdeltagere, og da der var gået op imod to uger, var der tilmeldt det ønskede antal deltagere.

Folkeskolen på Færøerne er organiseret i tre undervisningstrin, indskoling, mellemtrin og udskoling. Testpersonerne blev valgt tilfældigt ud fra disse undervisningstrin, og sammensætning af par blev udvalgt af skoleinspektørerne selv. Testpersonerne meldte sig frivilligt til testen. De fik alle hver sin mail, med instruktioner om hvordan de skulle gennemføre testen (*bilag 2*). Efter at testen var afviklet, blev det arrangeret at de testdeltagere, som havde medvirket i testen, fik tilsendt en godtgørelse i form af en bog fra skolebogforlaget, som også er en del af uddannelses centeret, som har ansvaret for FirstClass kommunikationssystemet.

Et website blev programmeret i php og MySql med instruktion til gennemførelse af testen, samt opgaver og testspørgsmål. Der blev programmeret to gange tre forskellige indgange efter de tre instruktionstyper. Sitet findes på [www.faroecraft.net/index13.php](http://www.faroecraft.net/index13.php).

Dette er en af seks indgange. I tabellen nedenfor likes der til alle indgangene. Sitet vil kun være tilgængeligt til udgangen af 2014. Jeg har oprettet en testbruger på FirstClass systemet, hvor man kan gå ind og se brugerfladen. Logonoplysningerne er:

**BrugerID: gmroynd – Adgangskode: dnug**

FirstClass sitet fremkommer som iFrame inde i nedenstående links. (*direkte link til sitet: <https://fcws.skulin.fo>*)

|                                 | Single - user                                                                      | Pairs                                                                              |
|---------------------------------|------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| Deduktiv instruktion            | <a href="http://www.faroecraft.net/index11.php">www.faroecraft.net/index11.php</a> | <a href="http://www.faroecraft.net/index21.php">www.faroecraft.net/index21.php</a> |
| Induktiv instruktion            | <a href="http://www.faroecraft.net/index12.php">www.faroecraft.net/index12.php</a> | <a href="http://www.faroecraft.net/index22.php">www.faroecraft.net/index22.php</a> |
| Deduktiv + induktiv instruktion | <a href="http://www.faroecraft.net/index13.php">www.faroecraft.net/index13.php</a> | <a href="http://www.faroecraft.net/index23.php">www.faroecraft.net/index23.php</a> |

Enkeltpersonerne og parrene blev jævnt fordelt ud på de tre typer instruktion. På tavlen nedenfor ses fordelingen af testpersoner og instruktionstype.

Tabel 3

|                                 | Single         | Pair       | Pair       | Pair       |
|---------------------------------|----------------|------------|------------|------------|
| Deduktiv instruktion            | 2 x M<br>4 x K | 2 x KK     | 2 x MK     | 2 x MM     |
| Induktiv instruktion            | 2 x M<br>4 x K | 2 x KK     | 2 x MK     | 2 x MM     |
| Deduktiv + induktiv instruktion | 2 x M<br>4 x K | 2 x KK     | 2 x MK     | 2 x MM     |
|                                 | 18 pers        | 6 par (18) | 6 par (18) | 6 par (18) |

Som man kan læse i tabel 3, ud for kolonnen med single users, er der et overtal af kvinder. Årsagen hertil er, at der tilmeldte sig ikke flere mænd, og derfor blev fordelingen på single users således.

Testsitet blev konstrueret således at selve websitet, som skulle testes, blev indlejret med iframe i en side med opgaver og spørgsmål. Dette blev lavet for at undgå at testpersonerne skulle skifte mellem flere vinduer på skærmen. Man kan diskutere hvor

hensigtsmæssigt det var, fordi brugerne i den valgte tekniske løsning, mistede friheden at fylde hele skærmen med sitet som skulle testes.

Det viste sig også at nogle problemer med browser kompatibilitet opstod, som man måtte forklare for testdeltagerne i en ekstra mail. De fik dog lavet testen tilfredsstillende, men man skal nøje vurdere til næste gang, om man skal benytte samme tekniske opsætning. Der kom flere meldinger om at man ikke kunne logge ind på FirstClass gennem testsiden, specielt i Safari browsere. Men da testen allerede var igang, blev det besluttet, at man ikke ændrede noget i programmeringen på testsitet, men i stedet lave en vejledning til deltagerne, som styrede dem udenom de tekniske problemer.

Men man skal tage dette i betragtning, da man behandler testresultaterne, fordi selve testmiljøet har skabt unødigt frustration og delvis flyttet fokus fra den oprindelige test.

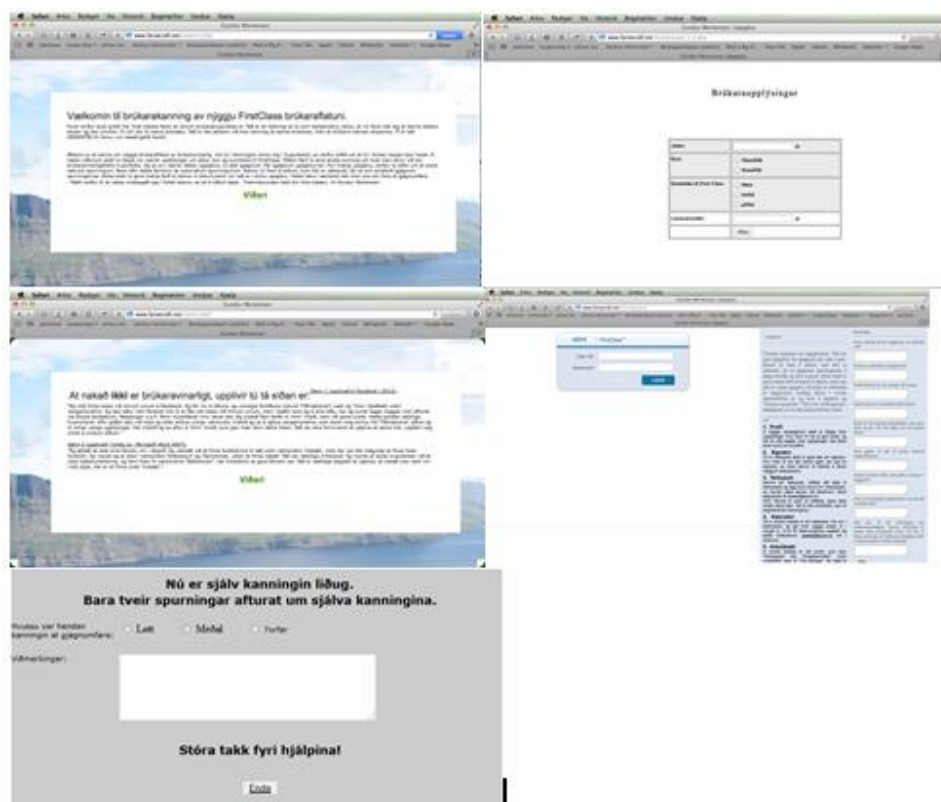


Screenshot af testsitet med FirstClass indlejret som iFrame.

Testen blev udført som en opgavebaseret remote asynkron test i henhold til Bruun og Stage (2012, Bruun, A; Stage, J ). Deres resultater viser, at med predefinerede opgaver bliver signifikant flere brugbarhedsproblemer identificeret, end hvis testbrugerne arbejder med egne autentiske opgaver. Instruktionsstypen er også af betydning, da den induktive

instruktionstype identificerer signifikant flere brugbarhedsproblemer end ved deduktiv instruktion. Mit eksperiment benytter sig af de samme tre typer opgavebaseret instruktion, men testbrugerne bliver sammenlignet enkeltvis/parvis, og parrene bliver sammenlignet efter sammensætning af køn og instruktionstype.

Hver enkelt testbruger fik en email tilsendt med detaljerede instruktioner om testen og en grundig forklaring på, at de ikke skulle føle at vi testede dem, men at det var de, som testede vores system med henblik på at vi kunne lave nogle videotutorials til brugerne. Ved at informere testdeltagerne så godt som muligt om hensigten med testen, formindsker man den kunstige situation, som altid vil være der i en testsituation. Men der vil altid være en vis grad af usikkerhed omkring det.



Screendumps fra websitet.

