



Evaluering av tilgjengelighetstest- metoder

Universell utforming

Rapportnr

1038

Forfattere

Aleksander Bai, Heidi Mork, Viktoria Stray, Nikolai Sverdrup, Till
Halbach, Kristin Skeide Fuglerud

Dato

31.03.2018

ISBN-13

978-82-539-0548-8

Forfatterne

Aleksander Bai, Heidi Mork, Viktoria Stray, Nikolai Sverdrup, Till Halbach, Kristin Skeide Fuglerud

Norsk Regnesentral

Norsk Regnesentral (NR) er en privat, uavhengig stiftelse som utfører oppdragsforskning for bedrifter og det offentlige i det norske og internasjonale markedet. NR ble etablert i 1952 og har kontorer i Kristen Nygaards hus ved Universitetet i Oslo. NR er et av Europas største miljøer innen anvendt statistisk-matematisk modellering og har et senter for forskningsdrevet innovasjon, Statistics for Innovation – (sfi)², med finansiering fra Norges forskningsråd. Det jobbes med et bredt spekter av problemstillinger, for eksempel finansiell risiko, jordobservasjon, estimering av fiskebestander og beskrivelse av geologien i petroleumsreservoarer. NR er ledende i Norge innen utvalgte deler av informasjons- og kommunikasjonsteknologi. Innen IKT-området har NR innsatsområdene e-inkludering, informasjonssikkerhet og smarte informasjonssystemer. NRs visjon er forskningsresultater som brukes og synes.

Samarbeidspartnere

Kantega AS

Institutt for Informatikk, Universitet i Oslo

Tittel	Evaluering av tilgjengelighetstest-metoder
Forfattere	Aleksander Bai, Heidi Mork, Viktoria Stray, Nikolai Sverdrup, Till Halbach, Kristin Skeide Fuglerud
Kvalitetssikring	Kristin Skeide Fuglerud
Dato	31.03.2018
År	2018
ISBN	978-82-539-0548-8
Publikasjonsnummer	1038

Sammendrag

Denne rapporten gir et overblikk over prosjektet som ble utført samt resultatene som fremkom. Totalt fikk vi evaluert 6 ulike metoder for å teste tilgjengelighet med over 50 ulike personer.

Emneord	Universell utforming
Målgruppe	Alle
Tilgjengelighet	Åpen
Prosjektnummer	320577
Satsningsfelt	Universell utforming
Antall sider	43

Forord

En stor takk til alle bidragsytere fra prosjektpartnere, til alle som deltok i evaluering av metodene, og ikke minst til UnIKT-programmet for finansiell støtte til prosjektet som ble gjennomført i 2017. UnIKT-programmet i Bufdir skal styrke arbeidet med digital deltakelse. UnIKT skal medvirke til at flere digitale satsninger kan bli til nytte for hele befolkningen. UnIKT skal stimulere til gode, inkluderende digitale prosjekt rettet mot allmenheten.

Innhold

1	Bakgrunn	9
2	Gjennomføring	10
2.1	Valg av metoder.....	10
2.1.1	Metode 1: WCAG-gjennomgang.....	11
2.1.2	Metode 2: SiteImprove	11
2.1.3	Metode 3: Cambridge Simulation Glasses.....	12
2.1.4	Metode 4: Skjermleser	13
2.1.5	Metode 5: Dysleksi-simulering.....	14
2.1.6	Metode 6: Personas	15
2.2	Praktisk gjennomføring	16
3	Evalueringer	18
3.1	Metode 1: WCAG-gjennomgang	18
3.2	Metode 2: SiteImprove	19
3.3	Metode 3: Cambridge Simulation Glasses	20
3.4	Metode 4: Skjermleser	21
3.5	Metode 5: Dysleksi-simulering	23
3.6	Metode 6: Personas.....	24
3.7	Alle metoder.....	25
4	Integrasjon	31
4.1	SiteImprove.....	31
4.2	Cambridge Simulation Glasses	32
4.3	Skjermleser.....	34
4.4	Dysleksi-simulering	35
4.5	Personas	36
5	Oppsummering	38
5.1	Designfase.....	38
5.2	Utviklingsfase.....	38
5.3	Testfase.....	39

5.4	Vedlikeholdsfase.....	39
5.5	Brukertesting.....	39
5.6	Organisasjon.....	40
6	Vedlegg A.....	41
6.1	USE-spørsmål	41
6.2	Scenario-oppgaver.....	42
6.2.1	Oppgave 1: Lavpriskalenderen.....	42
6.2.2	Oppgave 2: Er flyet i rute.....	42
6.2.3	Oppgave 3: Bli medlem i EuroBonus.....	42
6.2.4	Oppgave 4: Reise med katt	43
6.3	Intervjuguide	43

Figuroversikt

Figur 1. Skjerm bilde av en nettside og SiteImprove-utvidelsen i Chrome-nettleser. © SiteImprove	12
Figur 2. En ung kvinne som tar på seg flere briller for å simulere ulik grad av visuell funksjonsnedsettelse. © Cambridge Simulation Glasses	13
Figur 3. Skjerm bilde av nettsiden sas.no vist av en nettleser i kombinasjon med NVDA-skjermleser, fokus-utvidelse og 'speech viewer'-utvidelse.....	14
Figur 4. Eksempel på hvordan tekst kan bli seende ut når dysleksi-utvidelsen gjør om en tekst i et nettleservindu.	15
Figur 5. Bilde av imaginære «Linn» som kan brukes under personas-testing.....	16
Figur 6. Radar-diagram for WCAG-gjennomgang evaluering	19
Figur 7. Radar-diagram for SiteImprove evaluering.....	20
Figur 8. Radar diagram for Cambridge Simulation Glasses	21
Figur 9. Radar-diagram for skjermleser.....	23
Figur 10. Radar-diagram for dysleksi-simulering	24
Figur 11. Radar-diagram for personas	25
Figur 12. Radar-diagram for alle metodene.....	26
Figur 13. Evalueringsresultater for alle metoder i kategorien usefulness.....	27
Figur 14. Evalueringsresultater for alle metodene i kategorien ease of learning.....	28
Figur 15. Evalueringsresultater for alle metodene i kategorien ease of use.....	29
Figur 16. Satisfaction evaluering for alle metodene.....	29
Figur 17. Totalevaluering for alle metodene	30
Figur 18. Evalueringsresultater for SiteImprove for ulike roller.	32
Figur 19. Evalueringsresultater for Cambridge Glasses for ulike roller.	33
Figur 20. Evalueringsresultater for skjermleser for ulike roller.	35
Figur 21. Evalueringsresultater for dysleksi-simulering for ulike roller.	36

Tabelloversikt

Tabell 1. Oversikt over evaluerte metoder.....	10
Tabell 2. Evalueringsresultater for WCAG-gjennomgang	19
Tabell 3. Evalueringsresultater for SiteImprove.....	20
Tabell 4. Evalueringsresultater for Cambridge Simulation Glasses	21
Tabell 5. Evalueringsresultater for NVDA-skjermleser.....	22
Tabell 6. Evalueringsresultater for VoiceOver-skjermleser	22
Tabell 7. Evalueringsresultater for begge skjermlesere.....	22
Tabell 8. Evalueringsresultater for dysleksi-utvidelse	23
Tabell 9. Evalueringsresultater for personas	24

1 Bakgrunn

En av utfordringene innenfor IKT er å ta høyde for universell utforming (uu) under utvikling av nye tjenester og ny programvare. I Norge er dette påkrevd ifølge Forskriften for universell utforming av IKT. Studier viser at brukertesting ofte blir nedprioritert i systemutviklingsprosjekter på grunn av kostnadsbildet, tidsforbruk og/eller mangel på kompetanse. Spesielt brukertesting rettet mot grupper med nedsatt funksjonsevne blir ikke prioritert siden dette krever kompetanse og en del tidsressurser hva gjelder rekruttering.

IKT-miljøene i Norge har derimot fått større fokus på brukersentrerte prosesser de siste årene, og de fleste bedrifter har interne testere og designere som ivaretar generelle brukervennlighetsprinsipper ved programvareutvikling. Det er derimot relativt få utviklere og testere som har kompetanse på universell utforming, og som utfører brukertesting eller egen tilgjengelighetstesting ved programvareutvikling.

Det er tidkrevende å sette seg inn i standarder og verktøy som finnes, og det kan også være uoversiktlig og vanskelig å forstå de tekniske anbefalingene og å navigere i litteraturen. Dette gjelder spesielt for utviklere som har ekspertise på programvareutvikling, men også for testere som har mer fokus på brukervennlighet, men ikke nødvendigvis universell utforming eller tilgjengelighet.

I dette prosjektet har vi sett nærmere på metoder og verktøy for å teste tilgjengelighet, og vi har fått metodene evaluert av personer med ulike roller i smidige utviklingsteam. I denne rapporten presenterer vi en oversikt og vurdering av hver enkelt metode med tanke på hvor lett den er å lære, hvor lett den er å bruke, hvor nyttig og effektiv den oppleves, og hvor interessant den er å bruke i seg selv.

Dette tiltaket var et samarbeid mellom Norsk Regnesentral, IT-konsulentselskapet Kantega og gruppen Programmering & Software Engineering (PSE) ved Institutt for Informatikk, Universitet i Oslo.

2 Gjennomføring

Evalueringen av de ulike metodene ble gjennomført hos 7 ulike utviklingsprosjekter i 6 ulike bedrifter med totalt 53 ulike deltakere. Hver deltaker evaluerte mellom to og fire metoder, avhengig av forkunnskaper, ønske og tid.

2.1 Valg av metoder

Vi valgte ut seks universell utforming (uu)-metoder som dekket ulike grupper av funksjonsnedsettelse. Personer med synsnedsettelse støter ofte på store problemer fordi it-løsninger er i stor grad visuelle. Personer med kognitive funksjonsnedsettelse kan også ha store utfordringer fordi it-løsningene inneholder mye informasjon, og det kan også være vanskeligere å teste for. Med det i bakhodet ønsket vi å evaluere metoder som tester visuelle og kognitive aspekter, og metoder som dekker et bredt spekter av funksjonsnedsettelse. Det er også andre brukergrupper som kunne vært inkludert, som f.eks. personer med motoriske- eller hørselsnedsettelse, og spesifikke kognitive funksjonsnedsettelse, men pga. begrenset tid valgte vi disse tre områdene.

Det er mange metoder å velge imellom, og de metodene vi endte opp med ble valgt basert på erfaring fra tidligere prosjekter, og to piloter vi gjennomførte. I pilotene testet vi flere metoder, som f.eks. Wave-verktøyet utviklet av WebAim¹ (automatisk sjekker), flere ulike personas, ulike guidelines etc. Til slutt endte vi opp med de 6 metodene som er listet i tabell 1.

#	Navn	Type	Funksjonsnedsettelse
1	WCAG-gjennomgang	Sjekkliste	Flere
2	SiteImprove	Automatisk sjekker	Flere
3	Cambridge Simulation Glasses	Simulering	Visuell
4	Skjermleser	Assistive Technology / Simulering	Visuell
5	Dysleksi-simulering	Simulering	Kognitiv
6	Personas	Heuristisk	Kognitiv

Tabell 1. Oversikt over evaluerte metoder.

¹ <http://wave.webaim.org/>

Både skjermleser og Cambridge Simulation Glasses dekker tilgjengelighetstesting av det visuelle, mens dysleksi-simulering og personas dekker det kognitive. Både WCAG og SiteImprove dekker flere typer funksjonsnedsettelse.

2.1.1 Metode 1: WCAG-gjennomgang

Vi valgte å inkludere en WCAG-gjennomgang siden det er de facto-standard når det gjelder å teste tilgjengelighet på web.

For å utføre en WCAG-gjennomgang lagde vi et Excel-ark med alle 61 kriterier fra WCAG 2.0². Vi sendte Excel-arket ut i etterkant av evalueringene og minnet deltakerne på det etter 2 dager, og etter 2 uker. For hvert kriterium kunne de velge mellom NA (vet ikke eller ikke anvendbar), Pass eller Fail. Vi brukte den norske oversettelsen fra uu.difi.no og oppga linker til både uu.difi.no og W3C sin definisjon av WCAG 2.0. Vi valgte ut en nettside som alle deltakerne brukte en offentlig side som utgangspunkt. De fikk maks 10 minutter på å evaluere så mange punkter de rakk.

2.1.2 Metode 2: SiteImprove

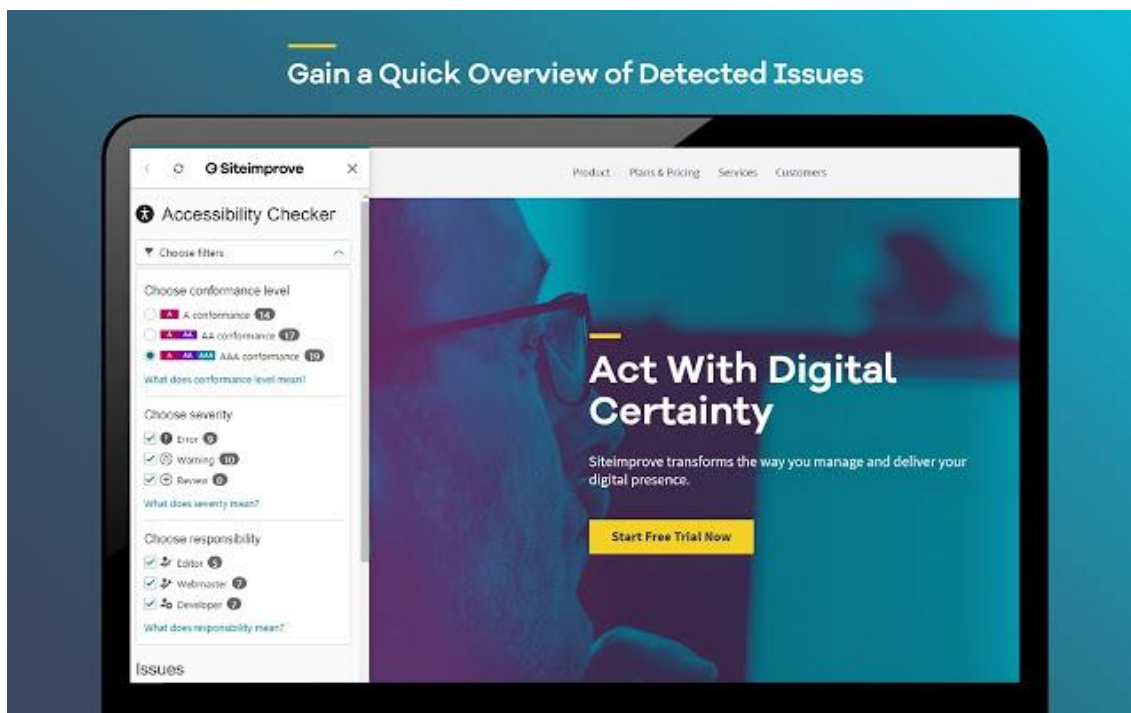
Vi vurderte flere automatiske sjekkere, men valgte SiteImprove³ istedenfor for eksempel Wave⁴ da vi fikk bedre tilbakemelding på grensesnittet og strukturen til SiteImprove i en pilot vi utførte.

