

Rapport PLAY POWER!

Skrevet av Lena Risvik, Hanieh Ali Bakhsh, Niklas Jacobsen og Rita Johnsen

1. Innledning

2. Presentasjon av prosjekt

2.1 Prototypen

2.2 Hva vi skal teste

3. Teori

3.1 Triangulering

3.2 Forskningsmetoder

- [3.3 Hypoteser](#)
- [3.4 Mini etnografi](#)
 - [3.4.1 Notater fra etnografiet#](#)
- [3.5 Valg av hypoteser](#)
- [4. Metode](#)
 - [4.1 Rammeverk](#)
 - [4.1.1 Elektronisk vs. fysisk](#)
 - [4.1.2 Kjønn](#)
 - [4.1.3. Alder](#)
 - [4.1.4 Random errors og systematic errors](#)
 - [4.2 Planlegging#](#)
 - [4.3 Gjennomføring](#)
 - [4.4 Usability testing](#)
 - [4.4.1 Evalueringen](#)
 - [4.4.2 Valg av representative brukere](#)
 - [4.4.4 Oppgaveliste](#)
 - [4.4.5 Målinger](#)
 - [4.4.6 Usability testen på film#](#)
 - [4.5 Observasjon](#)
 - [4.5.1 Notater fra observasjonen#](#)
 - [4.6 Intervju](#)
- [5. Analyse av data](#)
 - [5.1 Analyse av kvantitative data](#)
 - [5.1.1 Deskriptiv statistikk](#)
 - [5.2 Analyse av kvalitative data](#)
 - [5.2.1 Hovedtrekk](#)
 - [5.2.2 Sammenlikning av hovedtrekkene med hypotesene](#)
- [6. Resultater av analysen](#)
 - [6.1 Knytte resultatene opp mot bruk](#)
 - [6.2 Bias og resultat](#)
- [7. Evaluering av prosjektet](#)
- [8. Kilder:](#)

1. Innledning

Vi har valgt oppgave 5 som er et prosjekt for Oslo sitt kommende barnemuseum, opprettet i kjølevannet av 22 juli. Hovedhensikten er å hjelpe barn gjennom lek. PLAY POWER! som prosjektet heter (har skiftet navn fra PLAY HELPS!), kan f.eks. være styrkende for barn i en sørgeprosess. Vi valgte denne oppgaven fordi vi synes barn er en spennende målgruppe. Barn er åpne og direkte, gir deg ærlige tilbakemeldinger. De legger ikke skjul på hva de mener.

Vi har utviklet en applikasjon som er basert på den fysiske boomwhackers. Vi vil bruke denne applikasjonen til å teste tre hypoteser og prøve å finne ut hva som eventuelt kan forbedres med vår prototype. For å finne disse hypotesene studerer vi barna som interagerer med applikasjonen og de fysiske boomwhackers, ved å bruke et mini-etnografi. Når vi har utformet hypotesene bruker vi eksperimentell design som rammeverk til å designe eksperimentet. Deretter gjennomfører vi en usability test, observasjon og intervju. De kvantitative målingene analyseres gjennom statistikk. I tillegg til dette analyserer vi den kvalitative målingene for å få et bredere perspektiv på analysen. Dette bruker vi for så for å avkrefte eller bekrefte hypotesene våre som vi dannet etter mini-etnografiet.

2. Presentasjon av prosjekt

2.1 Prototypen

Prototypen er en musikkapplikasjon som baseres på det allerede kjente musikkspillet Boomwhackers. Applikasjonen består av kun en "side", og det er derfor ikke mulighet til å navigere. Applikasjonen inneholder 8 "bilder" som ligger horisontalt over hverandre. Alle representerer hver sin boomwhacker, og når du interagerer med de, via touchskjemer, spiller de tilhørende tone. De har tilnærmet lik farge som boomwhackers'ene har. Applikasjonen tar automatisk opp det brukeren spiller slik at de kan høre på det de har spilt i etterkant. I tillegg til de 8 "bildene", så er det to knapper plassert nederst på skjermen. En knapp "reset'er" det som er tatt opp, og brukeren kan spille inn på "blanke ark". Den andre knappen er for avspilling, når brukeren trykker på denne, spilles eventuelt tidligere spilte toner tilbake til brukeren i samme tempo som brukeren først spilte i. Vi har valgt å bruke knapper som har likhet med andre vanlige brukte knapp-symboler. Tilbakespillings-knappen er en trekant, som er vanlig å bruke som "play-knapp". Og reset-knappen ser ut som symbolet som vanligvis brukes til "refresh" i nettlesere. Dette er så brukeren forhåpentligvis skjønner bruken av knappene enklere.

2.2 Hva vi skal teste

Vi skal teste applikasjonen på barn og se hvordan de bruker applikasjonen i forhold til en fysisk utgave. Vi vil også teste om alder eller kjønn har betydning for hvordan de interagerer med produktet. Vi synes dette kan være interessant å se, fordi da kan vi presisere hvilken aldersgruppe applikasjonen passer til. Vi kan også da si noe om applikasjonen vår er et godt alternativ til/erstatning for boomwhackers som kanskje krever mer samarbeid, kroppskontroll og en mer viderekommende musikkunnskap. Dette er ting som kanskje ikke er så godt utviklet på barna enda.

Prototypen har mange bruksområder, og vil derfor kunne nå ut til en stor målgruppe. Den kan både bli brukt som et instrument og et leketøy. Ved å ha en enkel funksjonalitet gjør vi den mer brukervennlig for barn. I denne oppgaven er målgruppen avgrenset til barn mellom 3-12år. Vi har etterhvert konkretisert målgruppen ettersom vi har blitt mer sikre på hvordan vi skal teste prototypen.

3. Teori

3.1 Triangulering

For å komme frem til de mest nøyaktige resultatene er det nødvendig at flere researchere forsker på prosjektet samtidig og over tid. Det bør brukes forskjellige tilnæringsmetoder, også kalt forskningsmetoder, for å hente inn data. En tilnærming vil ikke gi et endelig svar på virkeligheten. Derimot vil resultatene bli mer realistiske hvis man over tid og ut ifra forskjellige forskningsmetoder oppnår konsensus (enighet om funnene). Dette vil gi et bevis for at «slik er det», altså en forskningssannhet. Det å samle inn data på forskjellige måter kalles triangulering.¹

3.2 Forskningsmetoder

Å bestemme seg for hvilke forskningsmetoder som passer best i et studie, avhenger av forskjellige faktorer. Det er viktig å se på hensikten med studiet, tid, penger, mengden av deltagere, og forskernes tidligere erfaringer.² Vi har i prosjektet hatt begrenset tid på en måned til testing og skriving av rapport. Målgruppen vår er barn, og vi har måttet finne de studiene som passer best til denne gruppen. Spørreundersøkelser, case-studier, og dagbok-studier har for eksempel vært noen metoder som vi har vurdert, men som vi har konkludert med har egnet seg mindre på grunn av bruk, alderen til målgruppen og tid. Case-studier og dagbok-studier krever at vi går i dybden på en rekke deltagere over en lengre periode enn det vi kan tillate i dette prosjektet. Disse studiene ville gitt oss en større innsikt i barnas spesielle atferd og generelle hverdagsliv. I større grad ville dette hjulpet oss til å finne ut av brukeres aktiviteter og behov, som er mer relevant i designfasen, enn for å teste en applikasjon som allerede er laget. Samtidig ville vi også trengt utstrakt tillatelse fra foreldre for enten å observere barna hjemme eller benytte oss av personlig materiale fra barna. Dette mener vi ville blitt alt for ressurskrevende og komplisert i dette prosjektet.