Da alt materiale, som testdeltagerne skulle få tilsendt og websitet med opgaver og spørgsmål, var færdigt programmeret, blev tre personer på min arbejdsplads bedt om at

gennemgå testen med kritiske øjne og kigge efter, om opgaver og spørgsmål var til at forstå og se efter om det teknisk kunne godkendes. Efter at de tre havde været igennem testen, blev alle deres notater diskuteret igennem og justeringer lavet på testsitet, både af indholdsmæssige tekstformuleringer og tekniske justeringer.

Derefter blev der arrangeret en pilottest, hvor fem andre arbejdskollegaer frivilligt påtog sig at være testpersoner. Tre personer gennemførte testen som enkeltpersoner, hvor de fik hver sin instruktionstype. To personer gennemførte testen som par med kombineret instruktionstype (induktiv+deduktiv). Efter at alle havde gennemført testen, samlede jeg dem, for at høre deres oplevelse af testen. Vi kiggede i fællesskab igennem deres testsvar, som blev hentet ud af databasen. Ud fra denne samtale med mine pilottestpersoner, blev der finjusteret noget på spørgsmålene og efter det blev testen godkendt og klar til at blive sat i gang.

Alle testdeltagerne fik tilsendt en personlig mail med velkomsthilsen, instruktioner og link til selve testen. I mailen til parrene var der instrukser om at de selv skulle aftale indbyrdes, hvornår de ville afvikle deres session, samt at de skulle skiftes til at sidde ved computeren og lave opgaver. Enkeltbrugerne fik instrukser på at de skulle være opmærksomme på deres handlinger under testen, for bedre at kunne formulere sig skriftligt, når de stødte på problemer, dvs. de blev bedt om at bruge tænke højt-metoden i en varieret form(2008, Rubin). Parrene fik ligeledes instrukser om at konferere med hinanden under testen i henhold til Constuctive Interaction (1986, Miyake, N.). Så de også kunne få en indbyrdes forståelse for, hvad der sker under testen og dermed bedre kunne formulere sine svar, når de skulle skrive om deres eventuelle brugerproblemer. Disse metode detaljer blev dog ikke præsenteret for testpersonerne.

Alle testpersonerne fik en tidsramme på en uge at afvikle testen i, og blev bedt om at afsætte en cirka time på det. Efter fire dage blev der sendt en påmindelse ud til deltagerne. I samme mail blev der informeret og vejledt om nogle tekniske problemer, som var opstået efter at testen var sat i gang. Dette var problemer, som ikke var blevet registrerede under pilottesten. Problemerne var alvorlige nok, men da var det for sent at lave ændringer i programmeringen. Brugere blev informerede om problemerne og om hvordan man kunne styre uden om dem, for at udføre testen. Dette bliver taget med i

behandlingen af indsamlede data senere i opgaven. Efter en uge, fik testdeltagerne tilsendt endnu en påmindelse om at gennemføre testen. Alle testdeltagerne fik samme påmindelse, fordi testresultaterne er anonyme og dermed ikke kan sammenkædes med bestemte personer. Påmindelsen var forfattet sådan, at de der allerede havde sendt ind deres svar, var undskyldt. I dagene efter fik jeg ind nogle ekstra datasæt.

## Data analyse:

Dataanalysen blev udført af mig selv, efter at tidsfristen med ekstra tid var gået. Alle dataene blev eksporteret fra MySQL databasen ud til Excel ark og derefter systematiseret og delt op efter brugersessionerne (bilag 3).

| s2  |                                                                                                                                           |                                       |                                                   |                                                                                    |                                                    |                     |                    |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------|---------------------------------------------------|------------------------------------------------------------------------------------|----------------------------------------------------|---------------------|--------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1   | 335                                                                                                                                       | 1                                     | 335                                               | 1                                                                                  | 335                                                | 1                   | 335                | 1        | 335      | 1        | 335      | 1        | 335      | 1        | 335      | 1        | 335      | 1        | 335      | 1        | 335      |
| 335 | 40                                                                                                                                        | 40                                    | 40                                                | 40                                                                                 | 40                                                 | 40                  | 40                 | 40       | 40       | 40       | 40       | 40       | 40       | 40       | 40       | 40       | 40       | 40       | 40       | 40       | 40       |
| L   | L                                                                                                                                         | L                                     | L                                                 | L                                                                                  | L                                                  | L                   | L                  | L        | L        | L        | L        | L        | L        | L        | L        | L        | L        | L        | L        | L        | L        |
| 0   | 0                                                                                                                                         | 0                                     | 0                                                 | 0                                                                                  | 0                                                  | 0                   | 0                  | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 0        | 0        |
| K   | K                                                                                                                                         | K                                     | K                                                 | K                                                                                  | K                                                  | K                   | K                  | K        | K        | K        | K        | K        | K        | K        | K        | K        | K        | K        | K        | K        | K        |
|     |                                                                                                                                           | 5                                     | 4                                                 | 4                                                                                  | 4                                                  | 4                   | 4                  | 4        | 4        | 4        | 4        | 4        | 4        | 4        | 4        | 4        | 4        | 4        | 4        | 4        | 4        |
|     |                                                                                                                                           | Arbeidsþéttur                         | Kalendari                                         | Teldupost                                                                          | Signatur                                           | Profil              | Navn.              |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
|     |                                                                                                                                           | Leggja skjal av teldunni út á mappuna | Stovna avtalu                                     | Sendu post og viðheftan til og mynd                                                | Seta signatur á allar mailar                       | Dagfært upplýsingar | Intention          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
|     |                                                                                                                                           | At skjali bleiv skrásett              | At atvatan bleiv skrásett og móttakarinn fékk boð | Tað riggaði allt ok at siggja til, men dugi ekki at siggja at posturinn er sendur. | Kom víðari beinanvegin                             | Eingin trupulleiki  | Problembeskrivelse |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
|     |                                                                                                                                           | Hendi einki                           | Royndi fleiri ferðir                              | Havi roynt fleiri ferðir, tó sama úrsilt.                                          | Um mailurinn ekki er móttikin, kom eg ekki víðari. |                     |                    |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
|     |                                                                                                                                           |                                       | Nei                                               |                                                                                    |                                                    |                     |                    |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
|     |                                                                                                                                           |                                       | Ónýtilig                                          | Ónýtilig                                                                           | Ónýtilig                                           | Ónýtilig            | Ónýtilig           | Ónýtilig | Ónýtilig | Ónýtilig | Ónýtilig | Ónýtilig | Ónýtilig | Ónýtilig | Ónýtilig | Ónýtilig | Ónýtilig | Ónýtilig | Ónýtilig | Ónýtilig | Ónýtilig |
| 2   | Hey Gundur, eg mangli at svara 2 spurningum nr 6 og 7. Eg dugi ekki at finna teir. Eri nýggi í starvinum og kenni ekki so gott skipanina. | 16.54.24                              | 10/05/14 16.58                                    | 1.15.47                                                                            |                                                    |                     |                    |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
|     |                                                                                                                                           | 16.37.35                              | 10/05/14 16.37                                    | 0.16.49                                                                            | 0.10.23                                            | 0.10.55             | 0.26.09            | 0.11.31  | 0.11.31  | 0.11.31  | 0.11.31  | 0.11.31  | 0.11.31  | 0.11.31  | 0.11.31  | 0.11.31  | 0.11.31  | 0.11.31  | 0.11.31  | 0.11.31  | 0.11.31  |

Det store regneark, med alle dataene, blev opdelt i flere ark, et ark for hver testbruger. (Bilag 3). Arkene blev igen delt op efter de instruktionstyper, sum testbrugerne var blevet instrueret i. Der var 6 enkelt brugere og 6 par til hver instruktionstype, induktiv, deduktiv og ind+ded instruktion. I alt 18 enkeltbrugere og 18 par (36 pers). Indenfor hver instruktionsgruppe af enkeltbrugere, var der jævn kønsfordelingen mellem kvinder og mænd. Parrene blev fordelt på den måde, at indenfor hver gruppe var der repræsenteret to par mænd, to par kvinder og to par mand og kvinde.

I testen blev testdeltagerne bedt om at oplyse: (*Bilag 8*)

- Alder
- Køn
- Undervisnings erfaring (år)
- Kendskab til systemet (år)

Da testbrugerne stødte på et brugervenlighedsproblem, skulle de svare på nedenstående spørgsmål, og sende dem til databasen ved at trykke på submit på testsitet.