For SiteImprove ga vi instruksjoner om å benytte Chrome sin utvidelse og hvordan de kunne installere denne. I tillegg ble deltakerne informert om at automatiske sjekkere som regel tester om webløsninger følger WCAG 2.0 standarden og andre standarder. Vi forklarte også at automatiske sjekkere har begrensninger, som f.eks. at de kan sjekke om det finnes alternativ tekst til bilder, men de kan ikke sjekke om beskrivelsen faktisk gir mening for de som bruker skjermlesere.

² <https://www.w3.org/TR/WCAG20/>

³ <https://chrome.google.com/webstore/detail/siteimprove-accessibility/efcfolpjhicnikpmhnmphjhpicljc?hl>

⁴ <https://wave.webaim.org/extension/>



Figur 1. Skjermbilde av en nettside og Siteimprove-utvidelsen i Chrome-nettleser.
© Siteimprove

2.1.3 Metode 3: Cambridge Simulation Glasses

Brillene simulerer ikke en spesiell øyetilstand/sykdom, men generelle problemer med fokus, nedsatt sensitivitet og tåke/skygge på øyet. Disse generelle problemene forekommer typisk med alder og med mange øyetilstander/sykdommer, eller hvis man ikke bruker riktig tilpassete briller. Deretter forklarte vi at ved å simulere generelle synsnedsettelse kan man fort oppdage problemer med kontrast og tekststørrelser. Grunnen til at vi valgte Cambridge Simulation Glasses⁵ var fordi flere i prosjektet hadde positive erfaringer med brillene. I tillegg ønsket vi å ha med en metode som ikke var digital, og som kan føles mer på kroppen.



⁵ <http://www.inclusivedesigntoolkit.com/csg/csg.html>

Figur 2. En ung kvinne som tar på seg flere briller for å simulere ulik grad av visuell funksjonsnedsettelse.
© Cambridge Simulation Glasses

Ved å sette flere briller oppå hverandre får man dårligere og dårligere synsopplevelse. Før vi begynte kalibrerte vi hver deltaker for å finne ut hvor mange briller han eller hun trengte. Vi tok utgangspunkt i 1% testen som sier at hvis løsningen fungerer med et gitt antall briller, så kan 99% av befolkningen bruke løsningen⁶. I snitt hadde de fleste 2 briller, men noen brukte også 3 briller.

Cambridge Simulation Glasses ble bestilt fra deres nettsider, og totalt bestilte vi 250 briller.

2.1.4 Metode 4: Skjermleser

Personer som er blinde eller har sterk synsnedsettelse kan ikke benytte synet til å lese innholdet på en nettside, men bruker i stedet programvare som parser innholdet på siden. Innholdet blir så lest opp eller vist på en blindeskriftslist etterhvert som brukeren navigerer med tab eller spesialkommandoer på siden. En slik programvare kalles for skjermleser. Vi benyttet NVDA⁷ for de som hadde Windows maskiner og VoiceOver⁸ for de som hadde MAC maskiner.

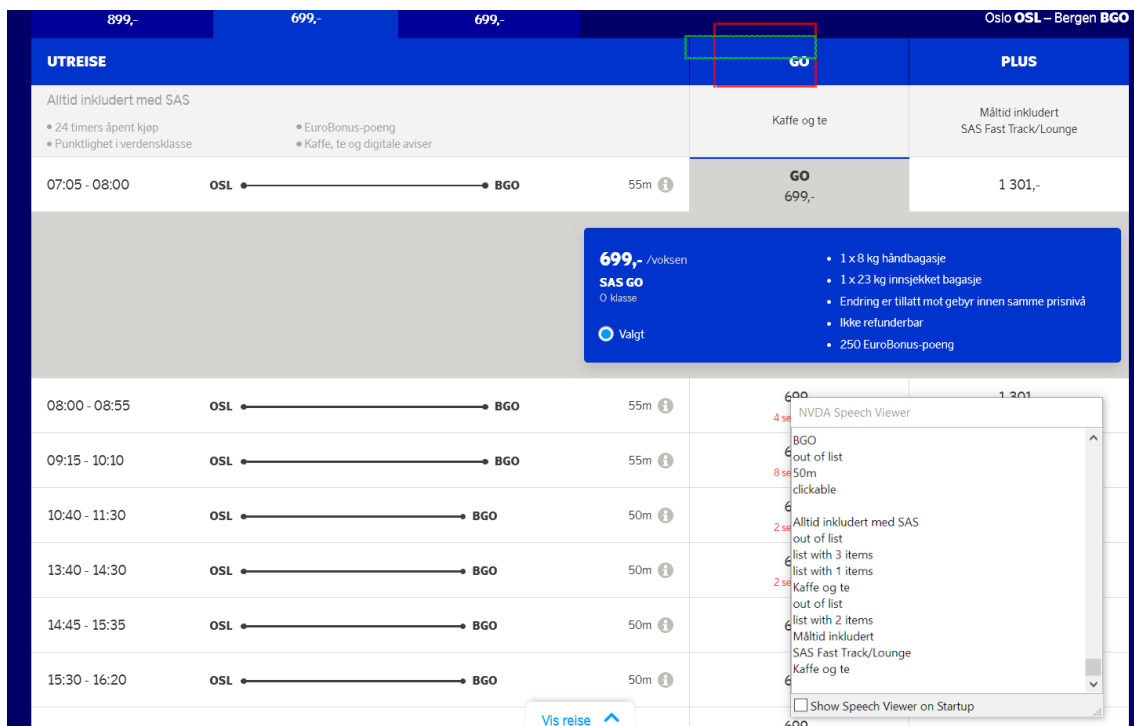
I forkant av utprøvingen sendte vi ut en epost med link til noen ressurser og instruksjoner som vi ba deltakerne om å lese igjennom i forkant. Vi sendte forskjellig epost ut ifra om de benyttet Windows eller Mac. F.eks. så ba vi dem med Windows om å installere NVDA og en utvidelse kalt 'focus highlight'. I tillegg ba vi dem se en kort video⁹ på 5 minutter som ga tips om hvordan man kan bruke en skjermleser som ikke-blind person. Blant annet at man burde slå på 'speech viewer' istedenfor å høre på hva skjermleseren sier.

⁶ <http://www.inclusivedesign toolkit.com/csg/csg.html#p20>

⁷ <https://www.nvaccess.org/>

⁸ <https://www.apple.com/lae/accessibility/mac/vision/>

⁹ <https://www.youtube.com/watch?v=Vx1vSd5uYS8>



Figur 3. Skjerm bilde av nettsiden sas.no vist av en nettleser i kombinasjon med NVDA-skjermleser, fokus-utvidelse og 'speech viewer'-utvidelse.

2.1.5 Metode 5: Dysleksi-simulering

For å simulere opplevelsen av dysleksi benyttet vi en Chrome-utvidelse kalt 'Dyslexia simulation'¹⁰. Utvidelsen er utviklet i samarbeid med dyslektikere og gir et inntrykk av konsekvensene mange dyslektikere opplever ved å lese.

Vi forklarte deltakerne at utvidelsen på ingen måte er en god simulering av dysleksi, men at den gir en liten pekepinn om hva personer med dysleksi kan oppleve når de besøker nettsider. Scriptet flytter rundt på bokstaver, slik at det krever betydelig konsentrasjon for å lese innholdet på en nettside. Det er viktig å påpeke at det ikke nødvendigvis er slik dyslektikere opplever det, men at ikke-dyslektikere får oppleve at tekst blir vanskelig å lese. Det blir vanskelig å lese ustrukturert informasjon, lange ord blir ekstra vanskelige og det krever mye mer konsentrasjon å lese. Det er denne konsekvensen som gir en slags simulering av dysleksi. I tillegg ga vi litt bakgrunnsinformasjon om hvor mange som har dysleksi i Norge og litt mer om hva det innebærer.

¹⁰ <https://chrome.google.com/webstore/detail/dyslexia-simulation/cnobhbaajmbcbfdiakhlkciemjgac?hl=en>

Dylxesia is ctaiieczrhred by dclifiutty wih Inrieang to read fluently and with acuacrtc
cnproeishmeon desptie nrmoal ienglelitnce. This idnleucus dlitfcuify wih pgloocnihoal
aweenarss, pcaoholgoinl deodcing, prcisneosg seped, otrphihragoc cdniog, atdioury
short-term mormey, laagunge silks/varbel chonoeoisrpn, and/or rpiad nmnaig.

Dvpeneltoamel rdeinag didosrer (DRD) is the msot comomn lnainerg dbistaiiy.
Dyesxlia is the msot rzgoicneed of redniag dседorris, howeevr not all rniadeg
dierodsrs are lkeind to deslyixa.

Smoe see disxeyla as dsnticit form rednaig defiiiculfits rstelnuig form otehr cuseas,
such as a non-naulrogoicel deincicfey wih vision or hirenag, or poor or iquiteaande
redanig ictosunirtn. Three are tehre psopreod cviogntie suytpbes of dyilsxea
(audtoiry, vsauil and aoitntaetnl), agolulth ivdauidinl caess of diyelxsa are bteetr
exaiplined by sfceiipc uylidnreng noucrepysaliooghil dicteifs and co-ourrcncig
lanenig dbialteisiis (e.g. aotntetin-dfieict/htctiipyervay dsidreor, math dlsbtaiiy, etc.).
Augohlth it is crnsoeided to be a rvpeictee luggaane-bsead lrenanig diilibatsiy in the
rsaeroeh lirttreuae, dxielsya also aftecfs one's exprsisvøe lauggnae siklls.
Reserhcaers at MIT funod taht pepole wih dlseyixa etbihiexd ipeairmd vocie-
rgionicteon aiieltibs.

Figur 4. Eksempel på hvordan tekst kan bli seende ut når dysleksi-utvidelsen gjør om en tekst i et nettleservindu.

2.1.6 Metode 6: Personas

Utvikling og bruk av personas er en velkjent metode for å gjøre seg kjent med en brukergruppe. Innen IKT-området benyttes personas ofte for å representere brukere i målgruppen. Personas kan benyttes i forbindelse med kravspesifikasjon, utvikling, testing eller markedsmessige avgjørelser for IKT-baserte produkter eller tjenester.

Vi ga deltakerne instruksjoner om personas som metode, som at kartlegging av behov er avgjørende når man utformer personas for å gjøre brukergrensesnittet best mulig for brukere. Vi valgte å benytte en persona med kognitive utfordringer. Det er mulig å utvikle en persona med diagnosen dysleksi, men ved å fokusere på utfordringen «lese- og/eller skrivevansker» enn diagnose, favner man da flere. Det er f.eks. langt flere som har utfordringer med lesing og skriving enn de som har diagnosen dysleksi. På tilsvarende måte kan det være mer hensiktsmessig å beskrive at en persona sliter litt med å huske ting enn å bruke begrepet begynnende demens. Det kan være mange situasjoner og tilstander som gjør at man sliter med å huske ting. Mange kan identifisere seg med å være en «rotekopp», og det kan ha mange årsaker, f.eks at man er sliten eller stresset, eller at man har konsentrasjonsvansker på grunn av ADHD.



Figur 5. Bilde av imaginære «Linn» som kan brukes under personas-testing¹¹.

I evalueringene brukte vi en persona utviklet av NR i samarbeid med Karde AS der personen¹² hadde karakteristika som typisk forbindes med ADHD.

2.2 Praktisk gjennomføring

I forkant av evalueringene sendte vi ut en epost hvor vi ba deltakerne om å svare på et bakgrunnsskjema. Bakgrunnsskjemaet og svarene er dokumentert detaljert i en annen rapport¹³. Vi sendte også ut en auto-generert nøkkel til deltakeren, slik at svarene kan linkes sammen, men uten at det er mulig å spore nøkkelen tilbake til en gitt deltaker.

Selve gjennomføringen foregikk på deltakeren sin egen bærbare PC eller Mac. Dette var fordi det gjorde logistikk enklere, men også fordi deltakeren er mer vant til å benytte sitt eget utstyr. Det var også hensiktsmessig at alle metodene enkelt kunne installeres.

Vi brukte en offentlig og godt kjent nettside som utgangspunkt, med fem ulike scenarier deltakerne kunne utføre mens de brukte en metode. Vi fulgte samme oppsett for alle deltakere. Hver deltaker fikk 10 minutter på å teste en metode, før de fikk beskjed om å fylle ut et evalueringsskjema for metoden. Selve evalueringen tok i gjennomsnitt 3 minutter.

Etter at en deltaker hadde evaluert to til tre metoder, gjennomførte vi et sluttintervju. Vi utarbeidet en intervjuguide som vi fulgte, og i gjennomsnitt tok sluttintervjuet rundt 10 minutter. Totalt brukte vi mellom 60 og 70 minutter per deltaker.

Vi gjennomførte to piloter internt hos Kantega i forkant for å teste gjennomføringen. Justeringer ble gjort basert på disse pilotene. Først og fremst ble tidskontrollen

¹¹ <http://iktforalle.no/tilgjengelige-nettsider/veileder-hele.html>

¹² <http://iktforalle.no/tilgjengelige-nettsider/material/persona-linn.pdf>

¹³ Evaluering av tilgjengelighetstest-metoder, NR 1036

strengere, og vi besluttet å bruke en offentlig tilgjengelig nettside som istedenfor interne sider. Det var primært for å ha samme utgangspunkt for alle deltakere, men også fordi vi opplevde at deltakerne stod friere til å diskutere når det ikke var interne sider eller løsninger.