¹ Lazar, Jonathan, Jinjuan Heidi Feng og Harry Hochheiser. *Research Methods In Human- Computer Interaction*. (United Kingdom: John Wiley & Sons Ltd), 6.

² Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 20.

Hensikten med vårt studie er å finne ut hvilken nytteverdi produktet har, hvordan vi kan forbedre produktet og konkretisere målgruppen ytterligere. Spørreundersøkelser (surveys) er en annen studieform som vi har vært innom. Denne beskrivende undersøkelsen (descriptive investigation)³ kan være effektiv for å nå ut til en stor brukergruppe på kort tid. Et negativt aspekt derimot er at spørsmålene lett kan misforstås eller være formulert feil. Dette kan gi bias ved at brukeren svarer noe annet enn det som er ment, eller uriktig ved at brukeren mangler det riktige svaralternativet. Å sette sin lit til en slik undersøkelse på barn kan være risikabelt ved at det er vanskelig å få gode svar. Det krever ytterligere skriftlige kunnskaper, tålmodighet og selvstendig tenkning fra barn enn det vi kan forvente i denne alderen.

Vi har derfor besluttet at usability-testing ville være en god metode å bruke for vårt formål og for å gjennomføre testingen. Før vi hadde denne usability testen, gjennomførte vi et mini-etnografi. Dette mini-etnografiet har vært med på å hjelpe oss til å få en forståelse av det vi skal undersøke, og presisere det vi har lyst til å utforske med applikasjonen. Vi har brukt etnografiet for å finne passende hypoteser som vi vil se nærmere på og konsentrere arbeidet rundt. Vi har i løpet av denne måneden altså utført to studier: et minietnografi og en brukbarhetstest. I mini-etnografiet observerte vi barna, mens i brukbarhetstesten hentet vi inn informasjon med forskjellige undersøkelsesmetoder som intervju, observasjon, og en brukbarhetstest. Dette gjorde vi for å komme frem til de best mulig resultater og konsensus innad i gruppa.

3.3 Hypoteser

Hypotese er en presis problemstilling som kan testes empirisk. Hypotese kan ses på som en antakelse av virkeligheten. Teori kommer ofte etter en hypotese har blitt testet, hvor hypotesen gir et mer sant enn falsk bilde av virkeligheten.^{4 5}

I et eksperiment har man minst en nullhypotese og en alternativ hypotese. En nullhypotese er en betingelse eller et krav som er tilfredsstillt, og ikke viser noen forskjeller mellom gruppene man undersøker hypotesen på. En alternativ hypotese er en påstand/setning som er det motsatte av en nullhypotese (gjensidig utelukkende).⁶

Vi har utledet tre hypoteser ved hjelp av et mini-etnografi som vi utførte i et bursdagsselskap. Dette vil si at vi samlet inn data om barna ved hjelp av observasjon,

³ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 21.

⁴ Wikipedia. *Hypotese* (Wikipedia, 24 august 2011), hentet 23 oktober 2011.

⁵ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 22.

⁶ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 23.

før vi fant ut hva vi ville forske videre på. Dette kalles en grounded theory teknikk. Grounded theory er data som systematisk blir samlet inn og analysert. Det er en induktiv forskningsmetode. Induktiv vil si at man på bakgrunn av et begrenset materiale slutter seg til en større generalisering av det man studerer. Det er en metode som er veldig ulik fra den tradisjonelle eksperimentelle forskningsmetoden. Der er det typiske å starte med noen velformulerte hypoteser basert på teori, for så å utføre et eksperiment for innhenting av data (noe vi også gjorde etter vi fant hypotesene). Denne dataen skal underbygge og bevise teorien. Grounded teori starter derimot med et empirisk studie og utvikler en velbegrunnet teori ut ifra det man erfarer i studiet. I løpet av denne teoretiske prosessen kan man gjennomføre flere datainnsamlinger og analyser for at den skal være mest mulig korrekt. Denne metoden kalles ofte «reverse-engineered» hypotesen.⁷

Grounded theory kan benyttes i en rekke metoder som etnografi, observasjon, case studier, og intervjuer. Den største forskjellen mellom kvalitativ forskningsstrategier som hovedsaklig er beskrivende eller utforskende og grounded theory er dens vektlegging av teori, basert og utviklet fra et kontinuerlig samspill mellom datainnsamling og analyse. Ingen hypoteser er formulert før studiet. Forskerne skal la dataene føre til hypotesene. Dette er en åpen og kreativ prosess, og man skal ikke ha tenkt seg ut noen potensielle løsninger på forskudd.⁸ Vi gikk derfor åpne inn i etnografistudiet. Vi hadde allerede tenkt ut en del hypoteser, men la disse fra oss. Vi rakk av tidsmessige årsaker bare å samle inn data til hypotesene under dette etnografiet. Det er som tidligere nevnt vanlig og ideelt å gjennomføre flere forskningsstudier før man utvikler teorien. Dette hadde vi også gjort om vi hadde hatt bedre tid.

3.4 Mini etnografi

Etnografi er en undersøkelsesmetode som beskriver og sammenligner ulike kulturer og samfunnstyper, ofte gjennom feltstudier og observasjoner⁹.

I HCI (Human Computer Interaction) blir etnografi ofte brukt som et første steg, for å forstå en gruppe mennesker og deres problemer, utfordringer og normer, med et mulig mål som for eksempel å lage et teknologisk produkt for dem eller sammen med dem. Etnografiske studier skjer ofte i hjemmet, jobb, utdanningsmiljøer, eller over Internett. Siden studiet er i så kjente omgivelser for deltageren finner man ofte ut hva deltageren virkelig gjør, og trenger ikke lengre å tro eller anta.¹⁰

Med mini-etnografisk studie mener vi bare å lage en liten versjon av studiet. Dette fordi

⁷ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 283

⁸ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 284

⁹ Wikipedia. *Etnografi* (Wikipedia, 15 Mai 2011), hentet November 2011.

¹⁰ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 246.

det vi ikke har tid eller mulighet til å gjennomføre et stort og ordentlig etnografistudie. Det vi tenkte i vårt mini-etnografiske studie var å la barna leke med både den fysiske boomwhackeren og applikasjonen uten å få noen instruksjoner eller innblanding fra oss. Vi vil se hva de gjorde og hvordan de lekte. Hvilke problemer som oppstod. Hvilke utfordringer de hadde og når de gikk lei produktet osv. Dette gjorde det lettere å se hva som kunne være interessant å forske/undersøke videre og det å danne relevante og målbare hypoteser.

