

**SINTEF IKT**

Postadresse: 7465 Trondheim
Besøksadresse: Forskningsveien 1
Telefon: 22 06 73 00
Telefaks: 22 06 73 50

Foretaksregisteret: NO 948 007 029 MVA

SINTEF RAPPORT

TITTEL

Prosedyre for evaluering av brukersentrerte aktiviteter

FORFATTER(E)

Asbjørn Følstad

OPPDRAGSGIVER(E)

EFFIN-prosjektet, finansiert av Norges Forskningsråd gjennom FIFOS-programmet

RAPPORTNR. STF90 A05128	GRADERING Åpen	OPPDRAGSGIVERS REF. NFR ref. nr. 153048/530	
GRADER. DENNE SIDE Åpen	ISBN 82-14-03808-1	PROSJEKTNR. 403318.04	ANTALL SIDER OG BILAG 25
ELEKTRONISK ARKIVKODE e/room/EFFIN/WP4/Prosedyre		PROSJEKTLEDER (NAVN, SIGN.) John Krogstie	VERIFISERT AV (NAVN, SIGN.) Jan Heim
ARKIVKODE NA	DATO 2005-07-27	GODKJENT AV (NAVN, STILLING, SIGN.) Bjørn Skjellaug, forskningssjef	

SAMMENDRAG

I prosjektet Efficiency through user involvement (EFFIN) har vi utarbeidet en evalueringsprosedyre for å vurdere kostnads- og nytteeffektiviteten av brukersentrerte aktiviteter i en utviklingsprosess. Evalueringsprosedyren er hovedsakelig orientert mot kvaliteten og påvirkningskraften til resultatet av brukersentrerte kravinnhentings- eller evalueringsaktiviteter.

Brukerne av evalueringsprosedyren tilhører to grupper:

1. Utøvere av brukersentrerte metoder (HCI-praktikere) som ønsker å evaluere egen praksis
2. Forskere og HCI-praktikere som ønsker å undersøke den generelle effektiviteten av en brukersentrert metode eller praksis

Evalueringsprosedyren er prøvd ut i forbindelse med evaluering av metoder for brukersentrert evaluering av mobile IKT-løsninger. Under utprøvingen ble det gjennomført evaluering av tre mobile løsninger: En applikasjon for tilgang til pasientdata til bruk på sengeposter ved sykehus, en applikasjon for parkeringsvakter og en applikasjon for å få tilgang til nyhetsklipp fra et Intranett, bla. tenkt brukt mot Stortingets intranett.

Utprøvingen viste at evalueringsprosedyren gir nyttig innsikt i fordeler og ulemper med ulike evalueringsmetoder i forhold til hverandre. I løpet av EFFIN-prosjektet vil vi også innhente erfaringer med prosedyren brukt som redskap for HCI-praktikeres evaluering av egen brukersentrerte aktivitet. Disse erfaringene vil danne grunnlag for en revidert versjon av metoden.

Prosedyren skal gjøre det mulig å vurdere hvorvidt en brukersentrert tilnærming er hensiktsmessig når man skal utvikle elektroniske publikumstjenester i offentlig sektor.

STIKKORD	NORSK	ENGELSK
GRUPPE 1	IKT	ICT
GRUPPE 2	Menneske-maskin interaksjon	HCI
EGENVALGTE	Evalueringsprosedyre	Evaluation procedure
	Brukersentrerte metoder	User-centred methods

INNHOLDSFORTEGNELSE

1	Innledning	3
1.1	Om kost/nytte	3
1.2	Nytten av brukersentrettede aktiviteter	4
1.3	Hvordan vurdere nytte?	5
2	Rammeverk for en evalueringsprosedyre	7
2.1	Kvaliteten av den brukersentrettede aktivitetens resultat.....	7
2.2	Resultatets påvirkningskraft i den videre utviklingsprosessen.....	9
2.3	Sluttproduktets brukskvalitet.....	10
2.4	Andre effekter av brukersentrettede aktiviteter.....	11
2.5	Kostnader tilknyttet en brukersentrettede aktivitet.....	11
3	Elementer i en omfattende evaluering	12
3.1	Overordnet analyse av resultatet	13
3.2	Umiddelbar vurdering av resultatets kvalitet	13
3.3	Vurdering av resultatets påvirkningskraft – utviklingsteamets prioritering av funnene	15
3.4	Vurdering av resultatets kvalitet etter implementering	16
3.5	Vurdering av sluttproduktets brukskvalitet	17
4	”Quick and clean” evalueringsprosedyre	19
4.1	Eksempel-case: Evalueringsprosedyren brukt på brukersentrettede evaluering av mobile IKT løsninger	19
4.2	Overordnet analyse av resultatet	20
4.3	Vurdere viktighet.....	21
4.4	Erfaringsinnsamling med utviklingsteam.....	21
4.5	Utviklingsteamets prioritering.....	22
4.6	Totalvurdering av den brukersentrettede aktiviteten	23
5	Vurdering og videreføring av evalueringsprosedyren	25

1 Innledning

I EFFIN-prosjektet har vi utarbeidet en evalueringsprosedyre for å vurdere kostnads- og nytteeffektiviteten av å benytte brukersentrerte metoder i en utviklingsprosess. Hensikten med dette er å gjøre det mulig å vurdere hvorvidt en brukersentrert tilnærming er hensiktsmessig når man skal introdusere og utvikle offentlige tjenester i offentlig sektor.

Brukerne av evalueringsprosedyren kan tilhøre to grupper:

1. Utøvere av brukersentrerte metoder (HCI-praktikere) som ønsker å evaluere eget brukersentrert arbeid
2. Metodeutviklere og HCI-praktikere som ønsker å undersøke den generelle effektiviteten av en brukersentrert metode eller praksis

Evalueringsprosedyren er egnet for vurdering av brukersentrert kravinnhenting og evaluering. Prosedyren gjør det bl.a. mulig å vurdere hvorvidt en brukersentrert tilnærming er hensiktsmessig når man skal utvikle elektroniske publikumstjenester i offentlig sektor.