3 Evalueringer

Vi benyttet USE som spørreskjema, som består av 30 ulike spørsmål (vedlegg A). USE er utviklet av Lund¹⁴ og kan deles inn i fire kategorier: Usefulness, Satisfaction, Ease of Use og Ease of Learning. Vi vurderte SUS¹⁵ og andre evalueringsskjema, men USE ble foretrukket siden den baserer seg på flere dimensjoner. Dette gir oss flere muligheter til å vurdere metodene opp mot hverandre innenfor ulike kategorier. Skalaen til USE er 1 til 7, hvor 1 er svært uenig og 7 er svært enig i påstandene. Vi valgte å beholde spørsmålene i sin opprinnelige form på engelsk for å unngå ekstra feil i forbindelse med oversettelser.

De fire kategoriene ga oss verdifull informasjon om hva deltakerne synes om en metode. Usefulness kan tolkes som hvor nyttig og effektiv en metode er i forhold til sin hensikt om å teste for tilgjengelighet. Satisfaction er hvor tilfredsstillende eller gøy det er å bruke en metode, og om man vil bruke den igjen. Ease of learning innebærer hvor lett det er å lære seg metoden og forstå den, mens ease of use er hvor lett det er å bruke metoden fra gang til gang. Sammen gir disse fire kategoriene et godt bilde over fordeler og ulemper med en metode.

3.1 Metode 1: WCAG-gjennomgang

Totalt fikk vi 19 personer til å evaluere en WCAG-gjennomgang. Det er den metoden færrest personer evaluerte, og vi klarte ikke å få flere til å gjøre en gjennomgang selv etter flere påminnelser. Motivasjonen til å gjøre en WCAG-gjennomgang var med andre ord ikke spesielt stor, noe som også reflekteres i evalueringen for metoden.

Som Tabell 2 viser, får WCAG-gjennomgang ganske dårlig skåre. Et gjennomsnitt på 4 for en kategori vil si at den er midt på treet (50 %). Samtlige av de fire kategoriene i USE er vurdert til under 4, og deltakerne synes både at sjekklisten er vanskelig å bruke (ease of use) og lite gøy å bruke (satisfaction) med verdier på [2.89, 0.50] (middelverdi og standardavvik) og [3.50, 0.29]. Det tilsvarer 31 % og 42 % av maks skåre som vist i Tabell 2.

Det som er positivt for WCAG er at man synes den er ganske nyttig med en gjennomsnittsverdi på [3.68, 0.54]. Det er også interessant at et del-spørsmål i kategorien usefulness skilte seg ut fra gjennomsnittet, nemlig spørsmålet «It is useful». Her var gjennomsnittet på 5.0, altså ganske mye høyere enn gjennomsnittet på 3,68. Det tyder på at de fleste synes WCAG-gjennomganger er nyttig i seg selv, men at helheten drar karakterene ned. Det reflekteres også ved en veldig lav satisfaction-verdi på [2.95, 0.52], som tyder på at de fleste ikke synes WCAG er spesielt artig å jobbe med.

Kategori	Gjennomsnitt	Standardavvik	Prosent
Usefulness	3,68	0,54	45 %

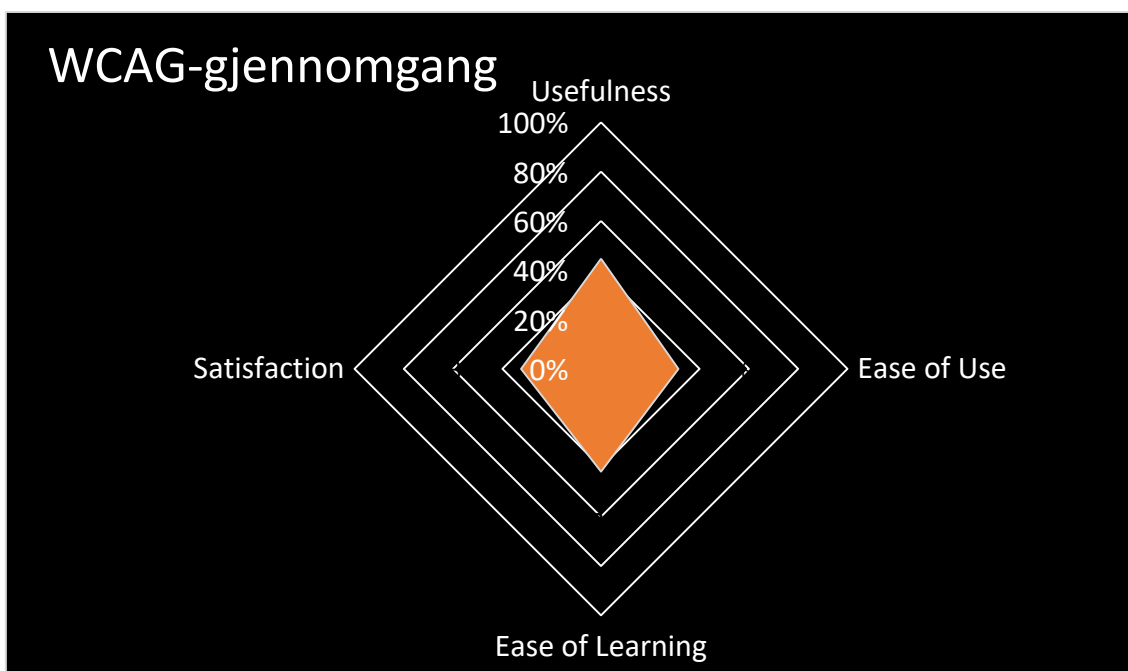
¹⁴ Lund, Arnold M. "Measuring usability with the use questionnaire12." *Usability interface* 8, no. 2 (2001): 3-6.

¹⁵ Brooke, John. "SUS: a retrospective." *Journal of usability studies* 8.2 (2013): 29-40.

Ease of Use	2,89	0,50	31 %
Ease of Learning	3,50	0,29	42 %
Satisfaction	2,95	0,52	32 %

Tabell 2. Evalueringsresultater for WCAG-gjennomgang

Hvis man ser på radar-diagrammet i figur 6, som er en populær måte å illustrere svarene i USE spørreskjemaet på, får man en fin visualisering av hvor godt WCAG gjennomgang som metode gjør det. Som forklart tidligere gjør ikke WCAG det spesielt bra, og metoden får en evaluering under 50 % av maks skåre på alle akser.



Figur 6. Radar-diagram for WCAG-gjennomgang evaluering

3.2 Metode 2: SiteImprove

Det var 35 personer som evaluerte SiteImprove, og det gjorde metoden til den nest mest evaluerte metoden. Som tabell 3 viser gjorde SiteImprove-metoden det veldig bra, med et høyt snitt for USE-spørsmålene i alle kategorier. De som evaluerte syntes at en automatisk sjekker, som SiteImprove, er en nyttig metode med [5.50, 0.30], som også er lett å lære [5.08, 0.24] samt bruke [4.85, 0.28].

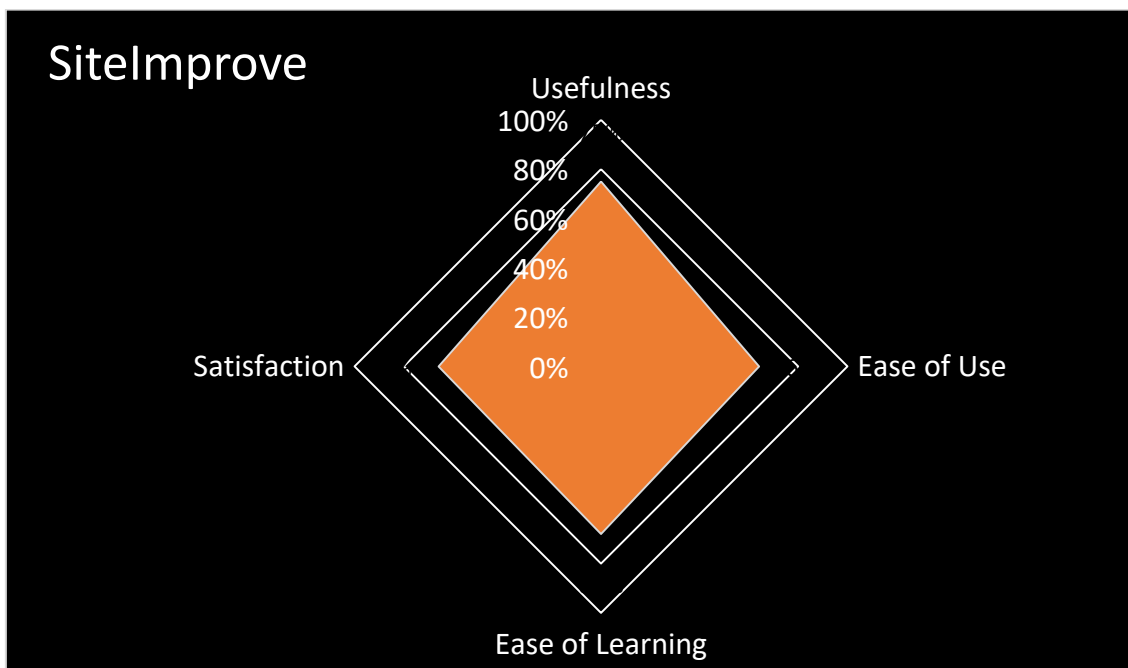
Det er ikke overraskende at man synes den er lett å lære (ease of learning) siden det er enkelt å installere utvidelsen i Chrome og det bare er å trykke på en knapp når man vil sjekke en side. Metoden gjør det også veldig bra på bruk (ease of use) og vi hadde egentlig forventet litt lavere verdier siden det er mange kompliserte termer og mye fagspråk som kan virker skremmende eller forvirrende om man ikke er vel bevandret i WCAG-terminologi. Evalueringene tyder også på at man synes SiteImprove er gøy å jobbe med, siden metoden får en verdi på [5.08, 0.24] for satisfaction.

Kategori	Gjennomsnitt	Standardavvik	Prosent
Usefulness	5,50	0,30	75 %

Ease of Use	4,85	0,28	64 %
Ease of Learning	5,08	0,24	68 %
Satisfaction	4,96	0,30	66 %

Tabell 3. Evalueringresultater for SiteImprove

Totalt gjorde automatisk sjekking med SiteImprove det meget bra, som også vises i figur 7 med et radar-diagram. Det er spesielt nyttigheten med SiteImprove som får veldig god evalueringer.



Figur 7. Radar-diagram for SiteImprove evaluering

3.3 Metode 3: Cambridge Simulation Glasses

Dette var den metoden som ble mest evaluert, og nesten alle testet denne metoden, totalt 50 personer. Ikke overraskende fikk metoden veldig bra verdier på både læring (ease of learning) med [6.80, 0.10] og bruk (ease of use) med [6.12, 0.43], siden det er en metode som bare krever at man tar på seg noen par med briller. Brillene får trolig litt lavere verdier på bruk siden metoden krever noe kalibrering før bruk. Når det gjelder nytten (usefulness) ved metoden får den en veldig god evaluering på [5.41, 0.54] som er tilsvarende som for SiteImprove-metoden. Spesielt delspørsmål «Is it useful» fikk veldig høy gjennomsnittsverdi på 6.38 som er godt over gjennomsnittet for hele kategorien.

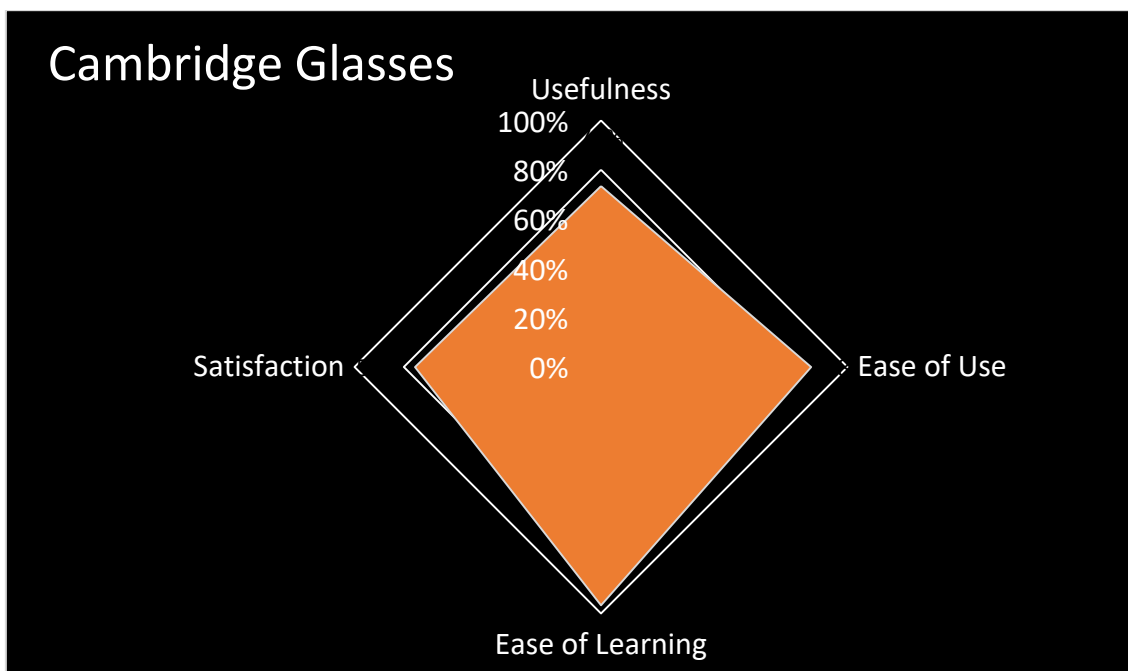
Når det gjelder gøy-faktoren får metoden en av de høyeste verdiene med [5.53, 0.69], noe som er svært positivt. Dette ble også reflektert i intervjuene hvor man trakk frem at metoden var morsom og gøy å bruke.