3.4.1 Notater fra etnografiet¹¹

3.5 Valg av hypoteser

Etter å ha observert barna fant vi ut at det kunne være interessant å se mer på kjønn og alder. Det var stor aldersforskjell på barna som deltok i mini-etnografiet, og vi så at de minste barna interagererte annerledes med applikasjonen/boomwhackerne enn de eldre barna. Det var også stor forskjell mellom jenter og gutter på hva de likte og hvordan de oppførte seg. Vi ble derfor nysgjerrige på om dette hadde noe å si for hvor godt de liker applikasjonen. Vi synes også det fortsatt er interessant å finne ut av hva de liker best av den fysiske og elektroniske applikasjonen. Derfor kom vi opp med følgende hypoteser:

H₀: Det er ingen forskjell på hvor lenge de leker med applikasjonen og de fysiske boomwhackers.

Uavhengige variabler: Type enhet: Fysisk (Boomwhackers) og elektronisk (applikasjon).

Avhengige variabler: Tid målt i minutter.

Condition 1 (vilkår): Fysisk (Boomwhackers).

Condition 2 (vilkår): Elektronisk (applikasjonen).

H₀: Det er ingen forskjell på hvor lang tid jenter og gutter leker med applikasjonen.

Uavhengige variabler: Type kjønn: Jenter og gutter.

Avhengige variabler: Tid målt i minutter.

Condition 1 (vilkår): Jenter leker.

Condition 2 (vilkår): Gutter leker.

H₀: Det er ingen forskjell på hvor lang tid barn i 1, 2 og 3 klasse leker med applikasjonen.

Uavhengige variabler: aldersgrupper: 1 klasse, 2 klasse og 3 klasse.

Avhengige variabler: Tid målt i minutter.

Condition 1 (vilkår): 1 klasse.

Condition 2 (vilkår): 2 klasse.

Condition 3 (vilkår): 3 klasse.

¹¹ Se vedlegg: Etnografnotater

4. Metode

4.1 Rammeverk

Vi har brukt rammeverket Experimental design for å planlegge Usability-testingen. I experimentell design har vi tre grupper studier: true-experiments, quasi – experiments, and non- experiments. Siden vi har minst en testbar hypotese og minst 2 vilkår velger vi true-experiment. Vi har bare en uavhengig variabel i alle hypotesene (type enhet, kjønn og aldersgruppe) derfor bruker vi basic design structure. Antall vilkår hypotesene kan være avgjørende når vi skal bruke between- group eller within- group design¹².

4.1.1 Elektronisk vs. fysisk

I den første hypotesen (elektronisk vs fysisk) har vi to vilkår. I denne hypotesen velger vi within-group, dvs at alle deltagere blir utsatt for alle vilkår. Dette fordi eksperimentet undersøker oppgaver med store forskjeller. Det er stor forskjell på fysiske boomwhackers og vår applikasjon, både fysisk form, musikk og utførelse. Et annet argument for dette valget er at oppgavene som blir valgt er mindre mottakelige for læringseffekten. Siden applikasjonen og den fysiske boomwhackeren er så ulike i form og bruksmål kan det bli vanskelig å få en læringseffekt. Men for å være på sikre siden prøver vi å unngå dette ved å bytte om rekkefølge for hva deltageren blir introdusert for først mellom gruppene. Hvis første gruppe blir først introdusert for applikasjonen så boomwackerne, får neste gruppe motsatt rekkefølge.

4.1.2 Kjønn

I den andre hypotesen (gutter vs jenter) har vi to vilkår. I dette tilfellet kan ikke gutter være jente og omvendt. Dette gjør valget for eksperiment strukturen veldig lett, da eksperimentet ikke kan bli undersøkt gjennom within-group design. Derfor velger vi between-group design, som vil si at deltageren bare blir utsatt for et av vilkårene.

4.1.3. Alder

I den tredje hypotesen (forskjellige aldersgrupper) har vi tre vilkår. I dette tilfellet kan alder være vanskelig å undersøke gjennom within-group design fordi man ikke har tid til å følge en person gjennom flere år og se at de forandrer alder. Skulle man likevel hatt denne tiden ville likevel eksperimentet fått en stor bias fordi deltageren ville fått en læringseffekt.

4.1.4 Random errors og systematics errors

¹² Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 42-45.

Før vi starter eksperimentet må man tenke på random (tilfeldige feil) og systematic errors (bias). Tilfeldige feil oppstår ved en tilfeldighet og er ikke forbundet med den faktiske verdien. Det er ingen måte å fjerne eller kontrollere tilfeldige feil, men man kan redusere

virkingen av tilfeldige feil ved å forstørre den observerte utvalgsstørrelsen¹³. I alle eksperimentene vi gjennomfører har vi ikke mulighet til å få så mange deltagere som er ønsket. Dette er i hovedsak på grunn av tiden, men også på grunn av begrensninger i resurser som å finne barn, og kompetansenivået vårt innenfor gjennomføring av tidligere studier. Systematic errors eller bias som det også heter i alle eksperimentene kan være for eksempel at barnet har et form for handikap, ADHD (lav konsentrasjonsevne) eller er autist. Det kan også være en bias at deltageren blir nervøs foran oss. Dette vil vi prøve å unngå ved å la deltagerne komme inn som en gruppe slik at de har noen kjente ansikter og kanskje derfor føler seg tryggere og mindre nervøse.

4.2 Planlegging¹⁴

4.3 Gjennomføring

Vi fikk ved hjelp av veilederen vår Katie gjennomført en brukertest, observasjon og et intervju på Majorstuen skole i Oslo. Vi var alle fire tilstede under undersøkelsen. Vi avtalte på forhånd at to skulle observere barna, mens en skulle ta tiden og den siste skulle være til assistanse for elevene. Vi fordelte det slik at Niklas og Lena satt i hvert sitt hjørne og observerte, Hani skrev ned deltager.nr, navn, tid, alder og kjønn, og Rita snakket og veiledet elevene gjennom testen. Brukergruppen var barn fra 1-3 klasse, og det var 5 jenter og 7 gutter i alt.

4.4 Usability testing

Evalueringen er av en fullt fungerende prototype, altså en high fidelity prototype, og kalles derfor summativ testing.¹⁵ Vi anser at vår prototype er en high fidelity ut ifra det som står i boka. Den oppfylder de kravene en high fidelity prototype må ha. Den har alle funksjonene som et fullverdig produkt skal innholde. Brukerne kan interagere med prototypen, og alle funksjonene fungerer i spillet. Den er brukerdrevet, som en high fidelity prototype må være. Vi har brukt prototypen til utforsking og testing og den ser ut som det ferdige

¹³ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 57.

¹⁴ Se vedlegg: Plan for gjennomføring av usability testen, observasjonen og intervjuet.