I Kapittel 1 av denne rapporten presenteres bakgrunn for evalueringsprosedyren, gjennom en analyse av hva det vil si at brukersentrerte aktiviteter er nyttige og en diskusjon av hvordan nytte bør undersøkes. Kapittel 2 beskriver et rammeverk for å evaluere brukersentrerte aktiviteter. Med utgangspunkt i rammeverket er det i Kapittel 3 utarbeidet en bred oversikt over elementer som kan inngå i en evaluering av brukersentrerte aktiviteter. Mange av disse elementene er svært ressurskrevende. I Kapittel 4 er derfor de minst kostnadskrevende elementene for datainnsamling satt sammen til en "Quick and clean"-evalueringprosedyre. Prosedyren er særlig egnet for HCI-praktikere som ønsker å evaluere eget brukersentrert arbeid. Vi viser også hvordan "Quick and clean"-prosedyren kan utvides med mer kostnadskrevende datainnsamling.

1.1 Om kost/nytte

Formelen for effektivitet som brukes i kost/nytte analyser er:

$$\text{Effektivitet} = \frac{\text{Nytte}}{\text{Kost}}$$

Kostnad er knyttet til investeringen gjort for å gjennomføre et tiltak. Nytte er knyttet til konsekvensen av tiltaket, og kan beskrives som differansen mellom produktets eller tjenestens verdi med tiltaket og dets verdi uten tiltaket.

Hvor mye brukersentrerte aktiviteter og metoder koster er forholdsvis lett å kalkulere, men *nytt* ("benefit") av brukersentrerte metoder er vanskelig å definere og beregne. Dette skyldes bl.a. at en brukersentrert aktivitet kun inngår som én blant mange andre aktiviteter i et utviklingsprosjekt. Det er vanskelig å vite om en eventuell positiv eller negativ effekt av prosjektet skyldes den enkelte aktiviteten. Nytteundersøkelser på tvers av prosjekter er også problematisk pga. at ulike prosjekter har sin egenart, noe som gjør det svært vanskelig å kontrollere for variasjon mellom prosjektene.

På grunn av utfordringene knyttet til å evaluere nytten av en brukersentrert aktivitet i et utviklingsprosjekt, vil evalueringsprosedyren i overveiende grad være orientert mot å vurdere å måle nytte. Den vil i begrenset grad være orientert mot å vurdere kostnad.

1.2 Nytten av brukersentrerte aktiviteter

I EFFIN-rapporten *Resultatet av litteraturgjennomgang i EFFIN, WP4; "Hva er nytten av brukersentrerte metoder og aktiviteter?"* (Heim et al., 2004) har eksisterende litteratur tilknyttet evaluering av brukersentrerte metoder og praksis blitt gjennomgått. På grunnlag av denne litteraturgjennomgangen foreslås det i denne rapporten at nytten av resultater som fremkommer av brukersentrerte aktiviteter kan vurderes i forhold til følgende dimensjoner:

- Nytten av aktivitetens resultat¹. Avhenger av resultatets:
 - Riktighet
 - Kompletthet
 - Viktighet
 - Påvirkningskraft
- Nytte i forhold til deltagelse i prosessen. Inkluderer:
 - Kompetanseheving
 - Forankring

Nytten av aktivitetens resultat

Resultatet av en brukersentrert aktivitet er typisk et sett "funn". Dette settet av funn kan vurderes med hensyn på nytte. Funn kan være brukerkrav, brukerproblemer eller forslag til redesign, avhengig av hva slags brukersentrert aktivitet det er snakk om. Et brukerkrav er en beskrivelse av en egenskap ved løsningen, der egenskapen vurderes å være nødvendig eller ønskelig for minst en brukergruppe. Et brukerproblem er en beskrivelse av en egenskap ved løsningen som begrunner at egenskapen hindrer minst en brukergruppe å bruke løsningen som forutsatt. Et forslag til redesign er en beskrivelse av hvordan løsningen kan endres, gjerne med referanse til et konkret brukerproblem. Begrepet "løsningen" betyr i denne rapporten produktet/tjenesten som utvikles.

Riktighet (Correctness). Hvor riktige er de funn som gjøres? Dette er en problemstilling som ofte behandles under begrepet validitet. Et funn er i denne sammenhengen "riktig" om det fører til en bedre løsning dersom funnet blir tatt til etterretning.

Kompletthet (Thoroughness). Resultatet man får kan ha større eller mindre grad av kompletthet, dvs. i hvilken grad funnene er dekkende for løsningen man uttaler seg om. Om man kan formulere 10 relevante krav til løsningen kan det være dekkende dersom det er alt som trengs, men om man identifiserer 10 problemer og det faktisk eksisterer 90 i tillegg er funnene man har gjort lite dekkende.

Viktighet (Importance, severity). Viktighet er knyttet til et funns betydning for løsningens brukskvalitet. Et funn har lav viktighet dersom det kun kan føre til en forbedring som er marginal i forhold til et annet funn - som gjør applikasjonen mye bedre. F. eks. vil et brukerproblem klassifisert som "kosmetisk" være mindre viktig enn et brukerproblem klassifisert som "alvorlig" eller "kritisk".

Påvirkningskraft (Impact, persuasive power). Et funn kan ha større eller mindre påvirkningskraft i utviklingsprosessen. Det vil først og fremst vise seg vet at funnet avstedkommer endringer i løsningen som utvikles – eller elementer i en kravspesifikasjon. Her er det mange faktorer som spiller inn. Presis formulering av funnene, funnenes troverdighet og hvor kostnadskrevenende det er å etterleve funnene, er blant de viktigste faktorene.

¹ I Heim et al. sin rapport behandles også "Endringspotensial" som et aspekt av resultatets nytte. Dette aspektet er ikke tatt videre i den foreliggende evalueringsprosedyren, og behandles derfor ikke i dette kapittelet.

Nytte i forhold til deltagelse i prosessen

Kompetanseheving. En brukersentrert prosess kan være nyttig dersom den bidrar til økt forståelse hos utviklingsteamet hva gjelder bruk, brukere og brukskontekst av applikasjonen. En vanlig iaktakelse er at representanter for utviklingsteamet synes å lære mer av å også observere brukere i en brukertest enn bare å lese rapporten med problemlistene i etterkant.

Forankring. Involvering av brukerrepresentanter i enkelte brukersentrerte aktiviteter kan gjøre at disse opplever økt eierskap til løsningen. Dette kan igjen ha stor betydning når et system skal taes i bruk, ved at brukerrepresentanter med økt eierskap kan bidra til åpenhet og endringsvilje i innføringsprosessen.