- Hvilket nummer havde opgaven, som du var i gang med?
- Hvad var titlen på opgaven?
- Forklar hvad du skulle gøre.
- Forklar hvad du forventede skulle ske.
- Prøv at belyse selve problemet, som opstod og giv et bud på, hvorfor det mon skete
- Hvad gjorde du for at komme forbi problemet?
- Kunne du komme videre, eller måtte du begynde på nyt?
- Kan du genkabe problemet og få det til at ske igen?
- Bedøm, ud fra instruktionen om brugervenlighed, hvor alvorligt du synes problemet er. Du kan kigge i den tilsendte mail, hvis du vil læse instruktionen igen.

Efter sidste indberetning, blev testdeltagerne bedt om at bedømme selve testen ud fra en skala fra 1-5, og med mulighed at skrive yderligere bemærkninger. Systemet var programmeret sådan, at det fik logget tiden, som testdeltagerne brugte på hver enkelt opgave, og også det samlede tidsforbrug på selve testen.

Da alle datasessionerne var blevet isoleret i egne ark, satte jeg mig at læse og analysere hvert enkelt ark. En problemliste blev udarbejdet ud fra alle indsamlede kommentarer fra testbrugerne (*bilag 4*). Alle indrapporterede brugerproblemer blev dobbeltchecket og forsøgt genskabt af mig selv. I de tilfælde, hvor jeg var i tvivl om hvordan enkelte formuleringer skulle forstås, blev der på sitet undersøgt en ekstra gang om man kunne genskabe problemet ud fra de data som forelå. Hvis det kunne genskabes ud fra givne



data, blev det taget med. Da jeg læste indberetningerne igennem, blev først alle indberetningerne sorteret efter overskrift på opgave. Dvs. alle indberetningerne, som kom under opgaven 'Profil', kom i en pulje osv. Efter denne grovsortering af indberetninger, satte jeg mig at finsortere og samle indberetningerne i små puljer indenfor hver overskrift, sådan at de der indberetninger, der lignede hinanden, kom i en pulje og hver gang en indberetning ikke kunne lægges i en pulje, blev der oprettet en ny pulje. På den måde fik jeg lavet nogle puljer af indberetninger, som jeg sidenhen analyserede og konverterede til usabilityproblem. Derefter blev usabilityproblemerne igen opdelt efter alvorlighedsgraderne 'Kritisk', 'Alvorligt' og 'Kosmetisk'. I testafviklingen fik alle deltagerne et unikt ID. Deltagerne blev identificeret med S for single user og P for par, og så et tal. ID numrene S1-S6 og P1- P6 har fået deduktiv instruktion, ID S7 - S12 og P7-P12 har fået induktiv instruktion og ID S13 –S18 og P13-P18 har fået kombineret instruktion. Disse ID numre blev registrerede sammen med brugbarhedsindberetningerne.

Efter konverteringen af brugbarhedsindberetningerne havde jeg så en liste med brugbarhedsproblemer, delt op efter alvorlighed og en registrering på hvilke testdeltagere havde oplevet de forskellige problemer (*bilag 4*).

## Resultat:

### Testdeltagelse

Det viste sig hurtigt, at testdeltagelsen ikke ville komme op på 100%, skønt hver enkelt testdeltager med tro og ære havde lovet at gennemføre testen indenfor givne tidsramme. Det giver grundlag for diskussion om hvor mange testdeltagere man skal medregne, ud over det ønskede antal, for at testen skal være troværdig. Man kan også spørge, om det er demografisk/kulturelt betinget, at testpersoner indvilliger at være med i en test, og alligevel ikke holder ord. Mere om dette under punktet diskussion.

## Testdeltagelse enkeltbruger

Tabellerne nedenfor viser testdeltagelsen, fordelt på hvilke instruktioner de havde fået. Tallene viser, at 13 ud af 18 (72 %) af enkeltbrugerne gennemførte testen, mens 5 ud af 18 (28 %) af brugerne ikke gennemførte eller undlod at svare. Kigger man ind til type af instruktion og indleverede svar, har 5 ud af 6 (83 %) enkeltbruger med deduktiv instruktion indleveret svar. For enkeltbrugerne med induktiv instruktion, gennemfører alle testbrugerne, 6 ud af 6 (100%). Der er kun 2 ud af 6 (33 %) enkeltbrugere med instruktionstypen ind+ded, der har indleveret svar.

Tabel 4

|                 | Indleveret svar | Procent |  | Ikke indleveret svar | %    |
|-----------------|-----------------|---------|--|----------------------|------|
| Single ded:     | 5 (6)           | 83 %    |  | 1 (6)                | 17 % |
| Single ind:     | 6 (6)           | 100 %   |  | 0 (6)                | 0 %  |
| Single ded+ind: | 2 (6)           | 33 %    |  | 4 (6)                | 67 % |
|                 | 13 (18)         | 72 %    |  | 5 (18)               | 28 % |

I gennemsnit gennemfører 13 ud af 18 (72 %) af enkeltbrugerne testen, mens 5 ud af 18 (28 %) ikke gennemførte eller undlod at svare.

## Testdeltagelse par

Vi ser det samme mønster ses hos parrene, hvor der er 4 ud af 6 ( 67 %) par med deduktiv instruktion, og 4 ud af 6 ( 67 %) par med induktiv instruktion, der har indleveret svar. Kun halvdelen af parrene, 3 ud af 6 ( 50%) med kombineret instruktionstype ind+ded, har indleveret svar. Her skal bemærkes, at da deadline for dataindsamlingen var gået, havde kun et par med kombineret instruktionstype indleveret svar. For at få bedre

sammenlignsgrundlag, prøvede jeg endnu engang at kontakte disse testpersoner om at gennemføre testen. Da kom der to svarsæt mere. Da jeg vidste hvilke seks par havde fået den kombinerede instruktionstype, kontaktede jeg dem alle via telefon, og fik dem at lave endnu et forsøg at gennemføre testen. Det ene par havde allerede svaret, to andre ville prøve igen, mens tre par ville ikke gennemføre testen med den begrundelse at testen tog for lang tid og at det var svært for dem at finde noget fælles tidspunkt at gennemføre testen. Det måtte jeg så bare acceptere.

Tabel 5

|               | Indleveret svar |      | Ikke indleveret svar |      |
|---------------|-----------------|------|----------------------|------|
| Pair Ded:     | 4               | 67 % | 2                    | 33 % |
| Pair Ind:     | 4               | 67 % | 2                    | 33 % |
| Pair ded+ind: | 3               | 50%  | 3                    | 50 % |
|               | 11              | 61 % | 7                    | 39 % |

### **Indleveret svar, kønsfordeling.**

I kønsfordelingen af enkeltbrugere, der indleverede svar, ser man straks, at kvinderne er dygtigere til at gennemføre testen. Af alle single testbrugerne, som indvilligede i at være med i testen, var der 7 mænd og 11 kvinder. Oplægget var at bruge 3 personer af hvert køn til hver enkelt instruktionstype, i alt 9 mænd og 9 kvinder, men det viste sig ikke at være muligt at få fat i alle mændene. Da der var flere kvinder, som tilmeldte sig, blev fordelingen derfor 7 mænd og 11 kvinder. I tabellen ser man også, at selv om mændene har indvilliget i at gennemføre testen, så gør de det alligevel ikke. 3 ud af 7 mænd ( 43 %) gennemfører, men 10 ud af 11 ( 91 %) kvinder gennemfører.