¹⁵ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 260.

produktet skal se ut. Det er et fysisk eksemplar, og kan brukes til å selge og markedsføre produktet.¹⁶

4.4.1 Evalueringen

I summativ testing evaluerer deltagerne effektiviteten av et spesifikt design.¹⁷ I vårt tilfellet er det musikkapplikasjon vi har laget for Android-mobiler og tablets. Målet med brukertesten er å se hvordan barna interagerer med produktet, og om det er betinget av kjønn eller alder. Et annet mål er å finne ut om applikasjonens brukergrensensitt kan forbedres og optimaliseres. Vi vil at funksjonene skal bli så lettfattelige og effektive å forstå som mulig. Vi vil også se om barna liker applikasjonen og om de tar den i bruk, og dette i forhold til en boomwhacker. Det kan hjelpe oss å konkretisere brukergruppen ytterligere.

Testen ble gjennomført av oss pga. tid- og ressurs-begrensninger. Utviklerne av programmet utfører ikke alltid testen, men leier ofte inn noen utenforstående. Dette med tanke på at utviklere mest sannsynlig har brukt mye tid på systemet, og støtter det slik det er. De kan bli inhabile i saken ved at de har en sterk tilknytning til programmet, og av den grunn muligens ikke vil være åpne for nye løsninger og forslag.¹⁸

4.4.2 Valg av representative brukere

Brukergruppen som vi evaluerte med var av størrelse et «medium» prosjekt, det vil si et antall på tolv personer. Disse skulle representere målgruppen til vår applikasjon. I denne testsesjonen er alder og kjønn satt som hovedfaktor og ikke kultur eller språk. Målgruppen ble derfor tatt fra et område i Oslo, dette sparte oss for tid og penger. Målet var å se hvordan brukerne interagerer med produktet i løpet av en fastsatt tid. Dette var for å finne ut om det kunne være betinget av bakgrunn. Vi så også etter om de oppdaget noen svakheter med applikasjonen som kunne forårsake problemer. Dette bestemte vi skulle være en del av observasjonen.

Det er viktig å tenke på kostnader og tid når man gjennomfører en forskningsmetode. Alle deltagere skal ha en slags betaling, og vi må tenke på hvor mange vi har tid og ressurser til å ha med i evalueringen. Det at vi fikk hjelp av Katie til å finne deltagere gjorde at vi sparte tid.

4.4.3 Valg av setting

¹⁶ Rogers, Yvonne, Helen Sharp and Jeniffer Preece: *Interaction Design: beyond human-computer interaction*, (2. utgave. Wiley) , 536

¹⁷ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 260.

¹⁸ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*,, 260

Evalueringen foregikk på Majorstuen Barneskole. Deltagerne går på skolen. Brukertesten ble satt opp i et klasserom. Barna plasserte vi midt i rommet på tre pulter. Vi gjennomførte observasjonen samtidig som brukertesten utspant seg. Brukerne ble delt inn i tre grupper: 1.klasse, 2.klasse og 3.klasse. Gruppene fikk 14 minutter tilsammen med produktene. 7 minutter var satt av til applikasjonen og 7 minutter til de fysiske boomwhackerne. De ble i tur og orden sendt inn i klasserommet, og gjennomførte oppgavene med produktene. Mikrofon og kamera tok opp data av hvordan brukerne utførte testen. I tillegg tok vi opp intervjuet med tre mobiler og noterte ned observasjonen. Alt dette har vi brukt i analysen.

4.4.4 Oppgaveliste

Summative tester pleier å fokusere mer på oppgavenivå-målinger (task-level measurements), beregninger og kvantitative målinger. Brukernes oppgave var å nyttiggjøre seg av grensesnittet i løpet av en fastsatt tid. Vi ville ikke ha for mange og vanskelige oppgaver fordi deltagerne er barn fra 5-8 år. Barn er mindre tålmodige og vi synes ikke vi kunne forvente at de skulle utføre for avanserte oppgaver heller. Vi ville også at de skulle ha det moro under testen, slik at de faktisk gjennomførte den. Dette var en stor grunn til at vi valgte en slik åpen oppgave. Vi hadde ikke et scenario på forhånd, i og med at funksjonene er så enkle som de er.

4.4.5 Målinger

- Kvanitativ data:

- Tiden (målt i minutter) på hvor lenge de er aktive med enten applikasjonen eller de fysiske boomwhackerne. Deltagerne skal nyttiggjøre seg av grensesnittet så lenge de ønsker, og si fra/gi seg når de er lei. Makstid på syv minutter pr. produkt. Siden barn er såpass forskjellige valgte vi likevel å gi en tilleggsoppgave til de som ikke følte at det å trykke på applikasjonen var en stor nok utfordring. Dette var bare en enkel oppgave hvor vi prøvde å få de til å spille en sang. Vitsen med denne oppgaven var å se hvilket nivå de ligger på.

- Kvalitativ data:

- Observasjonsnotater

- Film fra brukertesten

- Intervju: Etter at deltagerne ble ferdige med brukertesten gjennomførte vi et lite intervju. Dette kan også kalles en slags «refleksjonssesjon» der deltageren kommentere hva de synes om grensesnittet og eventuelle forslag til forbedringer. Vi valgte å gjøre dette intervjuet og refleksjonen rett etter testen, istedetfor at brukeren kommenterer gjennomføringen av oppgavene underveis. Dette gjorde vi slik at det ikke gikk utover barnas konsentrasjonen på selve testen og prestasjonen til brukeren.

4.4.6 Usability testen på film¹⁹

4.5 Observasjon

Vi har gjennomført en åpen og indirekte observasjon. Det vil si at deltageren vet formålet av observasjonen men ikke hvorfor de blir observert.

For at deltageren ikke skulle bli skeptisk, påvirket, eller stresset ved at vi observerte dem, plasserte vi de to observatørene i bakgrunnen og litt unna deltageren. Deltagerne får derfor mindre fokus på observatørene. Dette er viktig slik at vi ikke mister informasjonen mens vi observerer. Vi valgte å observere deltageren ved å bruke både notater og video. Som en oppvarming ga vi alle deltagerne hvert sitt startnummer. Dette fordi det gikk lettere for observatørene å notere ned hva de forskjellig gjorde deltagene. Men dette ble også en oppvarmings del, hvor vi og de presenterte oss, mens vi delte ut nummer og fortalte hva formålet med studiet var, og om dette var noe de ville hjelpe oss med.

4.5.1 Notater fra observasjonen²⁰

4.6 Intervju

Etter brukbarhets-testingen tok vi et lite intervju. For å få et perspektiv til på problemstillingen vår og for å triangulere. Dette gir oss muligheten til å spørre om noe spesifikt vi kanskje observerte under brukbarhets-testingen, vi får brukerens egne oppfatning, og ikke bare vår egen tolkning av situasjonen. Vi kan da se om det barna sier og hva de faktisk gjør har en sammenheng. Vi har da valgt å bruke et semi-strukturert intervju. Grunnen til at vi har valgt det er fordi vi har en klar formening om hva vi vil vite noe om, så vi kan formulere og sette opp spørsmål om det vi vil vite i forkant av testingen. Men barn er veldig uforutsigbare, så vi vil også beholde muligheten til å stille oppfølgings spørsmål og kunne grave dypere om barnet skulle si noe som er nytt eller spesielt interessant. Vi lagde derfor en intervjuguide²¹, som skal fungere som en veileder under intervjuet, men den er ikke absolutt, så vi kan vike i fra denne. Dette er grunnen til at det derfor ikke ble full-strukturert intervju eller ustrukturert intervju.