Kategori	Gjennomsnitt	Standardavvik	Prosent
Usefulness	5,41	0,54	73 %
Ease of Use	6,12	0,43	85 %

Ease of Learning	6,80	0,10	97 %
Satisfaction	5,53	0,69	75 %

Tabell 4. Evalueringresultater for Cambridge Simulation Glasses

Som figur 8 viser får Cambridge Simulation Glasses en veldig god evaluering på alle punkter. Den maksimerer nesten kategoriene læring og bruk, men har litt mer å gå på når det gjelder nytte og morsomhet.



Figur 8. Radar diagram for Cambridge Simulation Glasses

3.4 Metode 4: Skjermleser

Det var totalt 22 personer som evaluerte skjermlesermetoden, fordelt på 15 personer på NVDA-skjermleseren og 7 personer på skjermleseren VoiceOver. Totalt sett gjorde VoiceOver-skjermleseren det bedre enn NVDA, men resultatene er innenfor forventningsverdiene. Tabell 5 viser NVDA sin evaluering mens tabell 6 viser evalueringen av VoiceOver.

Tabell 7 viser den samlede evalueringen for begge skjermlesere, og man ser at skjermlesermetoden får en veldig bra evaluering på nytte (usefulness) med [5.49, 0.43]. Litt overraskende gjør skjermleser det bra på både læring [4.56, 0.25] (ease of learning) og bruk [4.36, 0.39] (ease of use). Vi hadde forventet lavere verdier her siden skjermlesere er kompliserte verktøy som det tar lang tid å beherske. Det kan kanskje ha sammenheng med at de fikk veiledning med å bruke utvidelser som gjør bruken av skjermlesere enklere, spesielt 'speech viewer' som beskrevet i avsnitt 2.1.4 For kategorien satisfaction så fikk skjermlesermetoden bra verdier på [4.80, 0.46].

Selv om skjermlesermetoden får høy verdi for nytte, så fikk den enda høyere verdier på delspørsmålet «Is it usefull» med en gjennomsnittsverdi på 6,32. Det er nesten like høyt som Cambridge Simulation Glasses. Skjermleser fikk også to verdier

som skilte seg ut i kategorien bruk (ease of use) med delspørsmålene «It is flexible» og «I can use it without written instructions». For «is it flexible» ble gjennomsnittsverdien 5.09, som er godt over kategori-gjennomsnittet, mens «I can use it without written instructions» fikk gjennomsnittsverdien 3.77 som er godt under kategori-gjennomsnittet. Ingen av disse resultatene er overraskende, men bekreftet forventningene vi hadde på forhånd om at skjermleser er nyttig, men komplisert å bruke.

Kategori	Gjennomsnitt	Standardavvik	Prosent
Usefulness	5,38	0,54	73 %
Ease of Use	4,32	0,52	55 %
Ease of Learning	4,50	0,35	58 %
Satisfaction	4,80	0,46	63 %

Tabell 5. Evalueringsresultater for NVDA-skjermleser

Kategori	Gjennomsnitt	Standardavvik	Prosent
Usefulness	5,75	0,38	79 %
Ease of Use	4,43	0,23	57 %
Ease of Learning	4,68	0,27	61 %
Satisfaction	4,84	0,55	64 %

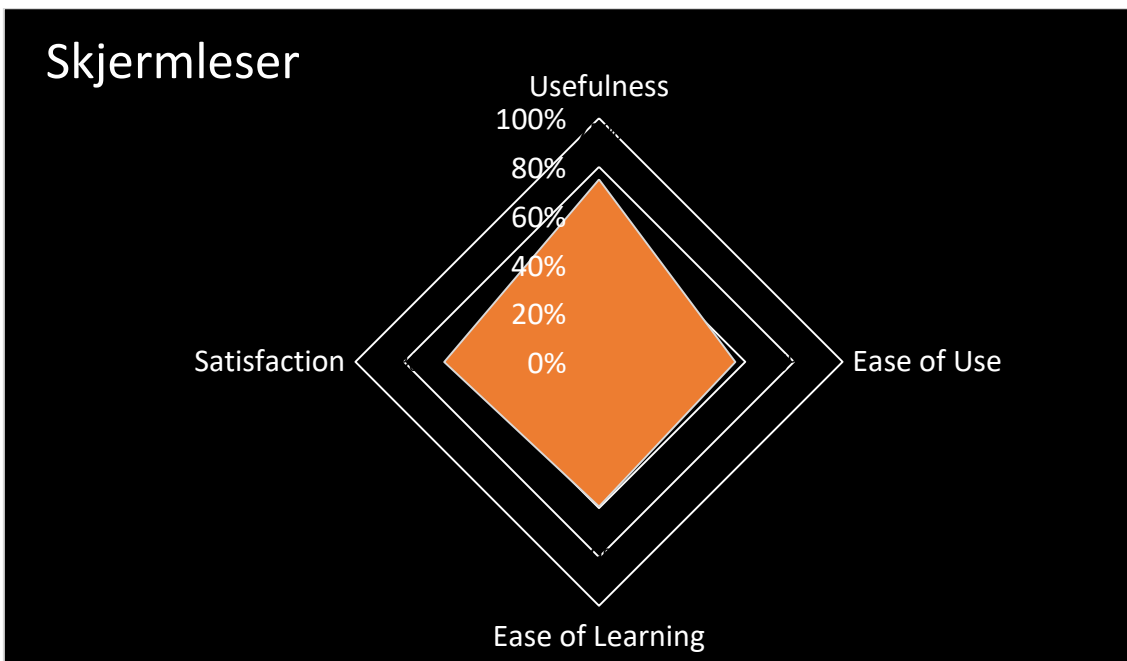
Tabell 6. Evalueringsresultater for VoiceOver-skjermleser

Kategori	Gjennomsnitt	Standardavvik	Prosent
Usefulness	5,49	0,43	75 %
Ease of Use	4,36	0,39	56 %
Ease of Learning	4,56	0,25	59 %
Satisfaction	4,81	0,46	64 %

Tabell 7. Evalueringsresultater for begge skjermlesere

Det er ikke overraskende at VoiceOver generelt får litt høyere verdier enn NVDA. VoiceOver er godt integrert i brukergrensesnittet, ettersom den er innebygd på IOS-enheter, mens NVDA er et gratis alternativ som kan installeres på Windows-plattformen. Det finnes andre skjermlesere for Windows som blir betraktet som bedre, men de har vi ikke fått evaluert i denne omgang primært pga lisenskostnader.

I figur 8 vises totalen for begge skjermlesere i et radar-diagram, og totalt sett gjør skjermleser-metoden det bra. Skjermleser har ikke overraskende lavere skåre på læring og bruk enn de andre to kategoriene, men gjør det likevel bedre enn forventet.



Figur 9. Radar-diagram for skjermleser

3.5 Metode 5: Dysleksi-simulering

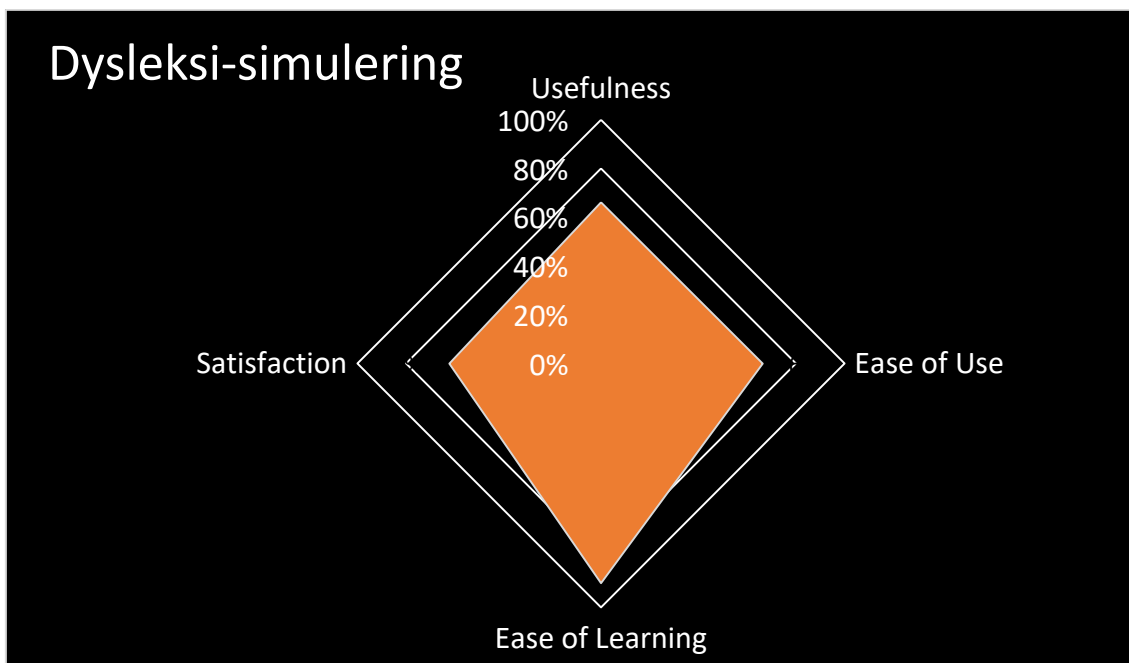
Totalt var det 29 personer som evaluerte dysleksi-utvidelsen, og som tabell 8 viser så gjør metoden det veldig bra. Det var en av metodene som gjorde det totalt sett best, etter Cambridge Glasses. På læring (ease of learning) fikk metoden verdier på [6.41, 0.18] som er meget høyt, bare slått av brillene. Dette kommer ikke så overraskende siden det stort sett bare er å trykke på en knapp. Både på nytten (usefulness) og bruk (ease of use) får metoden også veldig gode verdier på henholdsvis [4.97, 0.41] og [4.98, 0.49]. For bruk så er det den nest høyeste verdien, også her bare slått av Cambridge-brillene.

Kategori	Gjennomsnitt	Standardavvik	Prosent
Usefulness	4,97	0,41	66 %
Ease of Use	4,98	0,49	66 %
Ease of Learning	6,41	0,18	90 %
Satisfaction	4,73	0,38	62 %

Tabell 8. Evalueringsresultater for dysleksi-utvidelse

Overalt så gjør dysleksisimulerings-metoden det bra, spesielt innenfor læring og bruk som figur 10 viser. Det er en metode som får verdier godt over nøytralt (4). Metoden hadde muligens fått enda høyere verdier om det ikke hadde vært for noen småfeil i selve utvidelsen. Det gjenspeiles også i et av del-spørsmålene under kategorien bruk (ease of use), hvor deltakerne ble spurt «I don't notice any inconsistencies as I use it». Her ble gjennomsnittsverdien 3.86 som er betydelig lavere enn kategori-gjennomsnittet. Det har nok sammenheng med at utvidelsen ikke klarte å tilpasse alt av tekst (siden den bare endrer på tekst i spesifikke sideelementer).

Metoden får også gode verdier for hvor gøy det er jobbe med den (satisfaction) med verdier på [4.73, 0.38]. Spesielt synes deltakerne at metoden er morsom basert på delspørsmålet «It is fun to use» som får gjennomsnittsverdiene [5.41, 1.16].



Figur 10. Radar-diagram for dysleksi-simulering

3.6 Metode 6: Personas

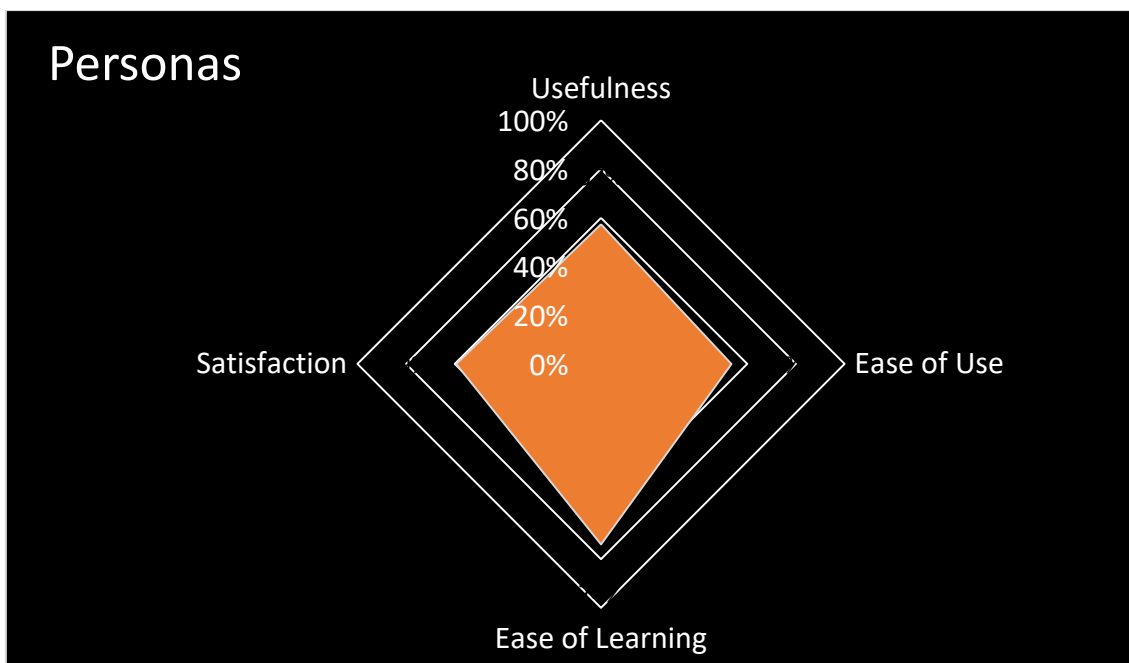
Den siste metoden ble evaluert av totalt 21 personer, og er en av metodene som fikk færrest evalueringer. Metoden fikk jevnt over gode verdier, men likevel en del dårligere enn de fleste andre på kategori nytte med verdier på [4.44, 0.39]. Dette er ikke direkte dårlige tall, men likevel en del under alle andre metoder med unntak av WCAG-gjennomgang.