Tabel 6: Kønsfordeling af enkeltbrugere, som har svaret

|   | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | S11 | S12 | S13 | S14 | S15 | S16 | S17 | S18 |        |      |
|---|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------|------|
| ♂ | ✓  |    |    |    |    | ✗  |    |    |    | ✓   | ✓   |     |     |     |     | ✗   | ✗   | ✗   | 3(7)   | 43 % |
| ♀ |    | ✓  | ✓  | ✓  | ✓  |    | ✓  | ✓  | ✓  |     |     | ✓   | ✓   | ✓   | ✗   |     |     |     | 10(11) | 91 % |
|   |    |    |    |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     | 13(18) | 73 % |

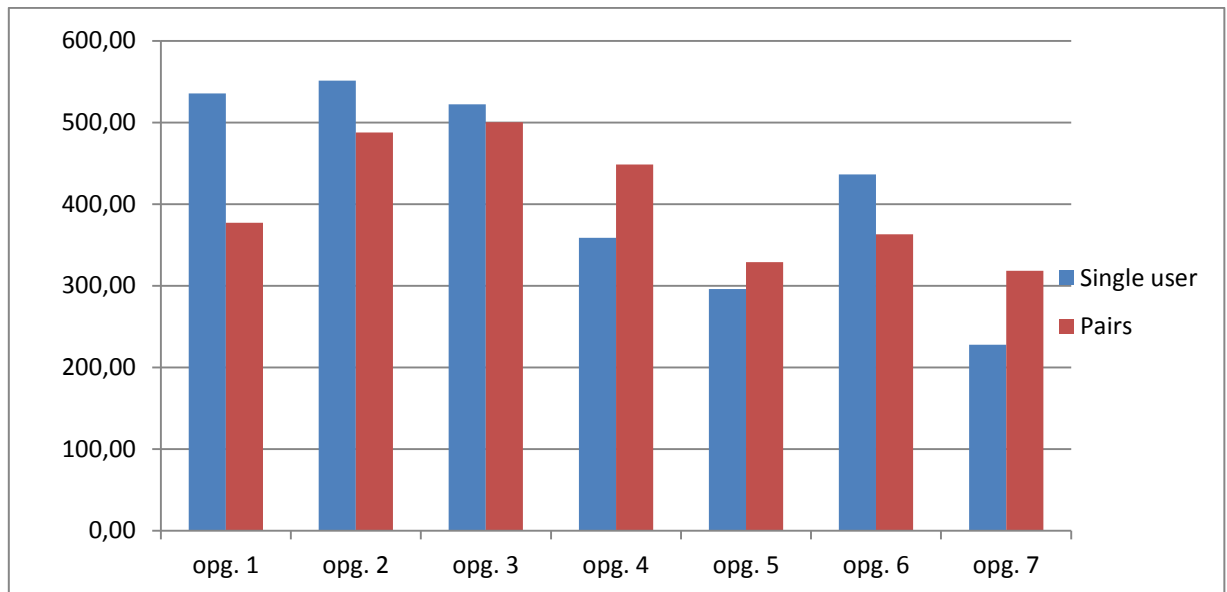
I kønsfordelingen af par, der indleverede svar, ser man at instruktionsgrupperne er repræsenteret i alle tre par sammensætninger (mand-mand ♂♂, mand-kvinde ♂♀, kvinde-kvinde ♀♀). I parsammensætningen, hvor begge personer er mænd, (♂♂), er der kun halvdelen af testparrene, som har svaret. 3 ud af 6 par (50%). Parsammensætningen mand-kvinde (♂♀) svarer 5 ud af 6 par (83 %) og parsammensætningen af kvinde-kvinde (♀♀) svarer også kun 3 ud af 6 par (50%). Ud fra dette kan man se at parsammensætningen mand-kvinde (♂♀) er den der giver højest testdeltagelse.

Tabel 7: Kønsfordeling af par, som har svaret

|    | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 | P11 | P12 | P13 | P14 | P15 | P16 | P17 | P18 |        |      |
|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------|------|
| ♂♂ |    |    |    |    | ✗  | ✓  |    |    |    |     | ✓   | ✗   |     |     |     |     | ✓   | ✗   | 3 (6)  | 50 % |
| ♂♀ |    |    | ✓  | ✓  |    |    |    |    | ✓  | ✓   |     |     |     |     | ✓   | ✗   |     |     | 5 (6)  | 83 % |
| ♀♀ | ✓  | ✗  |    |    |    |    | ✗  | ✓  |    |     |     |     | ✓   | ✗   |     |     |     |     | 3 (6)  | 50 % |
|    |    |    |    |    |    |    |    |    |    |     |     |     |     |     |     |     |     |     | 11(18) | 61 % |

Grafen (fig 1) på næste side, viser gennemsnitstiden single users og pairs løste de 7 opgaver. Y-aksen viser tiden i sekunder. Resultatet af en uparret t-test, viser ingen signifikant forskel mellem single users og pair, når det kommer til gennemsnitstiden at løse opgaverne. (t=-0,3039, df=5, p<0,773406).

(Opsamling af data og div. udregninger er ligger på bilag 5 og 6)



*Fig 1. Gennemsnits tid single users og pairs.*

## Problemliste

I min evalueringsprocess kiggede jeg på alle indberetningerne fra single users og fra pairs. Alle indberetningerne blev analyseret og oversat til usabilityproblem. For hver indberetning jeg analyserede, gik jeg igennem sitet og prøvede at genskabe problemet ud fra beskrivelsen, som testdeltagerne havde fundet. I de tilfælde, hvor jeg ikke kunne forstå indberetningen, eller hvor jeg ikke kunne genskabe problemet, blev disse indberetninger frasortet. Det skete dog kun i et tilfælde, og vælger jeg derfor ikke at tage det med i evalueringen.

Indberetningerne blev oversat til usabilityproblem efter kategorierne Kritisk, Alvorligt og kosmetisk problem. Til dette arbejde anvendte jeg et værktøj, som blev introduceret på fagpakken, Brugbarhed, på Aalborg Universitet forår 2011 hos professor Jan Stage.

### Værktøj til problemidentifikation (Jan Stage)

|           | <b>Forsinkes i at løse opgaven</b><br><i>(relativt til testpersonens normale arbejdstempo)</i> | <b>Forståelse</b>                                                                                                                                | <b>Frustration</b>                                                                                                                         | <b>Testleder</b>                                                                       |
|-----------|------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| Kritisk   | Forhindres i at løse opgaven                                                                   | Forstår ikke hvordan informationer i systemet kan bruges i løsningen af en opgave. Gentager de samme informationer forskellige steder i systemet | Bliver meget frustreret over, at hun ikke kan finde ud af det, og opgiver helt at løse opgaven.                                            | Får betydelig hjælp fra testleder (kunne ikke have løst opgaven uden)                  |
| Alvorligt | Forsinkes i adskillige sekunder                                                                | Forstår ikke hvordan en bestemt funktionalitet fungerer eller aktiveres. Kan ikke forklare systemets virkemåde.                                  | Bliver tydeligt irriteret over noget man ikke kan gøre eller huske noget eller noget ulogisk man skal gøre. Tror at hun har ødelagt noget. | Bliver ledt på vej af testlederen.                                                     |
| Kosmetisk | Forsinkes i nogle få sekunder                                                                  | Gør ting, som ikke kan forklares (det skal man bare)                                                                                             | Bliver lidt irriteret over måden noget skal gøres på.                                                                                      | Testlederen stiller et spørgsmål, som får testpersonen til selv at komme på løsningen. |

Efter at have analyseret alt materialet fra 13 single users og 11 par, fik jeg lavet en liste med usability problemer (*Bilag 4*). I denne liste, kunne jeg identificere et samlet antal 26 usability problemer, hvoraf ni problemer var i kategorien ”Kritisk”, tretten i kategorien ”Alvorligt” og fire problemer var i kategorien ”Kosmetisk”.

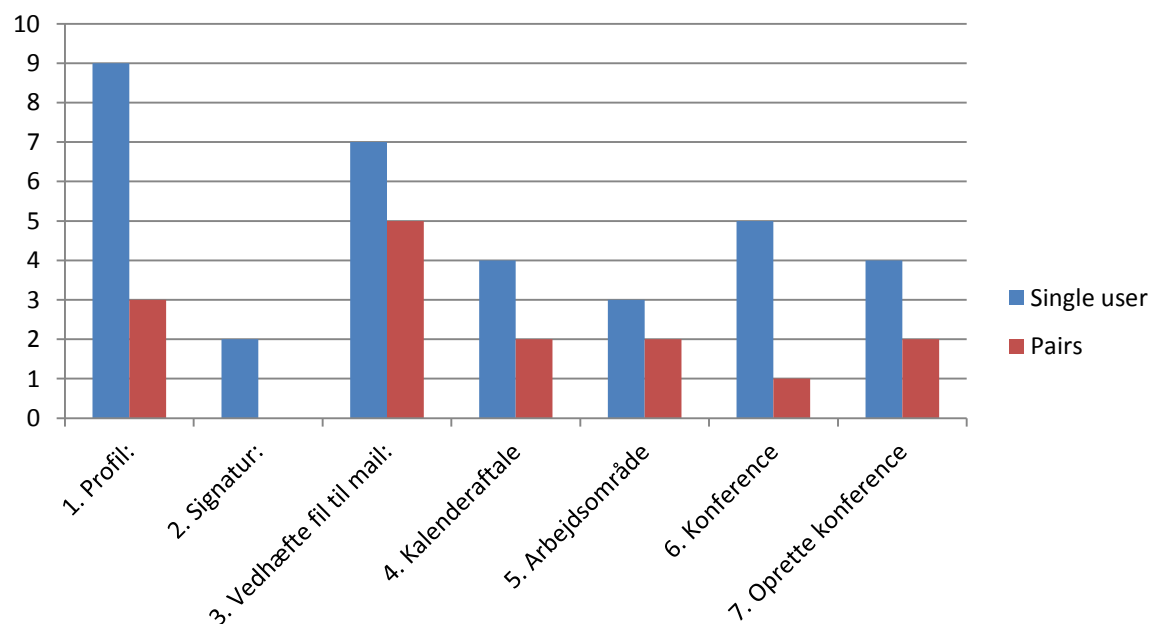
|           | Kritisk | Alvorligt | Kosmetisk |
|-----------|---------|-----------|-----------|
| Problemer | 9       | 11        | 4         |

I tabellen nedenfor (tabel 8) vises det samlede antal usability indberetninger, delt ud på hvem der har indberettet dem.