Vi tok opp intervjuet slik at vi i ettertid kunne høre på det og transkribere det, isteden for å bruke tid på å ta notater. Spørsmålene vi kommer til å stille blir da en blanding mellom

¹⁹ Se vedlegg: Usability testen på film

²⁰ Se vedlegg: Observasjonsnotater

²¹ Se vedlegg: Intervju til brukertesten.

lukkede spørsmål og open-ended spørsmål. For eksempel så vil vi gjerne finne ut av hva de likte best av applikasjonen og boomwhackers, så vi kan stille dem; -Hvilken av disse to dere har prøvd nå likte dere best? Her er det bare to alternativer å velge mellom, og derfor er det et lukket spørsmål. Etter vi har fått svar på det kommer vi til å prøve å finne ut hvorfor de likte det ene eller det andre ved å stille oppfølgings spørsmål. Oppfølgings spørsmålet på det spesifikke spørsmålet blir da open-ended, og forhåpentligvis kan barnet utdype hvorfor han/hun mener det ene eller det andre. Som jeg sa tidligere så var dette et lite intervju, dette er fordi det er barn vi skal intervju, og dette ville nok ikke blitt særlig produktivt å prate om i en lengre periode. Barn har ikke ferdig utviklede kognitive evner og har større vanskeligheter med å uttrykke seg muntlig, enn en voksen ville hatt. En mulig bias med å intervju barn er at barnet kanskje blir litt skremt da det befinner seg i en uvandt situasjon og vi er fremmede for barnet. For å gjøre dette biaset så lite gjeldende som mulig har vi bestemt oss for å holde intervjuet ganske uformelt, og prate med enkle ord som barnet forstår. Grunnet tidsbegrensninger så fikk vi ikke tid til å intervju en og en, så vi valgte å ta det felles med hele gruppen som var inne til testing (hhv. 5, 4 og 3stk). Intervjuet fikk da noen likheter med en fokusgruppe. En fordel med at vi måtte gjøre dette er at det er mindre sjans for at barnet blir redd og ukomfortabel når hun/han er med klassekameratene sine. En annen fordel er at det kan oppstå diskusjoner imellom barna, og vi kan muligens få en bedre forståelse for hvorfor de mener det de gjør. En ulempe er at barnet kanskje ikke svarer det han/hun mener selv, men sier det de andre på gruppen sier, dette er et mulig bias i dataene våre.

4.6.1 Transkribering av intervjuene²²

5. Analyse av data

Data-analyse er like arbeidskrevende som data innsamling. Å velge analyseringsmetoder for tolkning av testresultater er essensielt for brukerundersøkelsen, og for å få de riktige resultatet. I denne fasen skal vi gå fra råstoff til mer nyttig informasjon. Dette gjør vi ved bruk av disse kvantitativ og kvalitativ analyse av metodene. I analysen studerer vi barnas bruk av de fysiske boomwhackers og den elektronisk versjonen. Hvordan de bruker og hvor godt de liker disse komponentene, kan avhenge av bakgrunn som blant annet alder og kjønn. Hypotesene vi har kommet fram til har vi undersøkt ved hjelp av forskjellige metoder. Dette har vært viktig for å få et mest mulig nøyaktig bilde av virkeligheten. Under brukbarhetstesten målte vi først og

²² Se vedlegg: Intervjutraskribering

fremst hypotesene med tiden de brukte på produktene. Dette vil være vår kvantitative data. I tillegg til dette samlet vi inn data ved hjelp av observasjon under brukbarhetstesten, video av brukbarhetstesten og et intervju i etterkant av testen. Dette vil være grunnlaget for den kvalitative dataen. Analyse av både kvantitative data og kvalitative dataen utgjør det endelige resultatene fra undersøkelsen.

5.1 Analyse av kvantitative data

Det første vi må gjøre med dataene våre før vi begynner å analysere dem, er å nøye klargjøre dataene for analysen. Det er flere grunner til å gjøre nettopp dette, og flere grunner til at det er viktig. En grunn kan være inkonsistens med formateringen av dataen, på alderen til deltagerene våre kan det for eksempel stå både 6 og seks. Formateringen må være konsistent og ikke gi rom for feiltolkning. Dette er også viktig å gjøre med tanke på om vi skal bruke et program til å gjøre analysen og utregningene for oss. Da disse statistikk-programmene ikke er programmert til å lese tekst²³. Det man bør gjøre først er å se over og rense opp i dataene. All data, men hovedsaklig data som er skrevet inn av mennesker, da det er menneskelig å feile. Dette kan være ting som skrivefeil og også formateringsfeil. Det viktigste her er å se etter åpenbare feil. For eksempel er alderen på en av deltagerne 67, og vi har testet på barn i mellom fem og åtte år, så er det en åpenbar feil i dataene. Her er det imidlertid ikke så enkelt å rette opp, alderen kan både være 6 og 7 og det er ikke lett og vite hva som er feil/riktig. All dataene innsamlet om den deltageren bør forkastes.²⁴ I dette tilfellet ville for eksempel dette hatt en stor påvirkning for gjennomsnittsalderen til deltagerne i testingen. Dataene bør også forkastes om de er ufulstendige. Vi kommer også til å kode dataene, det vil si, at istedenfor å oppføre kjønnen til en deltager som «Mann» og «Kvinne», så vil vi bruke kode 1 for mann, og kode 0 for kvinne. En veldig viktig ting her er å være konsistent på kodingen, og være nøye på at alle som behandler dataene er klar over hvilken kode som refererer til hvilket kjønn.

Når vi organiserer dataene er det også veldig viktig å være klar over hva slags krav til organisering det statistikk-programmet vi bruker har.

5.1.1 Deskriptiv statistikk

Når vi har renset opp i dataene så kommer vi til å kjøre et nummer forskjellige grunnleggende deskriptive statistiske tester. Vi er for eksempel interessert i å vite om

²³ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 70.

²⁴ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 71.

det er noen forskjeller på brukstiden for jenter og gutter. Her er gjennomsnittet på brukstiden fra de to forskjellige gruppene en viktig ting å se på. Vi kommer også til å sammenligne de fysiske boomwhackersene opp mot applikasjonen vår med tanke på brukstid. Men bare gjennomsnittstiden alene er ikke her nok til å si noe om det er en forskjell eller ikke. For det kan være andre faktorer som spiller inn her. Så derfor må vi foreta en statistisk-signifikans test, som evaluerer om variansene kan forklares med de kontrollerte uavhengige variabelene (i dette tilfellet, boomwhackers eller applikasjonen). Eller om de ikke kan det. Signifikans testen vil derfor si noe om sannsynligheten av at de observerte forskjellene har forekommet med sjanse. Hvis sannsynligheten av at forskjellene skjer ved sjanse er forholdsvis lav, f.eks 5% eller lavere, så kan vi si oss forholdsvis sikre på at de forskjellene vi har observert er grunnet forskjellen mellom de uavhengige variablene. Dette er også direkte overførbart til de andre hypotesene våre.