1.3 Hvordan vurdere nytte?

EFFIN-rapporten *Evaluating user-centred methods: Results from an exploratory study among Nordic HCI-practitioners* (Bark et al., 2005) presenterer resultater fra en survey blant mer enn 200 HCI-praktikere. Rapporten gir innsikt i HCI-praktikernes synspunkter på

- Hvordan brukersentrerte aktiviteter evalueres i dag, og hvordan de ideelt sett burde evalueres
- Hvilke utkommer av brukersentrerte aktiviteter man bør fokusere på i en evaluering av aktivitetene
- Nytte av eksisterende brukersentrerte metoder, og faktisk bruk av de samme metodene.

Evaluering av brukersentrerte aktiviteter

Primært ønsker HCI-praktikere å vurdere nytten av en brukersentrert aktivitet ved bruk av brukertest av neste versjon av løsningen, evalueringsmøter og kundetilfredshetsundersøkelser.

Brukertest av neste versjon innebærer at man gjennomfører den samme brukertesten (med tilsvarende oppgaver og brukere) på Versjon n og Versjon n+1, og deretter sammenligner resultatet av de to evalueringene. Brukertest av neste versjon kan være nyttig for evaluering av evalueringsmetoder, men denne formen for evaluering er kostnadskrevene og vil kun være mulig å gjennomføre i enkelte prosjekter. En svakhet ved brukertest av neste versjon er at det er vanskelig å bestemme hvorvidt brukerproblemer i den nye versjonen er forårsaket av lav kvalitet på resultatet fra den brukersentrerte aktiviteten eller lav kvalitet på redesign/implementering.

Komparativ analyse kan være et godt alternativ til brukertest av neste versjon, for å evaluere kvaliteten av den brukersentrerte aktivitetens resultat. Komparativ analyse innebærer at man gjennomfører to eller flere brukersentrerte aktiviteter som skal gi det samme resultatet, og sammenligner resultatet fra disse. F.eks. ved å gjennomføre både en heuristisk evaluering og en brukertest av den samme løsningen. Fordelen med komparativ analyse i forhold til brukertest av neste versjon, er at man kan være sikker på at forskjeller mellom aktivitetenes resultater skyldes aktiviteten selv. Imidlertid er komparativ analyse, på samme måte som brukertest av neste versjon, kostnadskrevene.

At HCI-praktikere ønsker å vurdere nytte ved hjelp av evalueringsmøter og kundetilfredshetsundersøkelser viser en tro på at evaluering av brukersentrerte aktiviteter kan gjennomføres ved hjelp av post-factum gjennomganger med utviklingsteam og/eller utviklingsprosjektets kunde. Utviklingsprosjektets kunde er den virksomhet som har bestilt utviklingsprosjektet. Denne formen for evaluering åpner muligheten for langt mindre kostnadskrevene – og langt mer fleksible – former for evaluering enn komparativ analyse og brukertest av neste versjon. Med tanke på at evaluering av brukersentrerte aktiviteter bør være praktisk gjennomførbar for flest mulig, er det hensiktsmessig at en evalueringsprosedyre i stor grad hviler på datainnsamling fra utviklingsteam og utviklingsprosjektets kunde.

Fokus for en evalueringsprosedyre

I Bark et al. sin survey-undersøkelse ble HCI-praktikerne spurt om hva det vil være viktigst å fokusere på i en evaluering av brukersentrerte aktiviteter effektivitet. På grunnlag av besvarelsene ble det oppsummert følgende tre fokusområder:

- Resultatets kvalitet
- Resultatets påvirkningskraft i den videre utviklingsprosessen
- Sluttproduktets brukskvalitet (usability)

Det er viktig å merke seg at kvaliteten av resultatet fra en brukersentrert aktivitet bør vurderes atskilt fra sluttproduktets brukskvalitet. Årsaken til dette er at sluttproduktets brukskvalitet bare henger sammen med kvaliteten på resultatet av den brukersentrerte aktiviteten dersom resultatet faktisk har blitt tatt til etterretning i den videre utviklingsprosessen. Det er f.eks. mulig å tenke seg at resultatet fra en brukertest holder svært høy kvalitet, men at rapporten fra brukertesten har blitt ”lagt i skuffen”. Eller at et sluttprodukt kan holde høy brukskvalitet (gjennom tilfeldigheter eller et dyktig utviklingsteam) selv om de brukersentrerte aktivitetene gjennomført i utviklingsprosessen kun har levert resultater av lav kvalitet.

Nytte og bruk av eksisterende brukersentrerte metoder

Deltagerne i survey-undersøkelsen besvarte spørsmål om opplevd nytte og egen bruk av de mest vanlige brukersentrerte metodene. På grunnlag av deres svar er det mulig å si noe om de forhold en evalueringsprosedyre bør prøves ut under. Det er særlig to forhold som er verd å merke seg i forhold til utprøving av en evalueringsprosedyre for brukersentrerte aktiviteter:

- Det er typisk i de senere faser av et utviklingsprosjekt at HCI-praktikere har størst frihet i forhold til valg av brukersentrerte metoder. Det kan følgelig være hensiktsmessig å unngå å prøve ut en evalueringsprosedyre i et utviklingsprosjekts tidligste faser.
- Det er tilsynelatende kun små forskjeller på tvers av ulike typer utviklingsprosjekter i bruk og opplevd nytte av brukersentrerte metoder. Dette innebærer at en evalueringsprosedyre kan være av generell natur, og kan prøves ut i tilknytning til utviklingsprosjekter av så vidt ulike slag som mobile løsninger, web-applikasjoner eller fagsystemer.

2 Rammeverk for en evalueringsprosedyre

En prosedyre for evaluering av brukersentrerte aktiviteter skal hjelpe oss å si noe om:

1. Kvaliteten av aktivitetens resultat
2. Resultatets påvirkningskraft i den videre utviklingsprosessen
3. Sluttproduktets brukskvalitet

I tillegg er det positivt om evalueringsprosedyren kan si noe om andre nyttige effekter av brukersentrerte aktiviteter, som kompetanseheving i utviklingsteamet og økt forankring blant brukere og interessenter.