## Indberettede usabilityproblem, single vs pair

Tabel 8

| Afdækkede problemer       | single<br>user | pair | Singleuser<br>+ pairs |
|---------------------------|----------------|------|-----------------------|
| 1. Profil:                | 9              | 3    | 12                    |
| 2. Signatur:              | 2              | 0    | 2                     |
| 3. Vedhæfte fil til mail: | 7              | 5    | 12                    |
| 4. Kalenderaftale         | 4              | 2    | 6                     |
| 5. Arbejdsområde          | 3              | 2    | 5                     |
| 6. Konference             | 5              | 1    | 6                     |
| 7. Oprette konference     | 4              | 2    | 6                     |
|                           | 34             | 15   | 49                    |



Efter en uparret t-test, viser tallene i tabel 8, at der kommer signifikant flere indberetninger fra single users end fra pairs ( $t=2,49, df=12, p<0,028182$ ). Singleusers indberetter 34 problemer i forhold til 15 indberetninger fra pairs. Det vil sige, at single users indberetter 69 % af alle registrerede indberetninger, mens pairs indberetter 31 %. Fordelt over opgaverne, kommer single users med flere indberetninger i alle opgaver, og endda i fem ud af syv opgaver, er der 50% eller flere indberetninger i forhold til pairs. I to af opgaverne indberetter de næsten samme antal.

Det er interessant at notere sig, at det spørgsmål, som dukker op i forbindelse med fordelingen af indberetninger mellem single users og pairs, er om parrene er bedre til at snakke sig igennem problemerne og dermed løse dem, mens de går igennem opgaverne? Jeg vil komme nærmere ind på dette under punktet diskussion.

Tabellen nedenfor (tabel 9) viser det samlede antal af identificerede problemer delt op efter instruktionstype hos begge grupper testdeltagere. Tallene viser, at den deduktive instruktionstype afslører langt de fleste problemer. Ud af 26 identificerede usability problemer, finder testbrugerne med deduktiv instruktion 26 (100 %) ud af 26 problemer, mens brugerne med induktiv instruktion finder 16 (62 %) ud af 26 problemer, og testbrugere med kombineret instruktion ligger på 11 (69 %) ud af 26 fundne problemer.

| Singleusers + Pairs                                                                                                                     | Kritiske | Alvorlige | Kosmetiske | Total    |
|-----------------------------------------------------------------------------------------------------------------------------------------|----------|-----------|------------|----------|
| deduktiv instruktion<br>n= 9                                                                                                            | 9        | 13        | 4          | 26 (100) |
| Induktiv instruktion<br>n=10                                                                                                            | 1        | 11        | 4          | 16 ( 62) |
| Ded + ind instruktion<br>n= 5                                                                                                           | 1        | 7         | 3          | 11 ( 42) |
|                                                                                                                                         | 9        | 13        | 4          | 26 (100) |
| Tabel 9: Antal funde problemer delt op efter instruktionstype. (N= tal av singleusers + pairs). Tallene i parentes er procent af total. |          |           |            |          |



I næste tabel (Tabel 10) er single users isolerede. Tabellen viser fordelingen af fundne problemer fordelt ud på instruktionstype. Her ser man at deduktiv instruktion afslører 25 (96%) ud af 26 fundne problemer, mens induktiv og kombineret instruktion afslører henholdsvis 20 (72%) og 2 (8%) usabilityproblemer. Der skal tages hensyn til testresultatet for kombineret instruktion, da kun to ud af seks testdeltagere indberettede testsvar.

| Single user                                                                                                                      | Kritiske | Alvorlige | Kosmetiske | Total    |
|----------------------------------------------------------------------------------------------------------------------------------|----------|-----------|------------|----------|
| deduktiv instruktion<br>n=5                                                                                                      | 9        | 12        | 4          | 25 (96)  |
| Induktiv instruktion<br>n=6                                                                                                      | 8        | 8         | 4          | 20 (72)  |
| Ded + ind instruktion<br>n=2                                                                                                     | 1        | 1         | 0          | 2 ( 8)   |
|                                                                                                                                  | 9        | 13        | 4          | 26 (100) |
| Tabel 10: Antal funde problemer delt op efter instruktionstype. (N= tal av singleusers). Tallene i parentes er procent af total. |          |           |            |          |

I tredje tabel (Tabel 11) er pairs isolerede. Ligesom i tabellen ovenfor, viser den fordelingen af fundne problemer fordelt ud på instruktionstype. Man ser igen samme tendens som før, at deduktiv instruktion afslører 12 problemer (46 %) ud af 26 fundne problemer, mens induktiv og kombineret instruktion afslører henholdsvis 6 (23 %) og 16 (62 %) usabilityproblemer. I modsætning til singleusers, hvor den deduktive instruktion gav det højeste resultat fundne problemer, er det hos pairs gældende at det er den kombinerede instruktionstype, der giver flest funde problemer. Men stadigvæk giver den induktive instruktionstype ikke nær så mange fundne problemer som de andre.

| Pairs                                                                                                                      | Kritiske | Alvorlige | Kosmetiske | Total    |
|----------------------------------------------------------------------------------------------------------------------------|----------|-----------|------------|----------|
| deduktiv instruktion<br>n=4                                                                                                | 4        | 7         | 1          | 12 (46)  |
| Induktiv instruktion<br>n=4                                                                                                | 1        | 5         | 0          | 6 (23)   |
| Ded + ind instruktion<br>n=3                                                                                               | 6        | 8         | 2          | 16 (62)  |
|                                                                                                                            | 9        | 13        | 4          | 26 (100) |
| Tabel 11: Antal funde problemer delt op efter instruktionstype. (N= tal av pairs). Tallene i parentes er procent af total. |          |           |            |          |

### Fordeling af indberetninger i forhold til parsammensætning:

I tabel 12 vises der fordelingen af indberetninger af usabilityproblemer fra de forskellige par, opdelt efter parsammensætning. Her ser vi at kombinationen kvinde+kvinde (KK) indberetter 8 kritiske problemer, 16 alvorlige problemer og 3 kosmetiske problemer.

| Pairs                                                                                | KK n=3 | MM n=3 | MK n=5 |
|--------------------------------------------------------------------------------------|--------|--------|--------|
| Kritiske                                                                             | 8      | 4      | 9      |
| Alvorlige                                                                            | 16     | 2      | 11     |
| Kosmetiske                                                                           | 3      | 1      | 2      |
|                                                                                      | 27     | 7      | 22     |
| Tabel 12: Antal funde indberetninger i forhold til parsammensætning. n= antal af par |        |        |        |

Kombinationen mand+kvind (MK) indberetter 9 kritiske problemer, 11 alvorlige problemer og 2 kosmetiske problemer, og kombinationen mand+mand (MM) indberetter 4 kritiske problemer, 2 alvorlige problemer og 1 kosmetiske problemer. Tallene i tabellen skal læses som antal indberetninger og ikke som usabilityproblem. Vi ser at par-

kombinationen KK og MK indberetter nogenlunde samme antal problemer, men MM er viser et langt dårligere resultat.