H_0 : Det er ingen forskjell på hvor lenge de leker med applikasjonen og de fysiske boomwhackers.

En Paired-samples t-test indikerer at det ikke er en signifikant forskjell på hvor lenge de leker med applikasjonen og boomwhackers.
($t(11) = 1,29, p > 0,05$).

H_0 : Det er ingen forskjell på hvor lang tid jenter og gutter leker med applikasjonen. For å si noe om hypotesen.

En one-way ANOVA test som bruker tiden de lekte med applikasjonen som avhengig variabel og kjønn som uavhengig variabel indikerer at det ikke er en signifikant forskjell på hvor lang tid de forskjellige kjønnene bruker applikasjonen før de går lei.
($F(1,10) = 1,34, p > 0,05$).

H_0 : Det er ingen forskjell på hvor lang tid barn i 1, 2 og 3 klasse leker med applikasjonen.

En one-way ANOVA test som bruker tiden de lekte med applikasjonen som avhengig variabel og klasstrinn som uavhengig variabel indikerer at det ikke er en signifikant forskjell på hvor lang tid de forskjellige klassetrinnene bruker applikasjonen før de går lei.
($F(2,9) = 0,19, p > 0,05$).

Basert på dette kan vi ikke forkaste noen av de tre nullhypotesene våre²⁵.

²⁵ Se vedlegg Statistikk-rapport.ods for mer informasjon.

5.2 Analyse av kvalitative data

Kvalitative data kan enten være tekst basert eller et multimedia, og vil for eksempel være informasjon fra observasjoner og intervju notater, tilbakemeldingene i en spørreundersøkelse, eller video – og audio-materiale. For å studere kvalitativ data bruker man andre strategier enn med kvantitativ data. Noen av de viktigste karakteristikene er at analysen av innholdet (content analysis) involverer menneskelig tolkning. Fraværet av numerisk data og direkte målinger gjør kvalitativ data mer utsatt for bias i tolkingen eller manipulasjon fra tolkeren. For å kontrollere de manipulative innvirkningene bør man foreta en kodingenprosess der man bruker statistisk metode for å underbygge gyldigheten og påliteligheten til kodingen som er tolket.²⁶

Kvalitativ analyse består av tre stadier. Man starter vanligvis med innholdet i en kvalitativ data. Via analyse av materialet er målet å identifisere hovedmomentene i materialet. Deretter undersøker man disse punktene i dybden og finner de naturlige grunnene til hvorfor det er slik, og prøve å trekke linjer mellom disse. Tilslutt brukes kunnskapen vi har innhentet til å forstå helheten i innholdet, og å se sammenhengene i det.²⁷

Informasjonen vi har tilegnet oss gjennom observasjon, film og intervju fra brukbarhetstesten, er grunnlaget for den kvalitative dataen. Vi har gjennomgått all informasjonen og i hovedsak trukket ut de viktigste tendensene som vi har sett at har gått igjen i dataene. I følge våre hypoteser skal vi finne ut om barna er mer interessert i de fysiske boomwhackers eller i den elektroniske (applikasjonen), vi vil også se om kjønn og alder kan påvirke barnas adferd. Vi har analysert hver gruppe først, og deretter sammenliknet resultatene mellom gruppene, for så å komme tilbake til det overordnede målet om bruk igjen.

5.2.1 Hovedtrekk

Dette var likhetstrekk som gikk igjen i forskningsmetodene:

1.klasse

I den første gruppen som var fra 1 klasse, var det 5 barn som blir tatt inn til studiet. 4 gutter og ei jente. De var interessert i både de fysiske og elektroniske boomwhackerne. Vi kunne se en liten forskjell på hva de likte etter hvilket kjønn de hadde. Noen elementer som kunne vise oss dette resultatet er som følger:

- Var entusiastiske og trykket vilt på applikasjonen. Startet med en gang de satt seg. Selv om ikke hadde gitt de noen oppgaver ennå.
- For mange deltager i forhold til telefoner og applikasjoner, derfor måtte flere dele.
- litt uenigheter og dytting hos guttene.

²⁶ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 282.

²⁷ Lazar, Feng og Hochheiser. *Research Methods In Human- Computer Interaction*, 282.

- Elektronisk: tableten var mye mer interessant enn telefonene.
- Boomwhackers: først var farge viktig, men etter hvert brydde ikke guttene seg om farge. Mens jenta hadde to rosa rør under hele testingen.
- jenta sitter stille ved pulten under hele studiet.
- Under testing av de fysiske var barna mer ute av kontroll.
- Guttene likte boomwhackerne litt mer enn den elektroniske applikasjonen. Mens jenta ble litt mer ukomfortabel når guttene begynner å løpe rundt med de fysiske boomwhackers.
- Etter testing det var vanskelig å stoppe barna, endte opp med å løp etter dem for å få boomwhackerer tilbake og kontrollerer dem.
- Barna klarte ikke å sitte lenge stille, ble fort rastløse.
- Barna var mer rolige under den elektroniske testingen.

2. klasse

I 2 klasse var det 4 barn i alt. Tre gutter og en jente. Her var det enda større forskjeller på jenter og gutter, fysisk og elektronisk. Noen elementer som kan vise oss disse resultatene er som følger:

- Veldig entusiastiske til alle oppgaver de får.
- Alle skjønnte hvordan de forskjellige knappene skulle brukes.
- Organiserer en lek hvor alle spiller samtidig, varer ikke lenge.
- Guttene var ivrige og rastløse.
- Urolige, men fortsatt litt mindre urolig enn den første gruppen.
- Synes det er morsomt å ta opp lyd og spille av musikken etterpå.
- Mindre kranglig på de som deler tableten enn det var med først gruppe.
- Alle synes lyden på mobilen er lav.
- Boomwhackers: Guttene mister tålmodigheteten etter hvert og begynner å løpe rundt og slå vilt rundt seg.
- Jenta er mye mer reservert og rolig enn guttene. Veldig ivrig etter å vise oss at hun kan spille en sang.
- Jenta synes boomwhackers ble for mye støy, og at det var vanskelig å være en person når man skal spille en sang.

3. klasse

I 3 klasse var de tre barn i alt, hvor alle var jenter. Både på den elektronisk og de fysiske boomwhackers kunne du høre at deltagerne hadde en rytme, hvor de prøvde å spille i takt på hverandre. Dette gjorde at det ble mye mer organisert. Noen elementer som kan vise oss disse resultatene er som følger:

- Alle får hver sin applikasjon. Slipper å dele eller vente.
- Spiller med en gang forskjellige toner på applikasjonen.