2.1 Kvaliteten av den brukersentrete aktivitetens resultat

Kvaliteten av den brukersentrerte aktivitetens resultat evalueres i henhold til riktighet, komplettethet og viktighet.

Riktighet og komplettethet

Riktighet og komplettethet er to egenskaper ved resultatets kvalitet som gjensidig utfyller hverandre. En metode med ensidig fokus på riktighet vil gjøre at man gjør få, men sikre funn. En metode med ensidig fokus på komplettethet vil gjøre at man gjør mange funn, med den risiko at en forholdsvis stor andel av disse er feil.

Vurdering av resultatets riktighet innebærer å gi en oversikt over hvilke funn gjort gjennom den brukersentrerte aktiviteten som er riktige, og hvilke som er gale.

Vurdering av resultatets komplettethet vil si å utarbeide en oversikt over funn man (i ettertid) kunne forvente at aktiviteten frembrakte. Oversikten skal vise hvilke funn som faktisk ble frembrakt av den brukersentrerte aktiviteten og hvilke som manglet.

Det er en stor utfordring å undersøke et resultats riktighet og komplettethet.

For å undersøke resultatets riktighet må man frembringe sikker kunnskap om hvorvidt identifiserte brukerkrav eller brukerproblemer faktisk er reelle. Til dette formål kan man vurdere å benytte:

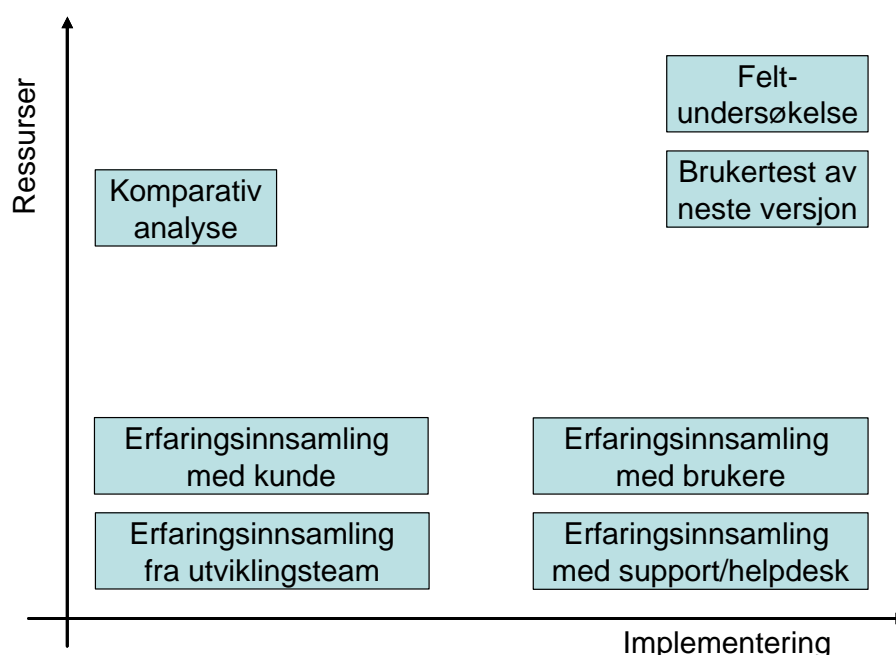
- Feltundersøkelse; der et representativt utvalg reelle brukere observeres i reelle brukssituasjoner
- Erfaringsinnsamling med brukerkontakter (support/helpdesk) eller brukere; der man benytter intervjuer, gruppeintervjuer, eller surveys for å samle kvalitative og kvantitative data om brukeratferd og brukerproblemer.
- Erfaringsinnsamling med utviklingsprosjektets kunde eller utviklingsteamet; der man benytter intervjuer eller gruppeintervjuer for å samle kvalitative og kvantitative data om kunden/utviklingsteamets vurdering av resultatet av den brukersentrerte aktiviteten
- Komparativ analyse; der resultatet fra en brukersentrert aktivitet sammenlignes med resultatet fra en eller flere tilsvarende brukersentrerte aktiviteter
- Brukertest av neste versjon; der man gjennomfører den samme brukertesten på Versjon n og Versjon n+1, og deretter sammenligner resultatet fra de to evalueringene

For å kunne undersøke resultatets komplettethet må man i prinsippet ha en "fasit" som gir oversikt over alle de riktige funnene den brukersentrerte aktiviteten skal frembringe. Ved undersøkelse av komplettethet brukes gjerne komparativ analyse, men også andre av tilnærmingene nevnt i forbindelse med undersøkelse av riktighet kan benyttes.

Valg av tilnærming for å vurdere riktighet og kompletthet avhenger av to forhold:

- *Implementering*. Evaluering gjennom brukertest av neste versjon, feltundersøkelse, eller erfaringsinnsamling med brukere og brukerkontakter (support/helpdesk) krever at funnene som inngår i resultatet som evalueres har blitt tatt til følge ved implementering.
- *Ressurser*. Evaluering gjennom brukertest av neste versjon, feltundersøkelse, eller komparativ analyse krever at man har tilstrekkelig ressurser til å gjennomføre disse.

I Figur 1 er ulike tilnærminger til å vurdere riktighet angitt i forhold til deres krav til implementering og ressurser. Vi ser at for å gjøre en evalueringsprosedyre mest mulig uavhengig av ressurser og implementering, vil det være fornuftig å basere prosedyren på informasjonsinnhenting fra utviklingsteam eller utviklingsprosjektets kunde. Imidlertid vil evaluering ved hjelp av de andre tilnærmingene være sikrere.



Figur 1: Ulike tilnærminger til å evaluere riktigheten og kompletthet av resultatet fra en brukersentrert aktivitet

Viktighet

Resultat fra brukersentrerte aktiviteter vil som regel leveres med en klassifisering av viktighet for hvert enkelt funn. Vurdering av resultatets viktighet vil si å gi en oppsummering av funnenes klassifisering. I tillegg kan det være aktuelt å etterprøve klassifiseringen som er gitt ved å gjøre gjentatt klassifisering.

Viktighet for funn i en brukersentrert evaluering kan klassifiseres på følgende måte:

- *Kritisk*. Vil sannsynligvis stoppe typiske brukere i å bruke løsningen til å gjennomføre oppgaven
- *Alvorlig*. Vil sannsynligvis medføre at typiske brukere opplever vesentlig forsinkelse ved gjennomføring av oppgaven, og at enkelte brukere stoppes fra å bruke løsningen til å gjennomføre oppgaven
- *Kosmetisk*. Vil sannsynligvis føre til at typiske brukere opplever en liten forsinkelse ved gjennomføring av oppgaven.