Litt overraskende får metoden veldig gode evalueringresultater på læring (ease of learning) med verdier på [5.44, 0.32], selv om det er en forholdsvis komplisert metode fordi den krever evne til å sette seg inn i de tilgjengelighetsutfordringene som den aktuelle personasen kan møte på. Som forklart i avsnitt 2.1.6 fokuserte vi på en person med kognitive utfordringer, og vi hadde forventet at de fleste synes dette var en vanskelig rolle å lære. Det viser seg heller at de fleste synes det var lett å lære metoden, men at den var vanskelig å beherske, som 'ease of use' kategorien viser i tabell 9. På bruk får metoden [4.21, 0.41] som er det nest dårligste evalueringresultatet, omtrent tilsvarende som for skjermleser.

Kategori	Gjennomsnitt	Standardavvik	Prosent
Usefulness	4,44	0,39	57 %
Ease of Use	4,21	0,41	54 %
Ease of Learning	5,44	0,32	74 %
Satisfaction	4,56	0,53	59 %

Tabell 9. Evalueringresultater for personas

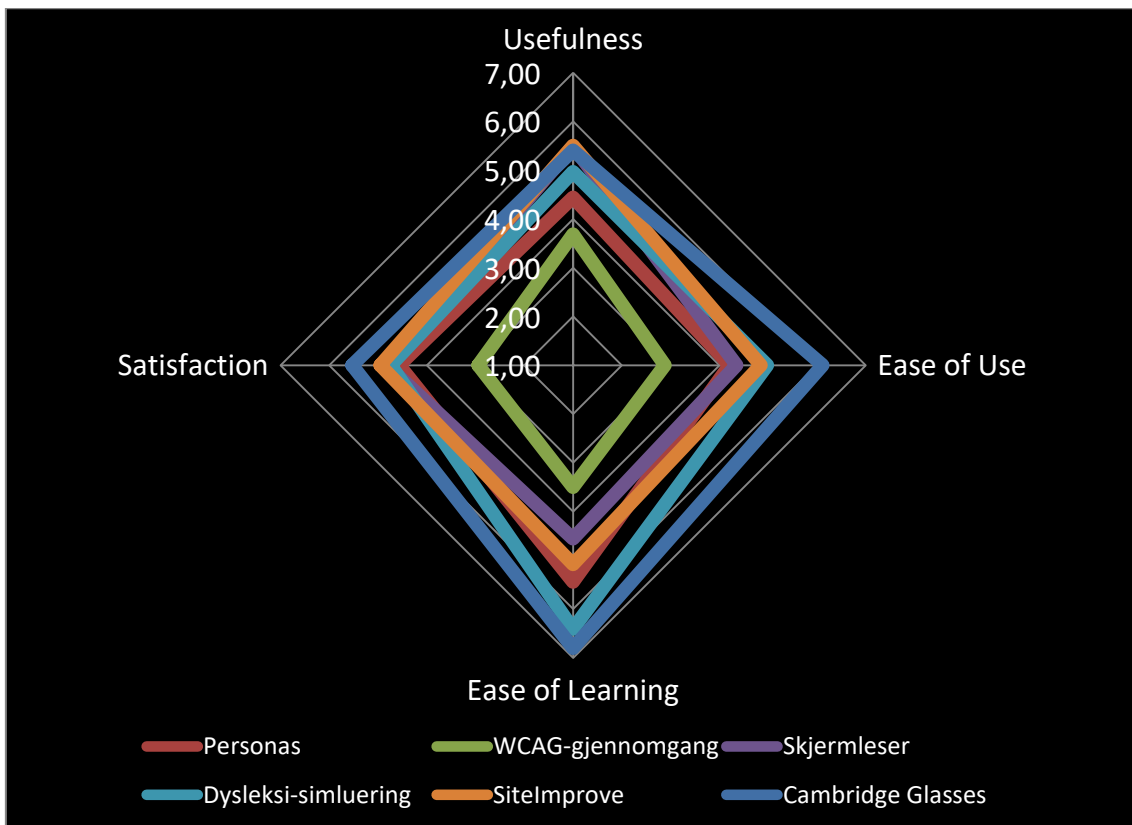
Som figur 11 viser gjør metoden det likevel ganske bra, med verdier over nøytralt. Vi møtte flere deltakere som hadde kjennskap til metoden og som brukte den aktivt selv i arbeidet, og vi synes det var veldig positivt å høre. Det var ikke alle deltakere som benyttet personas som den er tiltenkt, men vi tror ikke det er et problem så lenge man skaper bevissthet rundt de ulike brukergruppene. Det var derimot veldig få som hadde personas med kognitive utfordringer. Det som var overraskende var at så mange deltakere synes det var en veldig morsom teknikk, og dette resultat ble fanget opp både i sluttintervjuene, men også i et delspørsmål. For spørsmålet «It is fun to use» ble gjennomsnittsverdien [5.14, 1.49], som er betydelig høyere enn kategori-gjennomsnittet.



Figur 11. Radar-diagram for personas

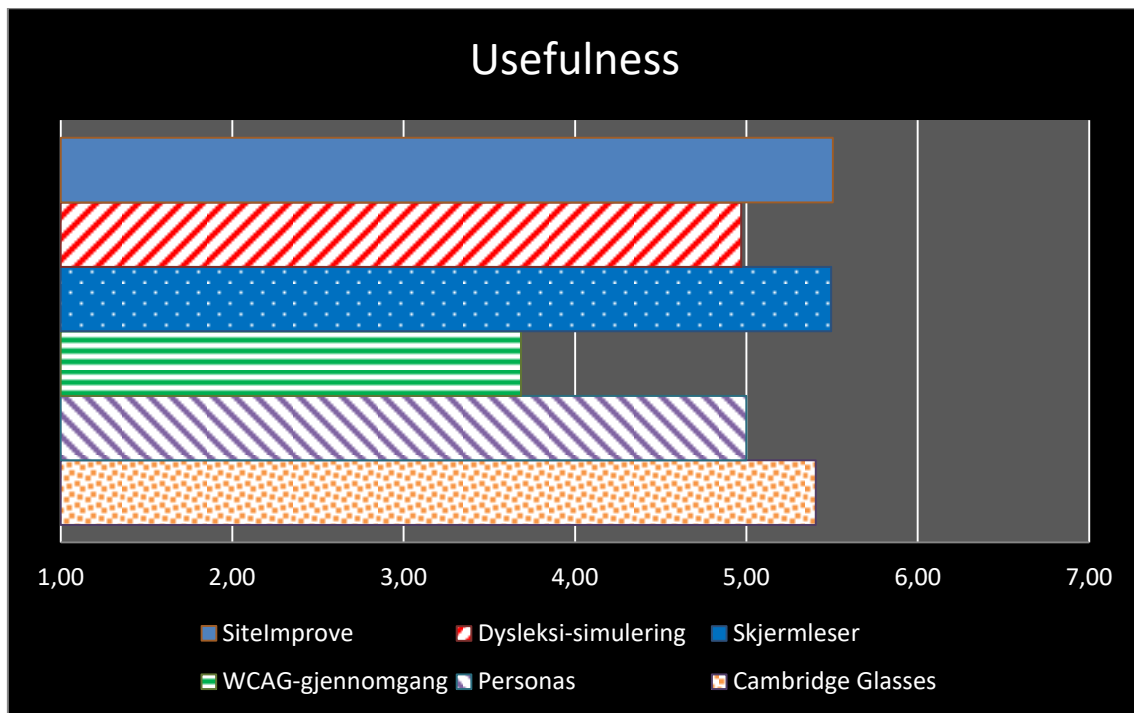
3.7 Alle metoder

For å sammenligne alle metodene opp mot hverandre, kan det være hensiktsmessig å plote de sammen for alle dimensjonene. I figur 12 vises et felles radar-diagram for alle metodene. Der ser vi at WCAG-gjennomgang er den metoden som skiller seg mest ut i alle dimensjoner, mens det er størst variasjon blant de andre metodene på lærekategorien (ease of learning) og bruk (ease of use).



Figur 12. Radar-diagram for alle metodene

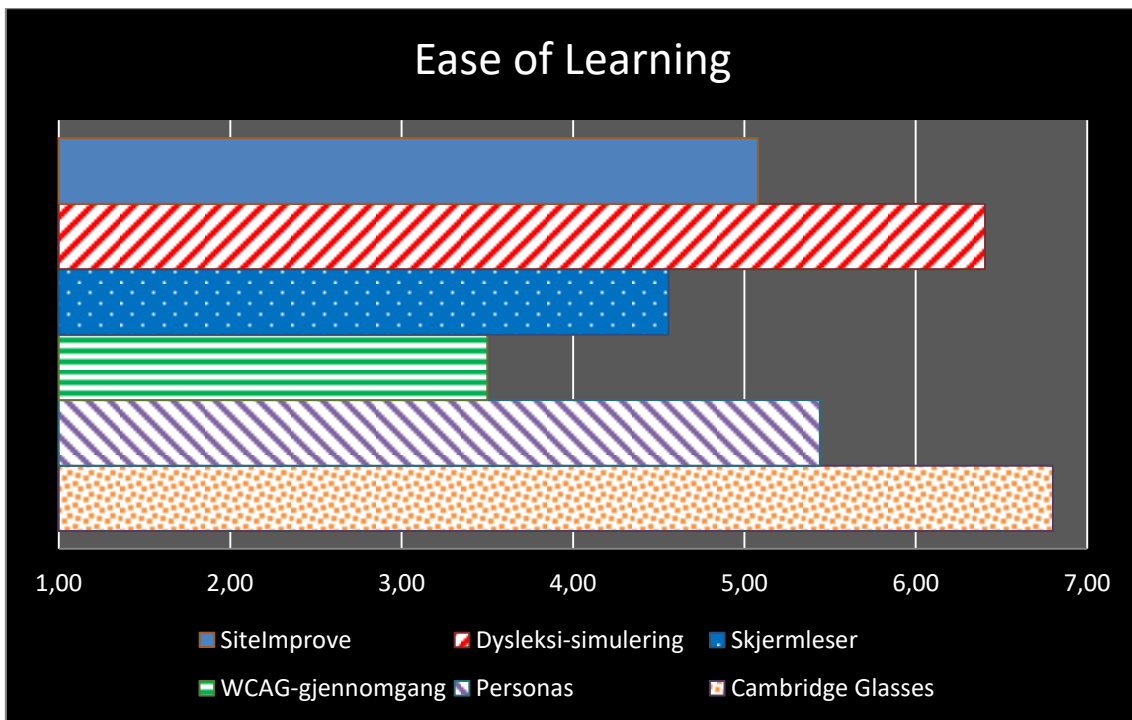
Hvis vi ser på de ulike dimensjonene hver for seg, så er det litt enklere å sammenligne de forskjellige metodene. I figur 13 vises USE-skåre for kategorien «nytte» for alle metodene. Med andre ord, hvor nyttig synes deltakerne at en gitt metode er for å teste tilgjengelighet. 4 er gjennomsnitt, og vi hadde forventet å finne de fleste mellom 4 og 5. Med unntak av WCAG-gjennomgang er det relativt likt mellom de fleste metodene. De får også veldig høye gjennomsnittsverdier rundt 5 og 5.5, altså litt høyere enn forventet.



Figur 13. Evalueringsresultater for alle metoder i kategorien usefulness

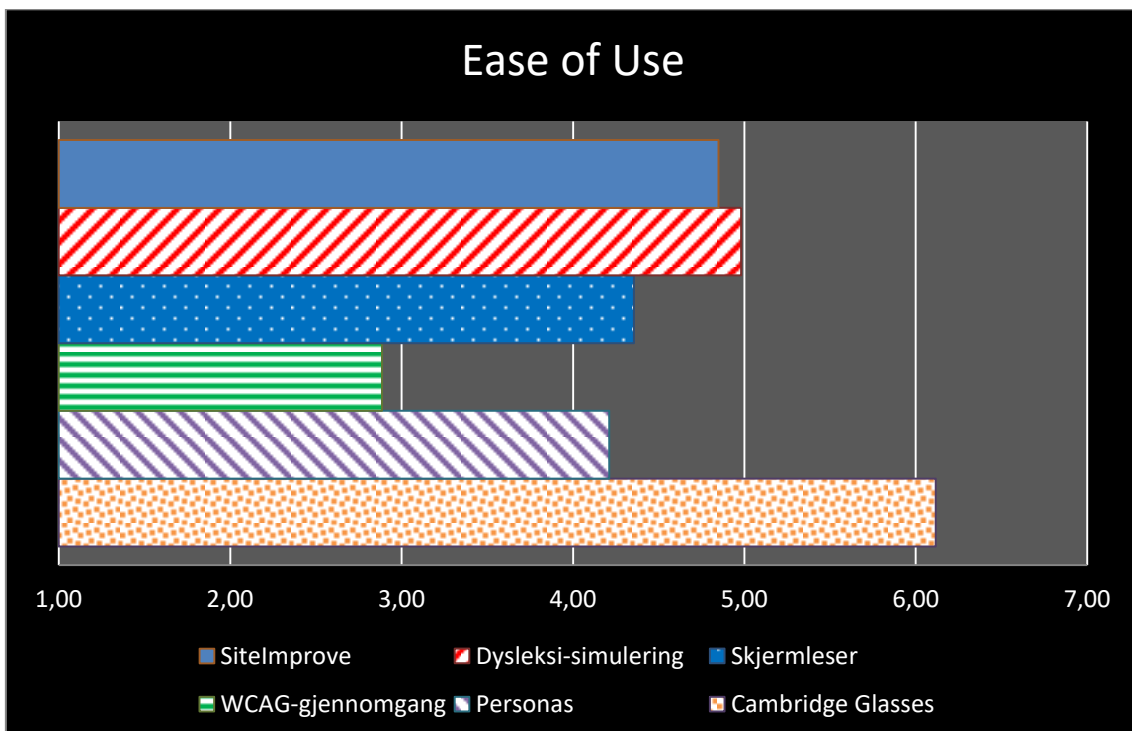
I figur 14 så kan man se resultatene for lære-kategorien. Ikke overraskende sa nesten alle at Cambridge-brillene var enkle å lære, siden de «bare er å ta på seg». Dysleksi-utvidelsen ble også betraktet som enkel siden det stort sett bare er å trykke på en knapp. Personas får overraskende bra vurderinger, på tross av at personas er en relativt komplisert metode. Kanskje litt overraskende får Siteimprove lavere vurdering enn Personas, men det kan være basert på en del tekniske termer som SiteImprove benytter.