## Diskussion:

Igennem denne opgave har jeg kigget på effekten af opgavebaseret asynkron remote usabilitytest, hvor jeg har koncentreret mig om tre forskellige instruktionstyper. Første instruktionstype, deduktiv instruktion, er baseret på at testbrugerne bliver oplært i at forstå hvad et usabilityproblem er, gennem præsentation af usability definitionerne: usefulness, efficiency, effectivity, satisfaction og accessibility (2008, Rubin). For at gøre det mere overskueligt for testdeltagerne, er usability definitionerne vendt om, således at når brugeren føler at noget føles ubrugeligt, ueffektivt osv., så skal de opfatte det som et usabilityproblem.

I den næste instruktionstype, Induktiv instruktion, får testbrugerne præsenteret et par eksempler på hvad et usabilityproblem er, og ud fra denne præsentation, skal de selv kunne gennemskue et usabilityproblem, når de støder på noget i testen.

Den tredje instruktionstype, deduktiv + induktiv instruktion, er en kombination af begge de første instruktionstyper. (2012, Bruun og Stage)

Mine testbrugere er blevet delt op i single users og pairs, for at undersøge om der er nogen forskel på testresultaterne mellem de to grupper. Yderligere er parrene delt op efter køn, så hver instruktionstype gruppe er blevet sammensat med to par kvinder, to par mænd og to par mand og kvinde.

I testsituationen blev single users instrueret i at "Tænke højt" under testen, og parrene blev instrueret i at snakke konstruktivt sammen "Constructive Interaction" under testen.

I dette afsnit vil jeg kigge nærmere på hvilken effekt de forskellige instruktionstyper har haft på testresultaterne for single users og for pairs.

## Effekt af instruktionstype

Den asynkrone remote usabilitytest afslørede et samlet antal usabilityproblemer på 26 problemer, hvor af ni af problemerne var kritiske, tretten alvorlige og fire kosmetiske. Fordelt ud på instruktionstyper, viser resultaterne, at hos single users afslører den deduktive instruktion 96% af alle problemerne, mens henholdsvis den induktive og den kombinerede instruktionstype afslører 72% og 8% af de fundne problemer.

Hos parrene viser resultaterne at den kombinerede instruktionstype afslører højest antal fundne problemer, 61,5%. Derefter kommer den deduktive instruktionstype med 46% og til sidst den induktive type med 23%.

Sammenlignet med resultaterne hos Bruun og Stage (2012, Bruun og Stage), viser det sig at den deduktive instruktionstype er mere effektiv i denne test, hvorimod Bruun og Stage kan stadfæste at den Induktive instruktionstype giver højest antal fundne problemer i deres test. De kommer frem til at den induktive instruktionstype afslører mest usabilityproblemer i forhold til de to andre instruktionstyper. Den deduktive afslører 17 problemer, mens den induktive afslører 29 problemer og den kombinerede afslører 25 problemer. I mine resultater ser man samme mønster i fordelingen af problemer. Deduktiv instruktion afslører 26 problemer, Induktiv og kombineret instruktion afslører henholdsvis 16 og 11 problemer. Men det er altså den deduktive instruktionstype, som er mest effektiv i mine testresultater.

Mine testbrugere er udvalgt fra en bestemt faggruppe, som består af folkeskolelærere i aktiv tjeneste. Det er muligt, at specielt denne faggruppe bedst trives med at få en deduktiv instruktion, når de skal til at modtage og forstå information, som de derefter skal arbejde videre med. Det ville være interessant at kigge nærmere på, og det har nogen indvirkning på resultaterne, at testbrugerne alle har samme baggrund. Og specielt hvad slags instruktionsmetode man skal vælge at bruge til de forskellige brugergrupper, når man vælger at lave asynkrone remote usabilitytests. Man skal også tage med, at de 54 lærere, som deltog i testen, udgør ca 7 % af hele lærerstaben i folkeskolen på Færøerne. Repræsentativt er det ganske højt for en test.

## Effekt af datakvalitet, single vs pair-wise

I tabellerne 9 og 10 på side 29-30, ser man antallet af usabilityproblemer fordelt på seriøsitet og efter instruktionstype. Når man ser bort fra den kombinerede instruktionstype, ser man at single users finder flere problemer i alle instruktionstyper. Parrene finder generelt færre problemer, og en mulig forklaring kan være at det skyldes den kommunikationsmåde, parrene er blevet instrueret i at bruge under testen.

Constructive Interaction. Når der er to personer, som snakker sammen om komplekse ting, hjælper de igennem denne dialog hinanden med at forstå og løse mulige hindringer, så de aldrig bliver til noget problem hos dem. Da de løser opgaverne sammen som par, mens de konstruktivt interagerer med hinanden, får de en fælles forståelse for at løse opgaverne på et tilfredsstillende sæt for dem, og derfor kommer de aldrig ud i den situation, at det faktisk er et usabilityproblem. Miyake kalder det for “motion”, når en person forslår for den anden en ny eller anden måde at løse problemet på. (1986,Miyake) Ved at vende og dreje problemet med hinanden, når det opstår, flytter man sig fra det ene sted hen til en tilfredsstillende løsning. Derfor kan man undre sig, at der ikke bliver flere usabilityproblemer indbrettet specielt fra parrene. Men det er altså ikke tilfældet i denne test. Det kan diskuteres, om selve testoplægget med opgaver har skyld i, at parrene ikke indberetter flere usabilityproblemer. Bliver parrene så fokuserede på at løse opgaverne, at de “glemmer” at skrive ind i formularen, når der er noget at indberette, eller klarer de sig i gennemsnit bedre end single users. For at komme nærmere til et svar på denne problematik, fik jeg efterfølgende lavet et interview med et af parrene, som var med i testen. Interviewet blev optaget og er blevet transkriberet og vedlagt som bilag nr 7.

Lydfilen ligger også vedlagt som bilag 7a.

Testpersonerne blev bedt om at forklare, hvordan de havde forstået hele test oplægget og vejledningen, som de fik tilsendt med mail, og om at forklare hvordan de havde udført testen. De forklarede, at de havde forstået det sådan at de skulle løse opgaverne sammen. Dvs. de skulle hjælpe hinanden at løse opgaverne. Da jeg så spurgte til indberetning af problemer, som de stødte på, svarede de at når de stødte på forhindringer, brugte de mere hinanden til at løse opgaven, end at diskutere hvorvidt det var et problem, som skulle indberettes. De havde forstået det sådan, at deres primære opgave var at løse opgaverne,

og derefter fortælle noget om dem. Det bekræfter min påstand om, at parrene supplerede hinanden i at løse opgaverne hinanden at, mere end at finde og indberette brugbarhedsproblemer. Det giver igen anledning til at sætte spørgsmålstejn ved min metode til vejledning af testbrugerne. Hvis jeg havde brugt en anden metode til vejledning af parrene, ville udfaldet nok blive et andet.

Når singleusers sidder og udfører testen i deres ensomhed, har de kun sig selv at interagere med. Og da man sidder alene med en opgave, og man støder på en hindring, har man så ingen at spørge om hjælp, og giver derfor hurtigere op, og kommer også hurtigere frem til den konklusion, at dette er et usabilityproblem, som de er stødt på. Single-users er også blevet instrueret i at 'snakke' højt med sig selv efter "Think a-loud" metoden. Da de går igennem opgaverne, giver "Think a-loud" metoden testbrugerne et formiddabelt værktøj at formulere sig, når de støder ind i problemer. Indberetningerne fra single users er rigtig godt formuleret, hvilket tyder på at metoden hjælper dem at formulere problemerne. Dette kan derfor være en grund til at der kommer næsten dobbelt så mange usabilityproblem indberetninger fra single users end fra pairs. Fig 1 på s. 37 viser dog, at single-users bruger mere tid på hver enkelt opgave end pairs.

### **Forskel i testresultater i forhold til parsammensætning:**

Als et al (2005, Als et al) påviser at Constructive interaction med børn, som kender hinanden, finder flere problemer end par som ikke kender hinanden. I min test har jeg ikke haft som parameter, at parrene skal kende hinanden, men de er kollegaer på samme skole, og derfor går jeg ud fra, at de kender hinanden. På resultaterne ser jeg, at parrene mand-kvinde finder flere problemer end mand-mand og kvinde-kvinde. Jeg ser også at antallet af indleverede svar har samme mønster. Mand-kvinde sammensætningen indleverer flest indberetninger, mens de andre parsammensætninger undlader at indberette svar.