Funn fra en brukersentrert kravinnhenting kan klassifiseres som:

- *Skal*. Løsningen skal innfri kravet
- *Bør*. Løsningen bør innfri kravet med mindre tungtveiende grunner taler mot dette
- *Ønskelig*. Det er ønskelig, men ikke påkrevd, at løsningen innfrir kravet

Resultatets viktighet kan angis som en oppsummering av funnenes klassifisering. En slik oppsummering gir tilbakemelding på hvor viktig det var å gjennomføre aktiviteten.

Dersom funnene gjort i en aktivitet i overveiende grad vurderes som mindre viktige, er det godt mulig at det ikke er hensiktsmessig å gjennomføre aktiviteten (med mindre aktiviteten kun skulle tjene som en kvalitetskontroll). Aktiviteter som i overveiende grad frembringer funn klassifisert som ”kosmetisk” eller ”ønskelig” bør sannsynligvis endres.

Hvis man sammenligner to aktiviteter med samme formål, kan antall funn med høy viktighet og prosentvis fordeling av klassifisering si mye om hvilken aktivitet som er den viktigste – og følgelig også den nyttigste.

Hvis man har mulighet, er det nyttig å etterprøve klassifiseringen av funnene fra en brukersentrert aktivitet. Helst som en gjentatt klassifisering, der den/de som gjør den nye klassifiseringen ikke kjenner den opprinnelige klassifiseringen. Gjentatt klassifisering kan gjøres av andre HCI-praktikere. Eventuelt av brukerrepresentanter, representanter for førstelinje/brukerkontakt, utviklingsprosjektets kunde, eller utviklingsteam.

2.2 Resultatets påvirkningskraft i den videre utviklingsprosessen

Resultatets påvirkningskraft i den videre utviklingsprosessen er av stor betydning for nytten av en brukersentrert aktivitet. Selv et resultat av høy kvalitet blir ikke nyttig før det avstedkommer endringer i spesifisering, design eller implementering. Skal man ha håp om å påvirke brukskvaliteten til løsningen som utvikles, er man nødt til å vektlegge påvirkningskraften for resultatet av den brukersentrerte aktiviteten.

Et god måte å undersøke resultatets påvirkningskraft på, er å samle inn data om hvordan funnene fra den brukersentrerte aktiviteten blir prioritert i det videre utviklingsløpet. Slik prioritering innhentes enklest i samarbeid med lederen for utviklingsteamet, og kan gjøres på forhånd eller post-factum. Prioriteringer gjort på forhånd gir mulighet for umiddelbar evaluering av den brukersentrerte aktiviteten. Imidlertid vil prioriteringer angitt post-factum være mer troverdige.

Det er nyttig å skille mellom funn gitt lav, middels og høy prioritet. Prioriteringer av funn fra brukersentrerte aktiviteter må gjøres noe forskjellig i utviklingsprosjektets spesifikasjonsfase og i konstruksjonsfase. Detaljer presenteres i Tabell 1.

Prioritet	Spesifikasjonsfase	Konstruksjonsfase
Høy	Resultatet ble tatt til etterretning i spesifikasjonen	Endring er gjennomført eller planlagt nærmeste X måneder
Middels	Resultatet ble tatt til etterretning i spesifikasjonen i endret/bearbeidet form	Endring kan være aktuelt, men ikke de nærmeste X månedene
Lav	Resultatet ble ikke tatt til etterretning i spesifikasjonen	Endring vil ikke prioriteres i overskuelig fremtid

Tabell 1: Mulige prioriteringer for funn i spesifikasjonsfase og konstruksjonsfase

Oppsummering av påvirkningskraften for en brukersentrert aktivitet kan gjøres ved å gi en oversikt over antall og prosentvis fordeling av funn mellom de ulike prioritetskategoriene. Brukersentrerte aktiviteter med lav andel høyt prioriterte funn bør endres. Ved sammenligning av to tilsvarende brukersentrerte aktiviteter vil sannsynligvis den aktiviteten med høyest påvirkningskraft være den nyttigste (forutsatt at den andre aktiviteten ikke genererer funn av vesentlig høyere kvalitet).

Det bør imidlertid påpekes at verdien av å gjennomføre en brukersentrert aktivitet med høy påvirkningskraft er avhengig av kvaliteten av funnene som aktiviteten genererer. En brukersentrert aktivitet med høy påvirkningskraft som feilaktig identifiserer brukerproblemer (dvs. identifiserer brukerproblemer som ikke er reelle), vil kunne medføre kostnadskrevende redesign uten at dette er påkrevd. Det er derfor viktig å huske på at påvirkningskraft er en nødvendig, men ikke tilstrekkelig betingelse for at brukersentrerte aktiviteter er nyttige.

2.3 Sluttproduktets brukskvalitet

Gitt at resultatet fra en brukersentrert aktivitet holder høy kvalitet, og resultatet har god påvirkningskraft vil dette bidra til en økning av sluttproduktets brukskvalitet. Det kan likevel være nyttig å undersøke sluttproduktets brukskvalitet i tillegg til å undersøke resultatets kvalitet og påvirkningskraft. Dette for å finne ut hvorvidt man har hatt riktig mengde og riktige former for brukersentrerte aktiviteter i utviklingsperioden.

Sluttproduktets brukskvalitet kan evalueres i forhold til følgende dimensjoner:

- *Måloppnåelse*. I hvilken grad klarer brukerne å løse sine oppgaver med løsningen
- *Effektivitet*. I hvilken grad klarer brukerne å løse sine oppgaver på en effektiv måte
- *Tilfredshet*. I hvilken grad er brukerne fornøyd med løsningen

Sluttproduktets brukskvalitet vurderes gjerne ved hjelp av brukertester eller feltundersøkelser. Feltundersøkelser gir den mest troverdige undersøkelsen, men er også mest kostnadskrevende. Mindre kostnadskrevende tilnæringer inkluderer datainnsamling blant brukere, utviklingsprosjektets kunde eller brukerkontakter (support/helpdesk), samt evaluering ved hjelp av analytiske evalueringsmetoder som heuristisk evaluering eller ”cognitive walkthrough”.