Skjermleser blir også vurdert overraskende positivt, og det kan være fordi deltakerne fikk instruksjoner om hvordan man kan benytte en skjermleser når man ikke har synsnedsettelse. WCAG-gjennomgang blir vurdert som vanskelig å lære, kanskje overraskende dårlig i forhold til de andre. Det er derimot ikke overraskende at WCAG-gjennomgang får lavere verdi enn 4 som skal angi nøytralt nivå.



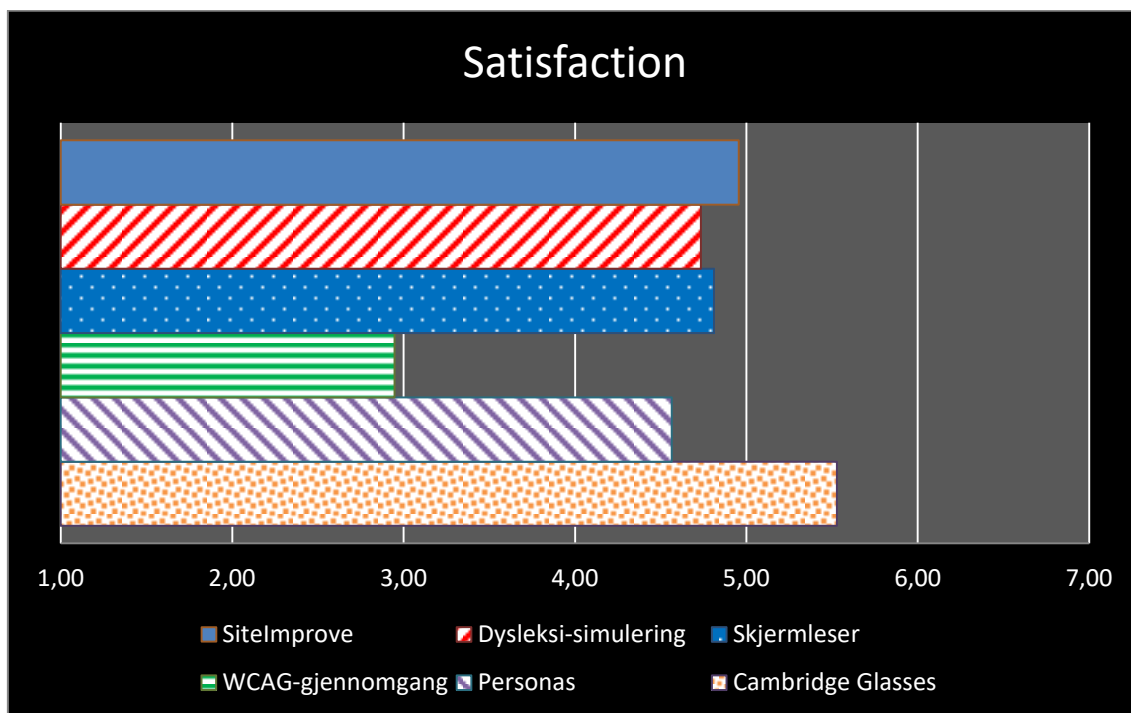
Figur 14. Evalueringresultater for alle metodene i kategorien ease of learning

Når det gjelder bruken av metodene, så er fire av seks metoder (SiteImprove, dysleksi-simulering, skjermleser og personas), evaluert til mellom 4 og 5 som vist i figur 15. Cambridge brillene får derimot en veldig god vurdering. Det er ikke overraskende at brillene får god vurdering, men litt overraskende at ikke Dysleksi-simulering får høyere. WCAG-gjennomgang blir igjen vurdert som svært dårlig.



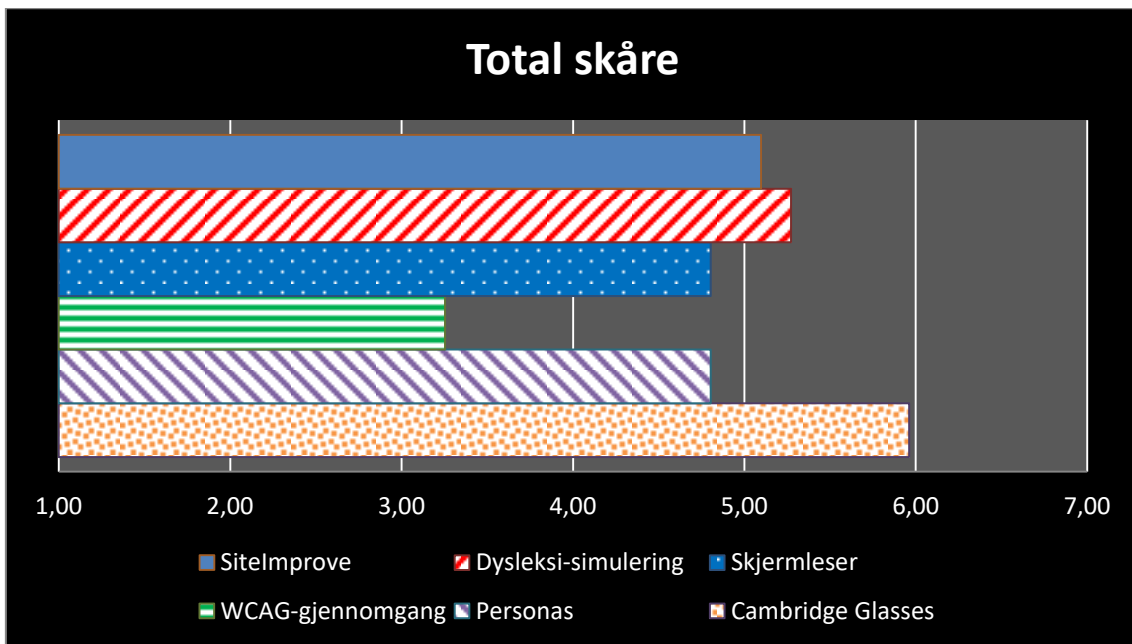
Figur 15. Evalueringsresultater for alle metodene i kategorien ease of use

Den siste dimensjonen som går på hvor morsomt eller fornøyd deltakerne er med en metoden vises i figur 16. Figuren støtter oppunder det sluttintervjuene fortalte oss: Cambridge brillene ble nesten enstemmig omtalt som veldig morsomt og innsiktsfullt å benytte, mens WCAG-gjennomgang ble opplevd som tungvint og slitsomt. De fire andre er ganske like, men likevel klart positivt ladet.



Figur 16. Satisfaction evaluering for alle metodene

Det kan også være hensiktsmessig å vurdere alle kategoriene sammen for å gi en helhetlig verdi eller skåre på metodene. En vanlig fremgangsmåte er å ta gjennomsnittet for alle kategorier, og figur 17 viser en total skåre for metodene opp mot hverandre. Her kommer det litt mer tydelig frem at det er WCAG gjennomgang som skiller seg negativt ut med en lavere evaluering enn resten, mens Cambridge Glasses skiller seg positivt ut med en høyere evaluering enn resten. De resterende metodene får også veldig god totalevaluering på rundt 5.



Figur 17. Totalevaluering for alle metodene

4 Integrasjon

Med utgangspunkt i intervjuguidene fikk vi mange gode tilbakemeldinger og synspunkter på de ulike metodene, hvordan de kan integreres i en smidig prosess og hvordan prosjekter håndterer tilgjengelighetstesting i dag. Totalt fikk vi gjennomført sluttintervju med over 40 personer.

Veldig mange av deltakerne syntes det var artig å se noe av det som finnes av metoder for å teste tilgjengelighet. Flere synes også at det er vanskelig å teste tilgjengelighet uten verktøy, men at noen metoder har en mer subjektiv tolkning enn andre. Vi fikk flere tilbakemelding på at metodene overlapper hverandre godt, men at de har hver sine fokusområder. Det synes vi var positivt å høre, siden vi var spente på å se om deltakerne selv tenkte over hvilke funksjonsnedsettelse de ulike metodene var ment å teste for, som forklart i avsnitt 2.1.

Felles for alle metoder og de fleste deltakere var at metodene skapte bevissthet om universell utforming og tilgjengelighet. Siden WCAG-gjennomgang ble gjort i etterkant fikk vi ikke samlet inn noen tilbakemeldinger for denne metoden.

4.1 SiteImprove

En automatisk sjekker som SiteImprove fikk gode tilbakemeldinger fra de fleste, og det deltakerne likte aller best med metoden var følgende:

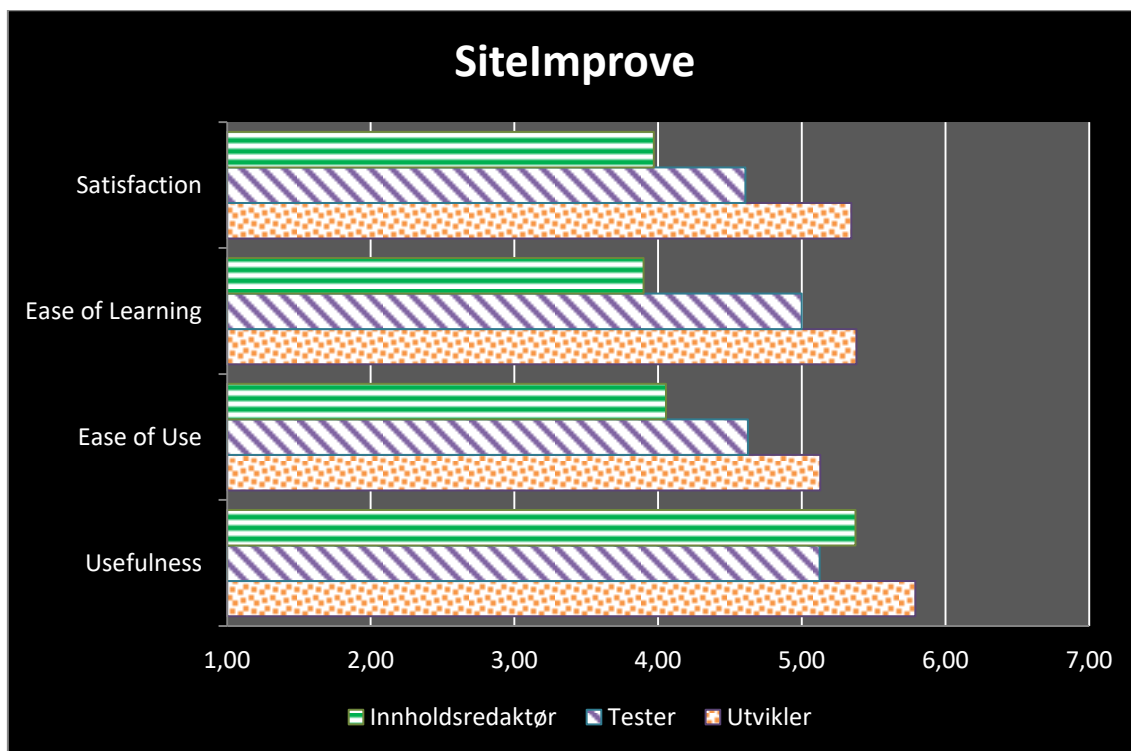
- Strukturert og oversiktlig
- Systematisk
- Objektive tilbakemeldinger
- Bra integrasjon med nettleserens egne utviklerverktøy (FireFox, Chrome)
- Avdekker mye på kort tid

Det som derimot ble påpekt som negativt med en automatisk sjekker som SiteImprove var følgende:

- Det er ikke så enkelt å skjønne nøyaktig hvor alle feilene er
- Ikke alle feil og tilbakemeldinger er like lette å tolke
- Man kan se seg litt blind på alle tilbakemeldingene
- Menyen kan ødelegge litt for selve siden som testes
- Fordel å ha kunnskap om WCAG-standarden

Denne metoden vil fungere fint å integrere i smidige team straks man har en fungerende versjon eller prototype. Metoden kan kjøres ofte, men er vanskelig å automatisere fordi det er en nettleser-utvidelse. Den har også utfordringer med å kunne nå alle sider, og det er usikkert hvor godt SiteImprove fungerer med dynamisk innhold og JavaScript komponenter.

Basert på observasjoner gjort gjennom evalueringene og intervjuene er det en klar trend at utviklere liker SiteImprove som et verktøy, og automatiserte sjekkere som en metode. Mange er vant til tilsvarende verktøy i forbindelse med å sjekke kodekvalitet og sikkerhetshull, og sånn sett passer en automatisert tilgjengelighetssjekker godt inn i deres verden.



Figur 18. Evalueringsresultater for SiteImprove for ulike roller.

I figur 18 vises evaluering av metoden SiteImprove fordelt på de tre rollene innholdsredaktør, tester og utvikler. De to andre rollene, leder og design har for få evalueringer (under 5) til å kunne analyseres. Alle disse rollene ble selvrappert av deltakerne i forkant vha. en spørreundersøkelse. For en utfyllende rapport om rollene, se NR rapporten 1036, «Holdninger rundt universell utforming i smidige team»¹⁶.

Figuren viser, ikke uventet, at det er store forskjeller mellom rollene, og at utviklere er de mest positive til denne metoden. Innholdsredaktør er stort sett de mest nøytrale til metoden, mens testere ligger midt imellom. Det er kanskje noe uventet at det er utviklere og ikke testere som er mest glad i metoden, men vi tror det er fordi SiteImprove ligner mye på tilsvarende nettleser-utvidelser som utviklere benytter ofte. Alle rollene synes metoden er nyttig med evalueringer over 5, men det virker som om utviklere er de som virkelig synes metoden er morsom og tilfredsstillende å bruke. Det gjenspeiles av evalueringene for kategorien "satisfaction", samt av intervjuene.