- Mye mer rytme på det de spiller enn tidligere. Virker som de prøver å spille en sang.
- Misforstår hensikten med rørene først. Synger i dem, i stedet for slå.
- Slår hverandres rør. Ligner litt på sånn guttene oppførte seg.
- Ikke opptatt av spesielle farger. Skjønner mer at fargene, bokstavene (oktavene), lengdene er forskjellige noter, er derfor interessert å prøve de forskjellige rørene.
- Prøver å få til et spill med rørene, men vanskelig da de ikke forstår instruksjonene. Gir derfor opp.

5.2.2 Sammenlikning av hovedtrekkene med hypotesene

Tekst og multimedia er gode kilder for informasjon og kan utdype forståelsen av adferd og interaksjon. Kvantitative målinger kan ikke hjelpe med å finne de underliggende utfordringene som brukeren opplever når han/hun bruker programmet. F.eks. hva de liker og ikke og forslag til forbedringer.²⁸ Vi har valgt ut informasjon som er relevant i forhold til det teoretiske ramverket vårt, altså hypotesene våres. Informasjonen vi har fremhevet baserer seg derfor på alder, kjønn og forskjellen på interaksjonen på elektronisk/fysisk boomwhacker.

Elektronisk/fysisk

Det var veldig jevnt hva de likte best av den elektroniske og den fysiske. Av de elektronisk produktene likte de fleste tableten best. Den var mer interessant enn telefonene, pga. at skjermen var større og lyden bedre. Det var vanskelig å dele på produktene når de var mange, jo færre og jo eldre de var dess bedre gikk det å dele. Delingen gjorde at det ble litt uenigheter og dytting hos guttene spesielt. På boomwhackers var farge viktig, men etter hvert som de lekte med de brydde ikke guttene seg om farge.

Barna var mer rolige under den elektroniske testingen enn den fysiske. Under den fysiske utgaven var barna mer ute av kontroll ved at de løp rundt og slo hverandre og rørene. Det er kanskje fordi den ekte versjonen er et fysisk spill, der man bruker mer av kroppen. Dette førte til at guttene reiste seg fort, og var mer aktive under denne delen av testingen. Applikasjonen er mindre i form og egner seg bedre til å spille i sittende tilstand. Guttene likte boomwhackerne litt mer enn den elektroniske applikasjonen. Mens jenta ble litt mer ukomfortabel når guttene begynner å løpe rundt med de fysiske boomwhackers.

Hensikten med de fysiske boomwhackerne blir i større grad misforstått enn den elektroniske. Flere av barna bruker rørene til å synge eller kikke gjennom i stedet for slå. I den andre gruppe som var fra 2 klasse, det er litt vanskelig å finne ut om forskjeller mellom barnas interesse i den elektroniske og den fysiske boomwhackers.

²⁸ 286

Kjønn

Vi oppdaget store forskjeller mellom jenter og gutter i forhold til den elektroniske applikasjonen og den fysiske boomwhackeren. Jentene var stille og sjenerte i de første gruppene. De synes boomwhackeren ble for mye støy og bråk. De var også veldig opptatt av fargene på de fysiske rørene. Guttene i første og andre klasse var veldig utolige. Når de skulle spille på boomwhackeren begynte de å slå hverandre med rørene og prøvde å ta hverandre. De brukte også rørene som sverd og begynte å fekte med disse. Jente i tredjeklasse var ikke så sjenerte. De brukte også rørene som sverd, men mer kontrollert enn guttene. De var ikke så opptatt av farger. Jentene var opptatt av rytme og prøvde å få til melodier.

Alder

På melodi og noter kan man se en betydelig forskjell på aldersgrupper. 1 klassingene slo bare rundt seg uten å tenke på melodi eller rytme. 2 klassingene hadde litt mer rytme, og man kunne høre litt melodier når de slo. 3 klassingene tenkte veldig på rytme, og slo mer i takt med hverandre. Dess eldre de er jo mer rytme er det i det de spiller. Spillet blir mer organisert. Prøver å spille sanger enn bare å trykke vilt på knappene. Mindre opptatt av spesielle farger i høyere alder. Skjønner mer at fargene og lengden representerer forskjellige bokstaver (oktaver), og er derfor interessert å prøve de forskjellige rørene. Prøver å få til et spill med rørene, men det blir litt vanskelig da de ikke kan noter og ikke vet hvilket rør man begynner med osv.

Det ser ut at jenter er mer interessert i applikasjonen en gutter. Særlig i vår testing det var en jente som kunne spille piano og hun var kjent med toner og musikk mer en de andre. Hun likte applikasjonen fordi det var letter å spille musikk med den og det var letter for henne for å gjenkjenne toner på applikasjoner.

Gutter spilt med fysiske boomwhackers på forskjellige måter, lage lyder, løpe rund, slå hverander, de brukt dem isteden for mikrofon osv. Men de likte godt å spill av lyder eller musikk som de opprettet selv ved bruk av applikasjonen.

Derfor vi fikk ikke til å se forskjell mellom gutter og jenter i 3.klasse men på den annen side vi kunne observer at hvordan spiller en grupp av jenter med applikasjonen.

6. Resultater av analysen

6.1 Resultatene fra dataen knyttet opp mot bruk

6.1.1 Kvalitative data:

Vi kan se at det i de kvalitative dataene er liten forskjell mellom den elektroniske og fysiske boomwhackerne. Det var veldig spredt hva barna foretrakk å leke med.

Ut i fra de kvalitative dataene var det en forskjell mellom jentene og guttene. Det oppdaget vi spesielt under testen av de fysiske boomwackerne. Guttene brukte boomwhackers feil, ved at det ble et slagvåpen og et leketøy, i stedet for et instrument. Jentene derimot, stabet de over hverandre og begynte å synge i de. Dette er kanskje fordi de ikke visste hva de skulle brukes til i utgangspunktet.

De kvalitative dataene indikerer også en forskjell i bruk og alder. De eldste oppførte seg anderledes med produktet, ved å bruke det mer som et instrument enn de yngste. De prøvde i større grad å spille ulike toner og få en fast rytme. En feil i eksperimentet er at det bare var jenter i 3.klasse gruppen. Det blir vanskelig å sammenligne alder når bare det ene kjønn var representert. Dette fordi det er forskjell på hvordan jenter og gutter interagerer med applikasjonen og boomwhackeren. Jentene virket i alle gruppene roligere og mer behersket i interaksjonen med produktene enn guttene gjorde.

6.1.2 Kvantitative data:

Resultatene av de kvantitative testene sier at det ikke er noen signifikant forskjell på hvor lenge de benytter seg av applikasjonen og boomwhackers før de går lei. Vi kan derfor ikke fra et statistisk perspektiv forkaste nullhypotesen: Det er ingen forskjell på hvor lenge de leker med applikasjonen og de fysiske boomwhackers. Her kan det være type II feil fordi vi har ganske få deltagere i testen, og dette kan føre til feil i de statistiske testene vi har gjort.