Parrene blev delt op i tre hovedgrupper. I første gruppe var der 6 par, bestående af kvinder, i næste gruppe var der 6 par af mænd og i tredje gruppe var der 6 par bestående af mand+kvinde. Testresultaterne viste, at parsammensætningen KK indberetter et samlet

tal på 27 problemer og MK indberetter 22 problemer, mens MM kun indberetter 7 problemer. Da vi kigger på fordelingen af par, som har svaret på testen, ser vi at 5 ud af 6 MK par svarer, men i de to andre kombinationer, KK og MM, svarer kun 3 ud af 6 par (50%).

Man bør overveje hvilken metode er mest hensigtsmæssig at bruge i et testmiljø. Baseret på hvor mange lærere man har adgang til som testpersoner, bør man vælge singleuser situationen med think aloud frem for pair-wise situationen med Constructive Interaction.

### **Begrænsninger og future work**

Dette eksperiment har nogle begrænsninger, som jeg vil tage frem i dette afsnit. Selve testen blev sat op i et php site med mySql database. Derudover blev test-brugerfladen kombineret således, at test-opgavene og svarformularen stod ude på højre side af vinduet og FirstClass brugerfladen, som skulle testes, var broderet ind med i et iframe. Idéen med dette var at forenkle brugerfladen, så testbrugerne skulle slippe for at skifte mellem flere vinduer under testen. Men de viste sig at være et u hensigtsmæssigt valg, da teknikken ikke ville det samme som mine intentioner. Hver gang en testdeltager havde udfyldt formularen med svar og trykt på send, blev hele vinduet genindlæst og nulstillet. Man blev logget ud af FirstClass, og måtte derfor igennem hele logon proceduren efter hver indberetning. Det er en sikkerhedsforanstaltning hos FirstClass, som gør at ved mindste forstyrrelse i netforbindelsen, så bliver man logget af.

Et andet problem viste sig også at være browserproblemer. Generelt fungerer FirstClass i alle browsere, men i dette kombinerede vindue, som jeg havde sat op, ville det bare ikke fungere. Det var specielt Safari browseren, som ikke ville samarbejde, men problemet dukkede periodisk op på de andre browsere.



Screenshot af testsitet med FirstClass indlejret som iFrame.

Disse tekniske forhindringer blev først observeret, da testen var gået i gang. P.gr.a tidspres, valgte jeg at sende en teknisk vejledning ud til alle testdeltagerne, hvor problemerne blev forklaret og hvordan man kunne komme forbi dem. Men det har helt sikkert påvirket testdeltagerne i den negative retning. Jeg kunne også have valgt at aflyse testen og konstruere den på nyt, men det ville være umuligt at nå det indenfor afsatte tidshorisont, og det ville også være umuligt at få testdeltagerne til at være med på det. Derfor blev det bestemt at gennemføre testen med de begrænsninger, som måtte være. Dette kan også være en årsag til at den samlede testdeltagelse kun var på 66.67%. Der var kun 61% af parrene og 72% af enkeltbrugerne, som indleverede svar. Den tekniske konstruktion af test-setuppet, er blevet konstrueret af mig selv alene, med inspiration fra litteraturen. Til fremtidige undersøgelser vil jeg bruge mere tid på at finde og bruge de tekniske løsninger, som allerede findes, istedet for at bygge det op fra scratch, for at undgå at tekniske problemer bliver forstyrrende elementer i en test.

Analysen af testen og konverteringen af indberetninger til usabilityproblem, blev udført kun af mig selv. Det, at det kun har været en person til analysere og konvertere testdataene, gør testresultaterne lidt ensidige. Hertzum og Jacobsen kalder dette evaluator



effekten: *“We believe that the principal cause for the evaluator effect is that usability evaluation is a cognitive activity, which requires that the evaluators exercise judgment”* (2003, Hertzum og Jacobsen). Havde der været flere evaluatore med i testen, ville det være muligt at sammenligne de forskellige analyser og derfra muligvis komme frem til et nuanceret resultat.

### **Testbruger commitment:**

Det er muligt at de tekniske forhindringer har været en årsag til at der ikke kom flere testsvar ind, men man kan ikke se bort fra, at selv om man har fået samtykke fra alle testdeltagerne om at de vil være med i testen, så er der alligevel flere deltagere, som vælger ikke at gennemføre testen, og vælger ikke at sige fra, at de ikke gennemfører. Det commitment, som jeg fik fra dem, synes ikke at være godt nok for alle. Det ville være interessant at undersøge om det er et lokalt problem, eller om det samme mønster viser sig globalt. Tullis et al berører samme emne i artiklen *“An Empirical Comparison of Lab and Remote Usability Testing of Web Sites”*. De har en hypotese, som siger at *“users in the remote test may feel more anonymous, and thus more willing to be critical, since they are not physically in the same place as the people conducting the test”* (Tullis et al).

Hver dag bliver vi bedt om at deltage i surveys og brugerundersøgelser via mail og besøg på diverse hjemmesider. Det er meget forskelligt, hvordan brugerne reagerer på sådanne anmodninger. De fleste springer over, men er der noget, som man er involveret i på en eller anden måde, så afsætter man tid på det alligevel.

Det, at jeg har kontaktet lærerne personligt og fået positive tilbagemeldinger, burde være nok i sig selv. Men alligevel sker det ikke. Er det den asynkrone remote test situation, som gør at testbrugerne alligevel ikke vælger at deltage. Hvis det havde været en remote synkron testsituation, ville testdeltagerne være nødt til at deltage, eller klart melde ud, at de alligevel ikke ønsker at deltage. Asynkront er det meget lettere at lade være, uden at tabe ansigt.

Emil Kruuse omtaler i sin bog *“Kvalitative forskningsmetoder”* om situationsfejl i forbindelse med undersøgelser, der har relation til bl.a. stedet. Her nævner han på p.260

Hyman, som siger, at jo tættere geografisk man er på dem, der skal observeres og interviewes, des større er chancen for, at de indsamlede oplysninger er valide. I min empiriske undersøgelse er det modsatte tilfældet. Testpersonerne i min undersøgelse er så langt væk geografisk, som de kan være. Den eneste kontakt jeg har haft med dem, er noget mail korrespondance. Hermed er det ikke sagt, at tæt geografisk tilknytning behøver at give mere troværdige informationer. Ønsket om at behage kan også betragtes som en mulig fejlkilde, idet nogle mennesker ønsker at behage i en sådan grad at de prøver at regne ud, hvad man gerne vil høre, frem for at sige noget, som de tror er ubehageligt for undersøgelsen og testlederen. (2007, Kruuse)

Til min test, valgte jeg at kontakte det præcise antal testbrugere, som jeg havde brug for til testen. Til næste test vil jeg nøje overveje at kontakte mindst 1/3 flere testbrugere end jeg har brug for, så jeg kan kompensere for det frafald af testpersoner, som er tilfældet i denne undersøgelse.

Kruuse nævner diverse fejlkilder, som man bør inddrage, når man skal bedømme validiteten af en kvalitativ undersøgelse (2007, Kruuse). Jeg vil påpege nogle mulige fejlkilder i min undersøgelse. Den første mulige fejlkilde, er om testspørgsmålene er godt nok formulerede, så testpersonerne forstår dem efter min hensigt. Har jeg glemt at spørge om noget væsentligt, eller har jeg stillet de forkerte spørgsmål. Har det påvirket testbrugernes indstilling til selve testen?

En anden mulig fejlkilde er utroværdighed. Kruuse nævner at det kan påvirke resultaterne af det kvalitative arbejde, eftersom det er den samme, der stiller spørgsmålene og fortolker svarene. Jeg mener dog selv at have været neutral i forhold til spørgsmålene og fortolkning af svarene. Det ville gavne testen, at der havde været en eller flere andre evaluatore med til at analysere og diskutere dataene.

## Future Work

De erfaringer jeg har fået ud af at eksperimentere med remote asynkron usabilitytest, vil jeg tage med mig til næste testprojekt jeg kommer at deltage i. Jeg har fået afklaret et par, for mig, vigtige ting mht. testopsætningen af remote asynkron usabilitytest. Det asynkrone remote testsetup har brug for at blive undersøgt nærmere. Hvordan kan man få bedre testbruger commitment, hvordan formuleres spørgsmålene, hvordan skal remote testdeltageren indberette sine svar? Er single-user opsætningen at foretrække frem for pair-wise i alle asynkrone tests? I dette projekt, har jeg arbejdet helt alene. Jeg er af den indstilling, at flere hjerner arbejder bedre end en, og et projekt af denne størrelse vil blive både bedre og mere givende for alle parter, hvis der er flere om at udføre det. Med hensyn til bearbejdning af testresultaterne, vil jeg inddrage personer med statistik ekspertise, udover at jeg selv også vil sætte mig grundigt ind i denne disciplin.