Utfordringen tilknyttet en vurdering av sluttproduktets brukskvalitet, er at det er vanskelig å si i hvilken grad brukskvaliteten skyldes de gjennomførte brukersentrerte aktivitetene eller andre forhold. F.eks. kan et system ha blitt utviklet i henhold til anbefalinger fra en HCI-praktiker, men likevel få lav brukskvalitet på grunn av et implementeringsproblem som gir uakseptabelt høy responstid.

Det er også andre utfordringer knyttet til å si noe om nytten av brukersentrerte aktiviteter, bare ved å vurdere sluttproduktets brukskvalitet. For å kunne vurdere nytten av en brukersentrert aktivitet bør vi i prinsippet vite hvordan brukskvaliteten av det endelige løsningen ville blitt *uten* at aktiviteten hadde blitt gjennomført. Dette er vanligvis ikke mulig.

Imidlertid vil det gjerne være slik at dersom sluttproduktet holder høy brukskvalitet vil den involverte HCI-praktikeren høste æren for dette. Og dersom sluttproduktet holder lav brukskvalitet vil HCI-praktikeren tilsvarende få skylden. Det er derfor verdifullt for HCI-praktikeren å skaffe seg sikrest mulig kunnskap om brukskvaliteten til systemer hun har vært med å utvikle.

2.4 Andre effekter av brukersentretre aktiviteter

Det finnes andre mulige effekter av brukersentrerte aktiviteter, som kompetanseheving og forankring. Disse to vil kort gjennomgås nedenfor, men vil ikke tas videre som del av evalueringsprosedyren for å begrense prosedyrens kompleksitet.

Kompetanseheving

Brukersentrerte aktiviteter kan bidra til kompetanseheving i utviklingsteamet tilknyttet brukerinvolvering og brukskvalitet. Dette gjelder særlig utviklingsteam som tidligere ikke har vært eksponert for denne måten å arbeide på. Økt kompetanseheving kan vurderes uformelt gjennom intervjuer eller oppsummeringsmøter med deltagerne i utviklingsteamet.

Dersom man ønsker et mer formelt mål på kompetanseheving kan man vurdere å gjøre en før-etter-vurdering av utviklingsteamet i forhold til såkalte "usability-maturity-models". Se f.eks. *Earthy, J (1998) "Usability Maturity Model: Human Centredness Scale" Rapport fra Telematics applications Project IE 2016.*

Forankring

Brukersentrerte aktiviteter kan bidra til økt forankring av løsningen blant brukergrupper og interessenter. Særlig i den grad aktivitetene direkte involverer brukere og interessenter som er valgt ut iht. demokratiske representasjon (f.eks. ansattrepresentanter eller representanter fra interesseorganisasjoner), eller iht. deres egenskap som beslutningstagere. Økt forankring kan vurderes uformelt på grunnlag av intervjuer eller oppsummeringsmøter med utviklingsprosjektets kunde, brukere, eller utviklingsteam.

2.5 Kostnader tilknyttet en brukersentrert aktivitet

Som nevnt innledningsvis er det forholdsvis lett å kalkulere hvor mye brukersentrerte aktiviteter og metoder koster. Dette vil derfor kun oppsummeres på et overordnet nivå.

Kostnader tilknyttet brukersentrerte aktiviteter kan tilhøre følgende kategorier:

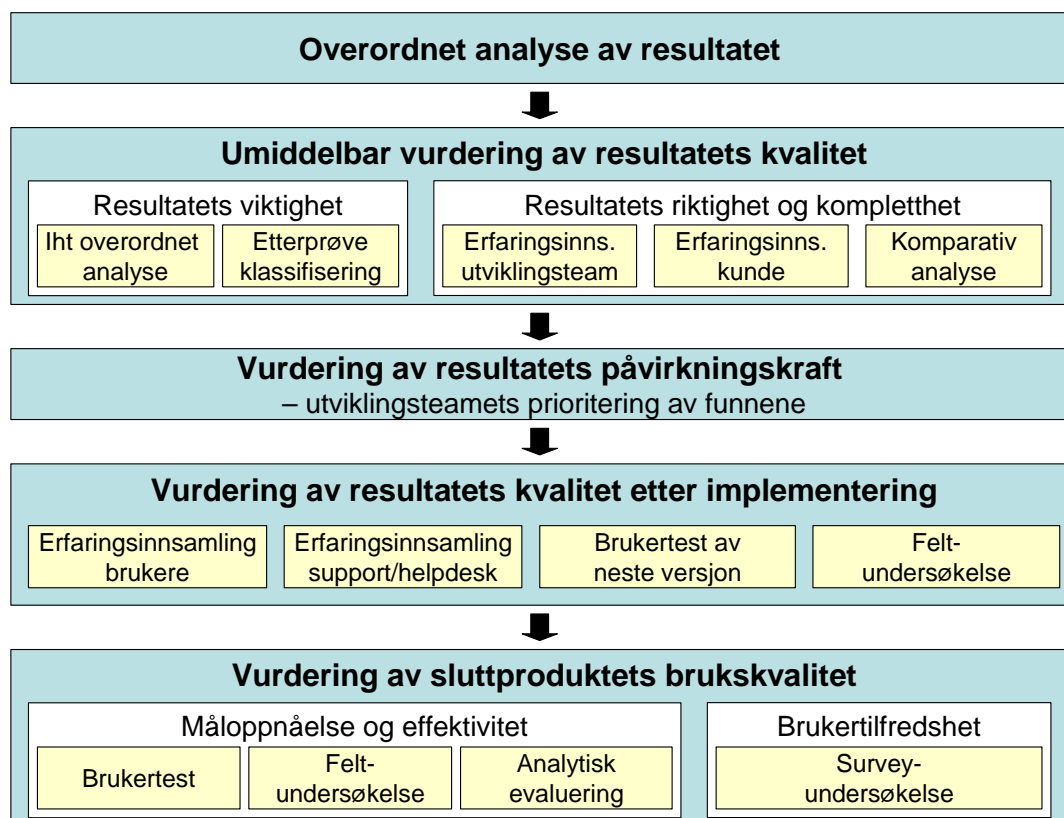
- Kostnaden for den eller de som gjennomfører aktiviteten. Kan være i form av arbeidstid (ved bruk av eget personell) eller direktekostnad (ved kjøp av arbeidstid).
- Direktekostnader tilknyttet gjennomføring. Inkluderer evt. honorering av eksterne brukere, transport, leie av rom o.l.
- Indirekte kostnader tilknyttet gjennomføring. Inkluderer evt. tidsbruk for interne brukere og interessenter.