De kvantitative testene sier også at det ikke er noen signifikant forskjell på hvor lenge gutter og jenter benytter seg av applikasjonen før de blir lei. Vi kan derfor ikke fra et statistisk perspektiv forkaste nullhypotesen: Det er ingen forskjell på hvor lang tid jenter og gutter leker med applikasjonen. Også her kan det være type II feil fordi også her har vi ganske få deltagere i testen, og dette kan føre til feilaktig statistikk.

På den tredje og siste hypotesen gjelder det samme som for de to første hypotesene. Og også her kan vi få en type II feil fordi vi ikke hadde nok deltagere. Det er flere grunner til at vi hadde så få deltagere, tidsbegrensninger og økonomiskebegrensninger var kanskje mest gjeldene på akkurat dette. En annen feil med testingen vår kan være at vi ikke hadde nok tid med hver gruppe, så vi måtte avbryte og sette tiden etter 7 minutter uavhengig av alle barna hadde gått lei eller ikke. Tiden som ble satt opp på de barna som enda ikke hadde gått lei ble da satt på 7 min, som var tiden vi måtte avbryte på. Dette er nok den største kilden til feil ved testene våre.

6.1.3 kvalitativ vs kvantitativ:

Det er ikke helt konsensus imellom det kvalitative perspektivet og det kvantitative perspektivet. Fra et statistisk ståsted kan man ikke forkaste noen av hypotesene med de testene vi har gjort, men sett ifra ett kvalitativt perspektiv så må man forkaste alle nullhypotesene. Hvilke av disse to perspektivene som har konkludert rett kan vi ikke svare på slik det er nå, og vi anbefaler at man utfører videre testing.

6.2 Bias og reliability av resultatene

Under usability testingen hadde vi for mange deltagere i første gruppe. Dette gjorde at vi ikke fikk så mye kontroll over testingen som vi gjerne ville ha hatt. Det var vanskelig å følge med på hva hver enkelt gjorde, samtidig var det mye krancling om de tre tabletene vi hadde. Dette kan ha innvirket på at målingene ble litt vanskelige og ukorrekte.

Før undersøkelsen skrev vi at en mulig bias kunne være at barna kunne bli for sjenerte hvis vi måtte gjennomføre brukertesten alene. Etter å ha gjennomført testen ser vi også at det å ha for mange deltagere inne av gangen igjen kan føre til en bias. Deltagerne i denne aldersgruppen er lett påvirkelig og kan fort forandre meninger. Hvis et barn sa at de likte applikasjonen kunne de andre barna også forandre mening.

En annen bias som kan tas i betraktning er om barna svarte de "riktige" svarene for å tilfredstille oss. Vi kan i denne testen ses på som en autoritet, og derfor være noen de ønsker å behage. De kan derfor ha fortalt oss det de trodde vi ville høre under intervjuene.

Gruppene vi utførte testen på hadde en ujevn fordeling av gutter og jenter. Dette var noe vi hadde lite kontroll over da Katie valgte ut deltagere for oss. Valgene av deltagere kan Katie f.eks. ha tatt på bakgrunn av hvilke foreldre som hadde signert samtykkeskjema. Dette førte til å at vi muligens kan ha fått noen biaser under testing og derav en redusert reliability. I den første gruppe var det for eksempel en jente og fire gutter. Denne typen situasjon kan påvirke resultatene. Vi så at jenta var mer reservert enn jentene i den siste gruppen. Jenta kan ha blitt påvirket av denne omstendigheten og guttenes utagerende adferd. Dette kan redusere påliteligheten av resultatet mellom gutter og jenter.

7. Evaluering av prosjektet

Vi brukte et mini-etnografi til å finne ut hva som kunne være interessante hypoteser. Her fant vi ut av oppførsel og at hva dem likte gikk mye på alder og kjønn. Det var også interessant å se forskjellen mellom applikasjonen og de ekte boomwhackerne. Når vi hadde utformet hypotesene brukte vi experimentell design som rammeverk til å designe

eksperimentet. Vi gjennomførte en usability testing, noen observasjoner og noen intervjuer. Da fikk vi utrolig mye data. Vi analyserte den kvantitative dataen med statistikk. Her fant vi ingen signifikante forskjeller på tiden deltagerne brukte. Alle hypotesene kan derfor ikke forkastes. Da vi analyserte den kvalitative data, ble det heller ingen stor forskjell på hypotesene. Vi kunne allikevel se en forskjell på alle hypotesene, som ikke kom tydelig nok frem da vi analyserte. Flere undersøkelser kan indikere at vi burde forkaste hypotesene.

Noen av forskjellene vi fant ut i intervjuet, tidene, observasjonene og barnas kroppsspråk viser at de er interessert i både fysiske og elektroniske, men de likte forskjellige elementer i hver. For eksempel syntes de at det var enklere å spille musikk med den elektroniske, men samtidig likte de å lage lyder med de fysiske og slå rundt seg med de. De likte fargene og tonene i både den fysiske og elektroniske utgaven, og i tillegg syntes de det var gøy å spille av musikk i applikasjonen.

Hvis vi hadde hatt lengre tid, ville vi gjennomført eksperimentet på flere deltagere. Da hadde vi fått en bredere og større undersøkelse, og kanskje et annet resultat. Manglende erfaring fra vår side har nok noe å si både på analysen av de kvalitative dataene og de kvantitative dataene. Både på effektivitet og mest sannsynlig med tanke på forskjellige biaser. Det er en viss mulighet for at vi ikke klarte å være fullstendig objektive når de behandlet de kvalitative dataene. Det hadde nok også hatt noe å si på gjennomføringen av observasjonen, intervjuet, og brukbarhetstestene. Vi hadde muligens klart å observere flere ting, ting som vi kanskje så på som urelevante, kan ha vært relevante. Det er også en mulighet for at vi hadde klart å se flere sammenhenger i bruken om vi hadde vært mer erfarne. På intervjuet ville det nok hatt nokså stor påvirkning om vi hadde vært erfarne. Vi hadde nok vært mer realistiske med tanke på hvilken grad et barn er i stand til å svare. Og kanskje også utformet bedre spørsmål.

Ellers er vi fornøyde med prototypen vi har utviklet, det ser ut til at det er lett å forstå hvordan den fungerer og hva de forskjellige knappene gjør. Vi oppdaget ingen feil eller "bug's" med applikasjonen under testingen.

8. Kilder:

Lazar, Jonathan, Jinjuan Heidi Feng og Harry Hochheiser. Research Methods In Human- Computer Interaction. United Kingdom: John Wiley & Sons Ltd, 2010.

Yvonne Rogers, Helen Sharp and Jeniffer Preece: Interaction Design: beyond human-computer interaction, 2. utgave. Wiley, 2009.

Wikipedia. Hypotese. Wikipedia, 24 august 2011, <http://no.wikipedia.org/wiki/Hypotese> (23 oktober 2011).

Wikipedia. Etnografi. Wikipedia, 15 mai 2011, <http://no.wikipedia.org/wiki/Etnografi> (5 november 2011).