Det vil være fornuftigt at bygge på de tekniske løsninger, der allerede findes indenfor feltet, i stedet for at bruge tid og energi på at konstruere den tekniske løsning selv fra grunden.

## Konklusion:

I denne opgave har jeg foretaget en undersøgelse af effekten af instruktionstyperne deduktiv, induktiv og kombineret deduktiv + induktiv instruktion i en opgavebaseret remote asynkron usability test, hvor testbrugerne er delt op som enkeltbrugere og par. Jeg har yderligere kigget på, om man kan påvise forskel i testresultaterne i forhold til sammensætning af par, som er opdelt efter køn?

De resultater jeg er kommet frem til, viser at indsamlingen af data klart er påvirket af hvilken instruktionstype bliver brugt. I single-user test setup viser resultaterne i min undersøgelse, at den deduktive instruktionstype finder 26 usabilityproblemer, mens induktiv og kombineret instruktionstype finder 20 og 2 usabilityproblemer. Dette resultat stemmer ikke overens med resultaterne hos Bruun og Stage, som viser at det er den induktive instruktionstype, som finder mest usabilityproblemer, og dernæst den

kombinerede type (20 problemer), mens den deduktive type finder mindst antal problemer (2 problemer).

I parsammensætningen er det derimod den kombinerede instruktionstype, der finder mest usabilityproblemer (16 problemer), dernæst den deduktive (12 problemer) og siden den induktive type (6 problemer). I testsituationen har parrene brugt konstruktiv interaktion at kommunikere sammen ved løsning af opgaverne. En årsag til at der kommer så få indberetninger fra parrene i forhold til single-users, kan være at denne konstruktive interaktion er medvirkende til, at parrene løser opgaverne bedre og derfor ikke møder så mange problemer, som single-users.

Det vil sige at single-user generelt finder langt flere usabilityproblemer end parrene, og leder det til spørgsmålet om den, i denne opgave, anvendte metode at lave remote asynkron usabilitytest med pair-wise opsætning, er brugbar. Ud fra hvad det er man skal teste, må man vurdere hvilken opsætning man skal vælge, fordi det er ikke altid lige givende at teste med par i modsætning til single-user. I dette tilfælde var systemet, som blev testet, et mail og kommunikationssystem, som er et personligt redskab for den enkelte bruger. Testsituationen, hvor man sidder sammen som par, er derfor meget kunstig, og vil ikke finde sted i virkeligheden, fordi den almindelige bruger vil værne om sit privatliv og aldrig lade en anden person sidde og kigge med på et personligt kommunikations site.

Den mest sandsynlige situation er at brugerne sidder alene ved deres computer/device, logget ind på deres personlige brugersite og derfra kommunikerer med de andre i systemet. Testsituationen med single-users er derfor den bedste gengivelse af virkeligheden. Er pair wise testing da noget, som vi kan bruge til noget? Brugerfladen er designet til single users og ikke til pairs. De normale brugere kommer aldrig at sidde og løse opgave sammen på et produkt, som er så personligt, som et mail og konference system.

I min problemformulering spørger jeg også til om man kan påvise forskel i testresultaterne i forhold til sammensætning af par efter køn. Mine resultater viser, at man kan ikke påvise nogen forskel på kvaliteten på testresultaterne, men derimod kan man

påvise et testbruger commitment ved deltagelse i testen. Her ser man at i kombinationen mand-kvinde deltager 5 ud af 6 par (83%), mens i de andre sammensætninger, mand-mand og kvinde-kvinde, deltager kun 3 ud af 6 par (50%).

Min konklusion er, at man bør nøje overveje brugen af pair-wise asynkron remote usability testing. I mit eksperiment har pair-wise asynkron remote testing ikke vist sig at være den rigtige opsætning til denne slags testning. Jeg har ikke fået nok af brugbar feedback i forhold til indsats. Men jeg har dog fået afklaret, at single users er langt nemmere at arbejde med end pairs. Den asynkrone testsituation er meget lettere at håndtere for single users, da de kan administrere deres egen tid, uden at tage hensyn til nogen anden. Parrene, derimod, er bundet til aftale og organisere deres tid til at afvikle testen i et tidsrum, hvor begge personer er til rådighed på samme tid.

## Litteratur:

Andreasen, M. S., Nielsen, H. V., Schrøder, S. O. and Stage, J. (2007) What Happened to Remote Usability Testing? An Empirical Study of Three Methods. Begole, B. *et al.* (Eds.), *Proceedings of Conference on Human Factors in Computing Systems 2007 (CHI 2007)*, pp. 1405-1414. New York: ACM Press.

Beck, J Pædagogik – Opdragelse og selvverd, *Nordisk Forlag A.S., København (1986)* p74.

Benedikte S. Als, Janne J. Jensen, Mikael B. Skov, (2005) Comparison of Think-Aloud and Constructive Interaction in Usability Testing with Children

Bruun, A., Gull, P., Hofmeister, L. and Stage, J. (2009) Let Your Users Do the Testing: A Comparison of Three Remote Asynchronous Usability Testing Methods. *Proceedings of Conference on Human Factors in Computing Systems 2009 (CHI 2009)*, pp. 1619-1628. New York: ACM Press.

Bruun, A. and Stage, J. (2012) The Effect of Task Assignments and Instruction Types on Remote Asynchronous Usability Testing. *Proceedings of Conference on Human Factors in Computing Systems 2012 (CHI 2012)*, pp. 2117-2126. New York: ACM Press.

Castillo, J. C. (1997). *The User-Reported Critical Incident Method for Remote Usability Evaluation*. Master thesis, Virginia Polytechnic Institute and State University (1997).

C. Lewis, J. Rieman (1994) Task-centered User Interface Design, A Practical Introduction, chap. 5

Fohr, S.D. The Deductive-Inductive Distinction, *Informal Logic Newsletter* (1980b), 3(1), 5-10.

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: Are effectiveness, efficiency and satisfaction really correlated? In: *Proceedings of CHI 2000 Human Factors in Computing Systems* (pp. 345–352). ACM Press.

Hertzum, M. and Jacobsen, N. E. The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction* 15, 1 (2003), 183-204.

Kruuse, E., Kvalitative Forskningsmetoder- i Psykologi og beslægtede fag. *Dansk Psykologisk forlag 6. udgave (2007)* p245-264

Martin, R., Shamari, M.A., Seliaman, M.E., Mayhew, P (2014) Remote Asynchronous Testing: A Cost-Effective Alternative for Website Usability *International Journal of*

*Computer and Information Technology* (ISSN: 2279 –0764) Volume 03–Issue 01, January 2014

Miyake, N. (1986) Constructive Interaction and the Iterative Process of Understanding. *Cognitive Science* , vol. 10(2), pp. 151 – 177

Molich, R. *Usable Web Design*. Nyt Teknisk Forlag, Odense, Denmark, 2007.

Nielsen, J; Usability 101: Introduction to Usability; <http://www.nngroup.com/articles/usability-101-introduction-to-usability/> January 4, 2012

Prince, M.J. and Felder, R.M. Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Engineering Education* 95 (2006), 123- 138.

Rubin, J. and Chrisnell, D (2008) Handbook on Usability Testing – How to Plan, Design, and Conduct Effective Tests . John Wiley & Son, Inc.

Tullis, T. (2008). Automated Usability Testing: A Case Study. *User Experience Magazine*, 7(3). Retrieved from [http://uxpamagazine.org/automated\\_usability\\_testing/](http://uxpamagazine.org/automated_usability_testing/)

Tullis, T., Fleischman, S., McNulty, M., Cianchette, C. and Bergel, M. An empirical comparison of lab and remote usability testing of web sites. <http://www.testapic.com/-medias/RemoteVsLab.pdf>

Thompson, J. A. *Investigating the Effectiveness of Applying the Critical Incident Technique to Remote Usability Evaluation*. Master thesis, Virginia Polytechnic Institute and State University, 1999.

Soucy, K; *Unmoderated, Remote Usability Testing: Good or Evil?* Published: January 18, 2010; <http://www.uxmatters.com/mt/archives/2010/01/unmoderated-remote-usability-testing-good-or-evil.php>