4.2 Cambridge Simulation Glasses

Dette var den metoden som fikk best evaluering totalt, og med ganske god margin, noe som figur 17 viser. Det som ble trukket frem som positivt med denne metoden under intervjuene var:

- Ga mer helhetlig inntrykk av kontraster og tekststørrelser
- Ga innsikt og forståelse
- Effektivt og enkelt
- Lavterskel

¹⁶ Holdninger rundt universell utforming i smidige team. : norsk regnesentral 2018 (ISBN 978-82-539-0546-4) 53 s. Report at the Norwegian Computing Center (1036), NR, UiO

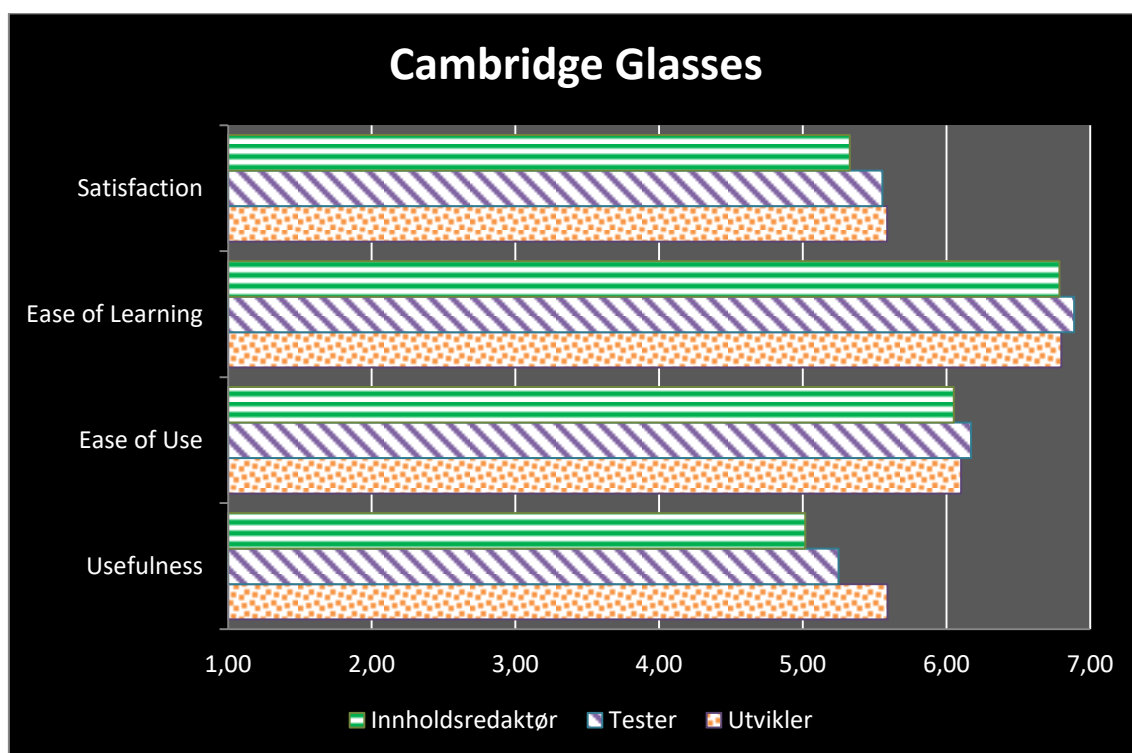
- Konsistent over tid
- Gøy å bruke

Selv om brillene fikk få negative tilbakemeldinger, var det flere gode innspill som påpekte svakheter ved metoden:

- Litt enkelt å jukse med
- Kan være vanskelig å finne riktig antall briller
- Vanskelig å få til å sitte ordentlig
- Slitsomt å bruke i lengden (mer enn 15 min)

Selv om dette var en metode som alle likte, så var det designere som ble mest frelst av metoden. En designer sa «*Folk hadde skjønt hva jeg prater om når det gjelder kontrast*», og dette ble delt av mange designere. Flere grupper nevnte også at å bruke brillene føles litt ut som å «Ta på seg brukerens briller» og at det ofte ga en liten a-ha opplevelse, spesielt for de som ikke var så kjent med uu og tilgjengelighet.

Denne metoden er en av de få som kan benyttes i alle faser i en utviklingsprosess, fra designskisser til sluttprodukt. Spesielt det at man kan benytte brillene allerede i en skisse- og design-fase var appellerende for mange. Noen foreslo å ta det med seg i møter, for å vise andre hvorfor noen kontraster er dårligere. Det kan være problematisk i noen prosjekter å få frem dette før man har en prototype, og noen opplever da at veien til å endre er lengre enn på et skissestadium.



Figur 19. Evalueringresultater for Cambridge Glasses for ulike roller.

Evalueringen av metoden Cambridge Glasses var ganske jevnt fordelt mellom de ulike rollene som figur 19 viser. Det er noen forskjeller i «usefulness» kategorien, hvor det er utviklere som er mest positive til metoden, men ellers er rollene samstemte.

4.3 Skjermleser

Skjermlesermetoden fikk gode evalueringer på de fleste kategoriene, og det som ble nevnt oftest i intervjuene som positivt med metoden var:

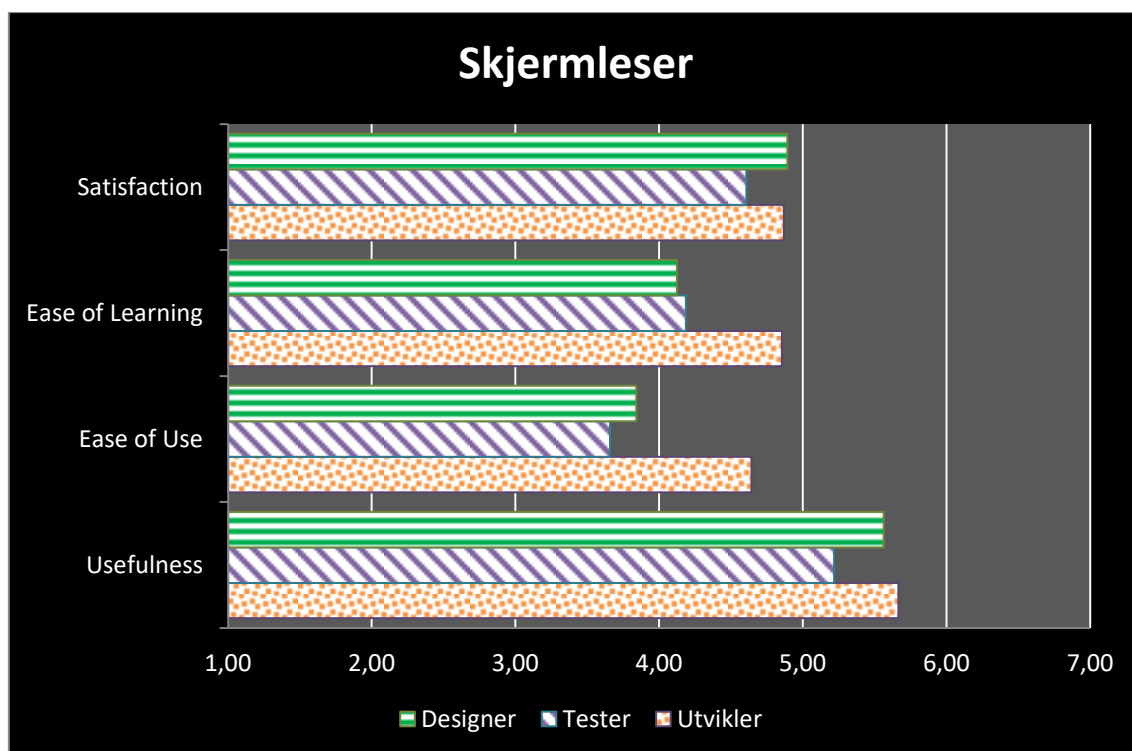
- Ga en aha-opplevelse
- Fikk mer forståelse for blinde/svaksynte og de som må bruke tastatur
- Avdekker flest kritiske feil

Det var også en del negative tilbakemeldinger under intervjuene, og de kom ikke som noen overraskelse:

- Vanskelig å bruke
- Krevende å sette opp

I tillegg nevnte flere at det krever litt mer enn bare en prototype for å kunne benytte skjermleser. Dette begrunnet de fleste med at en prototype ofte ikke har fokus på riktig struktur og kode fra et tilgjengelighetsperspektiv, og at en skjermleser derfor vil ha store problemer med å «lese» og «tolke» en slik prototype. Vi ser at dette er et typisk argument for å utsette bruken av skjermleser mot slutten, og det var ofte en person i prosjektene som hadde en slik innstilling. Dette er derimot et farlig argument, siden man da risikerer å måtte gjøre større kodeendringer for å rette feilene på et senere tidspunkt, og at terskelen for å rette opp både struktur, flyt og kode derfor blir høyere.

Dette er derfor en metode som burde integreres så fort som mulig, allerede fra prototype-stadiet. Det vil medføre at man tenker tilgjengelighet fra første stund, og man kan dra nytte av det i form av bedre brukerinteraksjoner for alle. Selv om prosjektet ikke har tid til å rette skjermleser- og tastaturnavigasjon-feil som oppdages i en tidlig fase, så får man i det minste registrert feilene som må rettes på et seinere tidspunkt. Dessuten er det viktig å rette de mest opplagte skjermleserfeilene før man kan gjennomføre testing med skjermleserbrukere for å unngå å kaste bort tid.



Figur 20. Evalueringsresultater for skjermleser for ulike roller.

Det var ikke nok innholdsredaktører som fikk evaluert skjermleser, men det var derimot flere designere som evaluerte metoden. Derfor er innholdsredaktører byttet ut med designere i figur 20, som viser evalueringsresultatene for skjermlesermetoden. Det er igjen utviklere som jevnt over mest fornøyd med metoden. Alle rollene synes metoden, ikke overraskende, er nyttig, men i kategoriene «ease of learning» og «ease of use» er det større forskjeller mellom utviklere og de andre rollene. Utviklere er generelt mer positiv til hvor lett det er å lære og beherske en skjermleser, og det er kanskje basert på at de er vant til å bruke kompliserte verktøy. Det kan også være fordi mange utviklere er vant til å bruke «tabbe testen» hvor man bruker TAB tasten for å sjekke at navigasjon med tastatur fungerer.

4.4 Dysleksi-simulering

Denne metoden fikk litt overraskende for oss, nest beste total skår, som figur 17 viser. Det er flere grunner til dette, men det de fleste trakk fram i intervjuene som positivt med metoden var:

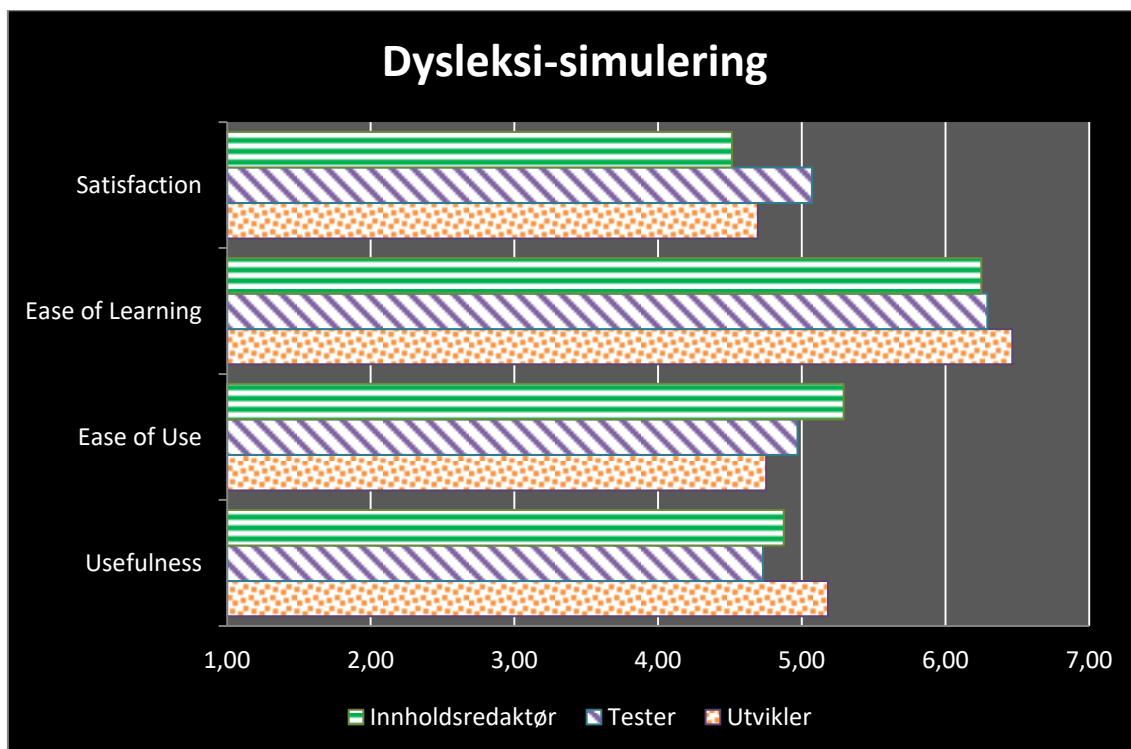
- Ga en aha-opplevelse
- Fint å kunne visualisere problemer med for mye tekst
- Enkelt
- Mer oppmerksom på struktur av sider
- God påminnelse om å skrive enkelt og godt språk

Det var derimot en del kritikk med metoden også, som f.eks.:

- Fungerte ikke på all tekst
- Vanskelig å skjønne når noe er feil
- Krever forkunnskaper om dysleksi

Det at metoden gav en aha-opplevelse og at verktøy for å teste for lesbarhet er lite kjent kan være en viktig grunn til at metoden fikk såpass god vurdering. Selv om metoden fikk veldig gode evalueringer, så var det flere som ikke forstod helt utfordringen med dysleksi eller hvordan man skal løse utfordringer knyttet mot dysleksi. For eksempel sa en deltaker at «dysleksi er litt sånn 'stakkars, men hva skal vi gjøre med det?」 Det var også problematisk for noen å skjønne at metoden simulerer opplevelsen og ikke dysleksi i seg selv. For andre så var det en helt annen opplevelse, en deltaker beskrev innholdsredaktør-rollen som «en kamp i å skrive enkelt og forståelig», og syntes at denne metoden hjalp med å tydeliggjøre viktigheten av dette.