3 Elementer i en omfattende evaluering

Med utgangspunkt i rammeverket beskrives i dette kapittelet en omfattende oversikt over elementer som kan inngå i en evaluering av brukersentrerte aktiviteter. Enkelte av elementene er svært kostnadskrevende, mens andre stiller små krav til ressurser. Kostnadskrevende elementer er tatt med fordi disse gjør det mulig å få sikrere kunnskap om et bestemt fokusområde enn tilsvarende elementer med mindre ressurskrav.

Elementene inngår i følgende struktur – presentert i Figur 2:

1. Overordnet analyse av resultatet
2. Umiddelbar vurdering av resultatets kvalitet
3. Vurdering av resultatets påvirkningskraft
4. Vurdering av resultatets kvalitet etter implementering
5. Vurdering av sluttproduktets brukskvalitet



Figur 2: Elementer i en omfattende evaluering av brukersentrerte aktiviteter

I Kapittel 4 er de minst kostnadskrevende elementene for datainnsamling satt sammen til en "Quick and clean" evalueringsprosedyre. Denne er særlig egnet for utøvere av brukersentrerte aktiviteter som ønsker å evaluere eget arbeid. Denne basisprosedyren kan utvides med mer kostnadskrevende datainnsamling. Dette vil særlig være aktuelt for metodeutviklere og HCI-praktikere som skal ta overordnede beslutninger om hvilke metoder man ønsker å satse på – f.eks. i forbindelse med intern kompetanseheving eller utforming av konsulent tjenester. I denne rapporten vil det ikke beskrives i detalj hvordan basisprosedyren kan utvides med ulike elementer, men elementene er beskrevet i avsnittene nedenfor.

3.1 Overordnet analyse av resultatet

For å kunne gjennomføre en hensiktsmessig evaluering av en brukersentrert aktivitet, stilles følgende krav til aktivitetens resultatrapportering:

- Resultatet må være presentert som et sett funn
- Funnene må være unike og ikke overlappende
- Funnene må være klassifisert i henhold til viktighet

Det sentrale elementet i evalueringen av den brukersentrerte aktiviteten, er en overordnet analyse av aktivitetens resultat. Den overordnede analyse er det første man utarbeider i evalueringen; gjerne som del av rapporteringen fra den brukersentrerte aktiviteten. I analysen av fremstilles:

- Antall funn (identifiserte brukerkrav eller brukerproblemer)
- Fordeling av funn i henhold til klassifisering

Analysen kan gjerne presenteres som en tabell – se eksempel i Kapittel 4.

3.2 Umiddelbar vurdering av resultatets kvalitet

Vurdering av resultatets kvalitet innebærer vurdering av viktighet og riktighet/komplettethet.

Resultatets viktighet

Vurdering basert på overordnet analyse

Resultatets viktighet vil i stor grad bestemme hvor nyttig den brukersentrerte aktiviteten har vært. Dette gjøres enklest med referanse til fordelingen av funn i henhold til klassifisering (utarbeidet i den overordnede analysen – Kapittel 3.1). Viktighet kan vurderes etter følgende enkle tommelfingerregler:

- Dersom den brukersentrerte aktiviteten har frembrakt få eller ingen funn med høyeste klassifisering, har den neppe vært videre nyttig (annet enn som kvalitetssikring av en ferdig løsning)
- Dersom brukersentrerte aktiviteten har frembrakt et høyt antall funn med laveste klassifisering, har man antagelig brukt tiden på uvesentlige forhold

Etterprøving av klassifisering

Etterprøving av funnenes viktighets-klassifisering er nyttig, for å motvirke eventuell subjektiv ”bias” knyttet til den opprinnelige klassifiseringen. Etterprøving av klassifiseringen kan gjerne inngå som del av resultatet fra den brukersentrerte aktiviteten.

Etterprøving bør gjøres ”blindt” – dvs. at det gjøres en ny klassifisering, der den som gjennomfører klassifiseringen ikke kjenner det eksisterende klassifiseringsresultatet. Etterprøving kan gjøres ved at en eller flere personer som gjør en ny klassifisering individuelt, for så å samstemme sine klassifiseringer gjennom en kort gruppeprosess. Den omforente klassifiseringen analyseres i forhold til den opprinnelige klassifiseringen.

Resultatet av en etterprøving kan uttrykkes som:

- Andel funn med lik klassifisering av viktighet
- Andel funn med noe avvikende klassifisering av viktighet (f.eks. med kun ett trinns forskjell mellom ny og eksisterende klassifisering)
- Andel funn med svært avvikende klassifisering av viktighet

Utvelgelse av personene som gjennomfører etterprøvingen av resultatet vil være avhengig av sammenhengen den brukersentrerte aktiviteten utføres i. Personene kan være representanter for utviklingsprosjektets kunde, utviklingsteam og en eller flere brukergrupper. Eller de kan være personer med HCI-kompetanse.

Resultatets riktighet og kompletthet

Erfaringsinnsamling med utviklingsteam

På grunn av at brukersentrerte aktiviteter typisk skjer i dialog med utviklingsteamet, er erfaringsinnsamlinger med en eller flere representant(er) fra utviklingsteamet en lite kostnadskreven måte å få tilbakemelding på resultatets riktighet og kompletthet. Erfaringsinnsamlingen gjennomføres som halvstrukturert intervju eller gruppeintervju. Det er imidlertid viktig å huske på at når det gjelder resultatets riktighet og kompletthet er denne formen for tilbakemelding langt mindre sikker enn f.eks. komparativ analyse, brukertest av neste versjon, eller feltundersøkelse.

I de halvstrukturerte intervjuerne fokuseres det på følgende tema:

- Riktighet
 - Opplevs funnene fra den brukersentrerte aktiviteten som rimelige
 - Samsvar mellom resultatet av den brukersentrerte aktiviteten og resultater fra andre aktiviteter som inngår i utviklingen og utprøvingen av løsningen
- Kompletthet
 - Ble alle allerede kjente brukerkrav/brukerproblemer identifisert gjennom den brukersentrerte aktiviteten?