Denne metoden er utfordrende å integrere i en utviklingsprosess fordi den er tett knyttet mot tekst som ofte ikke blir ferdigstilt før mot slutten av lanseringen i mange prosjekter. Det kan derimot brukes som et argument for at man burde involvere innholdsredaktører og de som skriver tekster tidligere i prosessen, slik at man får et overblikk over hvordan løsningen er tilrettelagt for de som produserer innhold. Hvis strukturen er veldig fastlåst og innholdsredaktører har liten frihet, vil denne metoden kunne avdekke svakheter med dårlig struktur og lesbarhet.



Figur 21. Evalueringresultater for dysleksi-simulering for ulike roller.

De ulike rollene var relativt samstemte i evalueringen av metoden, som figur 21 viser.

4.5 Personas

Denne metoden er utfordrende, fordi den er forholdsvis enkel å forstå, men vanskelig å beherske. Det vil si, for å gi realistiske svar, krever den kunnskaper om eller god evne til å sette seg inn i de utfordringene personasen vil kunne oppleve med hensyn til tilgjengelighet på web. Det ble også gjenspeilt i de mange tilbakemeldingene vi fikk gjennom intervjuene. Det som ble nevnt som positivt med metoden var:

- Lett å sette seg inn i
- Enkelt å lære seg empati
- Kan fungere godt sammen med andre metoder
- Kreativ måte å jobbe på

Det var derimot ganske mye kritikk også, og det var bakgrunnen for at metoden var en av de som fikk dårlig skåre. Noen av kritikken som gikk igjen hos flere var:

- Vanskelig å beherske
- Rart å spille rollespill
- Vanskelig å sette seg inn i andre sitt perspektiv
- Vanskelig å legge fra seg egne vaner og erfaringer
- Innebærer mye synsing

Den største kritikken mot metoden er som en deltaker sa: «Personas må tolkes». De fleste var usikre på om tolkningen deres var riktig, men tatt i betraktning av at de ikke fikk mye tid på å sette seg inn i rollene så er ikke det uventet.

Personas er en av metodene som kan benyttes i alle faser i en utviklingsprosess, helt fra design skisser og prototyper til sluttprodukt. Spesielt i startfasen foreslo mange at den kan være nyttig for å fange problemer med f.eks. flyt og brukerinteraksjoner som medfører designendringer. Noen synes også metoden kunne prøves ut i både workshop og testing i par, og kanskje vil det gjøre det lettere å leve seg inn i rollen da.

Det var ikke nok andre roller enn utviklere som fikk evaluert metoden, og derfor viser vi ikke evalueringresultater for metoden fordelt på rollene.

5 Oppsummering

Uavhengig av hvilke metoder man bruker for å teste tilgjengelighet, synes mange at man burde teste ofte og hyppig. Et av spørsmålene i sluttintervjuene var «Hvor ofte kunne du tenke deg å teste for universell utforming?». Her synes de fleste at man burde gjøre det jevnlig, helt fra daglig automatisert testing og ukentlig testing til minst månedlig. De færreste synes man burde teste sjeldnere enn en gang i måneden, men tidligere undersøkelser viser at veldig få prosjekter tester for tilgjengelighet så ofte.

Lærdommen er altså å teste for tilgjengelighet hyppig og ha gode rutiner for det. En gang i måneden er det majoritet av deltakerne mener burde være et minimum for de fleste prosjekter.

Når det gjelder hvilke metoder man skal velge og når man skal anvende dem i utviklingsprosessen, så er det avhengig av mange faktorer. Noen tommelfingerregler kan vi likevel gi basert på våre observasjoner. Merk at selv om vi lister opp flere metoder for en fase, så betyr ikke det at man bør benytte alle for hele fasen eller for hver iterasjon, men at man kan veksle fra iterasjoner til iterasjon for å få variasjon.

5.1 Designfase

I en tidlig fase av en prosess hvor man ikke har en prototype å teste mot, er det mest hensiktsmessig å benytte metoder som kan brukes direkte mot designskisser:

- Cambridge simulation glasses
- Personer med kognitive utfordringer (ADHD, senior etc)

Begge disse metodene kan som forklart brukes mot designskisser og avdekke feil som dårlig kontraster, tekststørrelser, dårlig struktur på sider, flyt og overkompliserte løsninger etc.

Som designer er man i mange møter med ledere og utviklere, og må ofte forklare design-konsepter for ulike grupper, og ting må legges frem på litt ulik måte for de forskjellige gruppene. Her er det spesielt brillene som er et nyttig verktøy, og man kan til og med ta med seg noen par i møtene.

5.2 Utviklingsfase

Selv om vi har splittet opp i utviklingsfase og testfase er de som oftest i kombinasjon og foregår i samme iterasjon. Vi har likevel valgt å holde dem litt adskilt siden det ofte er forskjellige personer som utfører oppgavene. Det er spesielt tre metoder som er gode å bruke i utviklingsfasen, og spesielt av utviklerne selv:

- Cambridge simulation glasses
- Skjermlesere
- Automatiske sjekkere som SiteImprove

I tillegg til at disse avdekker tilsvarende feil som i designfasen, så vil de også oppdage strukturelle feil i koden, feil koding av elementer (spesielt HTML/CSS), brudd på

WCAG-retningslinjer etc. Spesielt strukturelle kodefeil er det viktig å fange opp tidlig for å unngå «kodegjeld» som blir kostbart på et seinere tidspunkt.

Automatiske sjekkere kan med fordel integreres i automatiske byggeprosesser som for eksempel kjøres hver natt eller hver gang kode sjekkes inn i versjonssystemet.

5.3 Testfase

Selv om vi har definert en egen testfase, så kan denne gjerne kjøres parallelt med utviklingsfasen. Ofte har dedikerte testere mulighet til å gjennomføre lengre testsesjoner, og derfor mener vi at flere metoder burde benyttes i denne fasen:

- Cambridge simulation glasses
- Skjermleser
- Personas med kognitive utfordringer (ADHD, demens etc)
- Dysleksi-simulering
- WCAG gjennomgang

Disse metodene dekker mye av det som blir fanget opp av metodene i utviklingsfasen, men personas med kognitive utfordringer kan være en effektiv metode å supplere med. Den fanger opp feil som ikke er så lett å avdekke med noen av metodene som er listet opp under utviklingsfasen, på tilsvarende måte som at en manuell WCAG gjennomgang fanger opp problemstillinger som ingen automatiske sjekkere klarer.

5.4 Vedlikeholdsfase

Dette er gjerne en fase som kommer etter at løsningen er lansert og den intense utviklingsfasen fram mot leveranse erstattes av faser med fokus på å rette feil og forbedre det som allerede er levert, i tillegg til å lage ny funksjonalitet. Både ressurser, budsjett og endringer er ofte i mindre skala enn for nyutvikling. Det medfører at man ikke trenger å teste like grundig som i nyutvikling, men med mindre det kommer større endringer. I dette tilfellet kan man argumentere for at man starter en nyutvikling av deler av løsningen og bør følge de andre fasene. I vedlikeholdsfasen anbefaler vi følgende metoder:

- Briller
- Automatiske sjekkere
- Dysleksi-simulering

Felles for alle disse metodene er at de er raske å gjennomføre, og gir en god indikasjon på tilgjengeligheten. Alle metodene fanger også opp ting som typisk endres hyppig under vedlikeholdsfasen, nemlig innhold og grafiske elementer. Her vil alle tre metodene avdekke utfordringer hvis løsningen er mangelfull.

5.5 Brukertesting

Selv om vi ikke har snakket spesielt om brukertesting så langt, betyr ikke det at man ikke bør gjennomføre brukertester om man bruker metodene vi har anbefalt. Tvert imot, så er behovet for brukertesting like sterkt! Det er først da man får verifisert om man har brukt metodene riktig, og avdekket de fleste feil og utfordringer. I beste fall går brukertestene smertefritt og man får en god indikasjon på hvor godt løsningen fungerer. I verste fall klarer man ikke å gjennomføre brukertester fordi man glemte å

teste med f.eks. skjermleser, og en blind person klarer ikke å navigere i løsningen eller forstå konteksten.

Vi anbefaler derfor å gjennomføre brukertesting gjennom hele utviklingsløpet i tillegg til metodene vi har diskutert i denne rapporten. Ved å benytte testmetodene så vil man luke ut de verste feilene før en brukertesting. Dette betyr lavere kostnader med tanke på endringer og mer effektive brukertester fordi helt enkle feil er luket ut på forhånd. Det er frustrerende for alle parter hvis brukertesting stopper opp etter kort tid fordi for eksempel skjermleser ikke er mulig å benytte. Det vil man fort avdekke med skjermlesermetoden.

5.6 Organisasjon

Det er også noen tilbakemeldinger vi har fått, som går på organisasjon som helhet og ikke prosjektet man jobber på. De tilbakemeldingene vi har fått, og som vi mener burde forankres hos ledelsen og gjennomsyre alle prosjekter er:

- Konsekvens hvis tilgjengelighet av løsninger er dårlig, og insentiver til å gjøre løsninger mer tilgjengelig enn «bare» WCAG-gjennomgang.
- Skape bevissthet rundt universell utforming, fra topp til bunn i organisasjonen.
- Stille krav til konsulenter og leverandør om kompetanse på uu.
- Opplæring i uu og metoder for å teste tilgjengelighet.
- Fokuserer på ulike funksjonsnedsettelse.
- Akseptere at det tar litt lengre å gjøre løsninger tilgjengelige, og at det dermed må avsettes tid i prosjektplanlegging.

Vi vil avslutte med en kommentar som flere deltakere har gitt oss: «Det er viktig å fokusere på 80/90% og ikke gi opp fordi man ikke klarer 100%». Eller sagt med andre ord: Alle monner drar!

6 Vedlegg A

6.1 USE-spørsmål

Spørsmålene som ble stilt i forbindelse med en USE evaluering er disse:

USEFULNESS

1. It helps me be more effective
2. It helps me be more productive
3. It is useful
4. It gives me more control over the activities in my life
5. It makes the things I want to accomplish easier to get done
6. It saves me time when I use it
7. It meets my needs
8. It does everything I would expect it to do

EASE OF USE

9. It is easy to use
10. It is simple to use
11. It is user friendly
12. It requires the fewest steps possible to accomplish what I want to do with it
13. It is flexible
14. Using it is effortless
15. I can use it without written instructions
16. I don't notice any inconsistencies as I use it
17. Both occasional and regular users would like it
18. I can recover from mistakes quickly and easily
19. I can use it successfully every time

EASE OF LEARNING

20. I learned to use it quickly
21. I easily remember how to use it
22. It is easy to learn to use it
23. I quickly became skillful with it

SATISFACTION

24. I am satisfied with it
25. I would recommend it to a friend
26. It is fun to use
27. It works the way I want it to work
28. It is wonderful
29. I feel I need to have it
30. It is pleasant to use

6.2 Scenario-oppgaver

Disse oppgavene ble gitt i forbindelse med metode evalueringene. Vi brukte maks to oppgaver per metode, og det var alltid en metode som var eksplorativ og som ikke fikk oppgaver.

6.2.1 Oppgave 1: Lavpriskalenderen

Du har lyst til å besøke din kjæreste som bor i Bergen en helg i februar. Det har ikke noe å si hvilken helg, men du vil ha så billige som mulige billetter.

1. Gå til www.sas.no
2. Finn lavpriskalenderen til SAS. Bestill en billig reise Oslo- Bergen som går fra fredag ettermiddag til søndag ettermiddag i februar. (Du vil betale med penger, ikke poeng)
3. Etter at du har fylt inn navnet ditt så ønsker du å legge til bonusprogram EBB600575700 [dette er fiktivt og vil ikke fungere].
4. Finn et ledig sete ved vinduet

6.2.2 Oppgave 2: Er flyet i rute

Du har tilbudt deg å hente en kamerat som kommer med fly fra London i kveld

1. Gå til www.sas.no
2. Finn siden der man kan se om flyet er i rute
3. Søk opp flightnummeret som er SK506
4. Sjekk stausen på flyet, er det i rute?
5. Du blir nysgjerrig på linken "Operert av 4 carriers", og klikker på denne
6. Hvilke selskaper opererer på denne ruten?

6.2.3 Oppgave 3: Bli medlem i EuroBonus

Du flyr ofte med SAS og har bestemt deg for å bli medlem i SAS EuroBonus

1. Gå til www.sas.no
2. Finn registreringsskjemaet for å bli medlem av Eurobonus
3. Fyll inn fornavn og etternavn
4. Når du kommer til feltet for fødselsdato skriver du inn datoen på gammel vane på formatet dag-måned-år (ddmmyy), og tabber videre til neste felt
5. Fyll inn resten av skjemaet

6. Åpne "Betingelser og vilkår"

1. Hvem skal du kontakte for å få vite mer om Eurobonus?

6.2.4 Oppgave 4: Reise med katt

Du skal på ferie til familiens sommerhus i Danmark og vil undersøke hvordan katten kan være med på flyturen

1. Gå til www.sas.no
2. Finn ut hvor raskt du må bestille plass til katten på reisen
3. Finn ut hvor mye det koster å frakte katten i kabinen

6.3 Intervjuguide

Vi fulgte intervjuguiden under, og stilte oppfølgings spørsmål der hvor deltakeren hadde mer informasjon å gi.

1. Hva synes du om de ulike testmetodene?
2. Hvilken metode synes du avdekket mest feil?
 1. Gjerne utdyp. Hva er grunnen til dette? Hva slags type feil?
3. Hvilken metode synes du var enklest å bruke?
 1. Utdyp. Hva var der spesifikt som var enkelt?
4. Hvilken metode synes du var vanskeligst å bruke?
 1. Utdyp, hva spesifikt var vanskelig?
5. Hvilke metoder kunne du tenke deg å bruke neste gang du skal teste for universell utforming?
 1. Hva er sannsynligheten for at dette vil bli gjort de neste tre ukene tror du?
6. Hvor ofte kunne du tenke deg å teste for universell utforming (uten at det er en Jira-task)?
7. Hva slags påvirkning tror du universell utforming har for løsninger dere lager?
8. Hvordan tror du best metodene kan integreres i de daglige rutinene?