Hvis ønskelig kan det i intervjuet evt. også fokuseres på betydningen av den brukersentrerte aktiviteten for *kompetanseheving i utviklingsteamet* og *forankring av løsningen blant brukergrupper og interessenter*.

Erfaringsinnsamling med utviklingsprosjektets kunde

Om man har anledning til å legge noe mer ressurser i evalueringsprosessen kan det være hensiktsmessig å gjennomføre en erfaringsinnsamling med utviklingsprosjektets kunde på samme måte som med utviklingsteamet. Erfaringsinnsamling med kunde forutsetter at dette er akseptabelt for utviklingsbedriften og kan innebære noe mer kostnad enn erfaringsinnsamling med utviklingsteam, dersom man ikke har umiddelbar tilgang til kundens representanter. Erfaringsinnsamling med utviklingsprosjektets kunde gjennomføres på samme måte som erfaringsinnsamling med utviklingsteam.

I noen tilfeller er den ansvarlige for brukersentrerte aktiviteter engasjert av utviklingsprosjektets kunde, ikke utviklingsbedriften. I så fall vil erfaringsinnsamling med kunden være den minst kostnadskreven aktiviteten, mens aktiviteten med utviklingsteamet vil kreve noe mer ressurser.

Komparativ analyse

Komparativ analyse innebærer at man undersøker resultatet oppnådd med en metode i forhold til resultat oppnådd med en eller flere andre metoder. Komparativ analyse krever innsamling av samme type data gjennom flere metoder kan derfor være kostnadskreven. Denne typen analyse er derfor ofte kun aktuell i tilknytning til metodeutvikling og mer omfattende metodeevaluering. I enkelte utviklingsprosesser vil det imidlertid være aktuelt å innhente samme type data gjennom ulike metoder, noe som kan gjøre det mulig for den enkelte HCI-praktiker å benytte komparativ analyse for å vurdere egen praksis. Fokus i en komparativ analyse er primært resultatets riktighet og kompletthet. Kvaliteten av konklusjoner som bygger på en komparativ analyse er avhengige kvaliteten på resultatet fra den andre metoden/de andre metodene som brukes som sammenligningsgrunnlag i analysen.

For å gjøre en komparativ analyse, gjennomføres først en overordnet analyse av resultatet fra hver enkelt metode (se "Overordnet analyse" over). Deretter utarbeider man for en oversikt over følgende:

- Antall/andel funn med den brukersentrerte aktiviteten som evalueres, som er felles med resultatet fra den/de andre metodene.
- Antall/andel funn med den brukersentrerte aktiviteten som evalueres, som er i motstrid med resultatet fra den/de andre metodene.
- Andel funn med den brukersentrerte aktiviteten som evalueres, i forhold til den totale mengden funn.

For hver av de tre punktene gis det oversikt både for alle funn under ett, samt gruppert i henhold til klassifisering av viktighet.

Tommelfingerregler til bruk ved komparativ analyse:

- Høy andel felles funn mellom metodene indikerer at resultatet er riktige og komplette. Det har særlig betydning at det er høy andel av felles funn med høy viktighet
- Lav andel funn i motstrid med funn fra andre metoder indikerer at resultatet er riktig. Ved motstrid mellom funn bør det gjennomføres en kvalitativ vurdering for å belyse hvilken av metodene som sannsynligvis har generert det riktige funnet.

Dersom man kun har anledning til å sammenligne metoden man bruker med én annen metode, bør man være sikker på at metoden man sammenligner med gir resultat av høy kvalitet. I litteraturen er det vanlig å regne resultat fra brukertester som av forholdsvis høy kvalitet; dette regnes derfor som et egnet sammenligningsgrunnlag. Resultat fra feltundersøkelser – der løsningen evalueres i den sammenheng den er utviklet for – kan i prinsippet holde enda høyere kvalitet enn brukertester, men feltundersøkelser av høy kvalitet kan være vanskeligere å gjennomføre (typisk pga. vansker og ressurskrav knyttet til å gjennomføre representativ observasjon av brukstilfeller).

Komparativ analyse må gjøres med forsiktighet og respekt for de utfordringene dette innebærer. Det er viktig å være klar over at en endelig vurdering av riktighet og kompletthet i de fleste tilfeller ikke er mulig. Det vil derfor som regel være vanskelig å si med sikkerhet hvorvidt resultatet fra en brukersentrert aktivitet er riktigere eller mer komplette enn resultatet av en annen. To metoder kan f.eks. gi riktige resultater uten at resultatene er identiske – man får forskjellig informasjon om samme løsning (dette skjer typisk når man sammenligner resultater fra heuristisk evaluering og brukertest). Det er også viktig å være oppmerksom på at det ofte kan forekomme stor variasjon i funn gjort med flere brukersentrerte metoder. Det samme gjelder for samme metode brukt flere ganger av ulike personer. (Se f.eks. *Molich et al., 2004, "Comparative usability evaluation", Behaviour & Information Technology, 23 (1), s. 65-74*). Årsaken til denne typen variasjon kan være at brukersentrerte metoder ofte genererer et høyt antall funn med lav viktighets-klassifisering.

3.3 Vurdering av resultatets påvirkningskraft – utviklingsteamets prioritering av funnene

Resultatets påvirkningskraft vurderes på grunnlag utviklingsteamets prioritering av funnene. Denne formen for prioriteringsøvelse vil også avdekke utviklingsteamets vurdering av funnenes riktighet.

Prioritering av funnene kan gjøres umiddelbart etter at rapporten fra den brukersentrerte aktiviteten er levert. Den kan også gjennomføres når neste versjon av løsningen er under utvikling eller ferdigstilt. Jo nærmere ferdigstilling av neste versjon man kommer, jo sikrere vil dataene som er samlet inn på denne måten være.

Innsamling av utviklingsteamets prioritering foregår ved at en testleder og en representant for utviklingsteamet går serielt gjennom alle identifiserte brukerkrav eller brukerproblemer i rapporten. Man skal ikke ta hensyn til de identifiserte kravene/problemenes klassifisering. Hvert krav/problem prioriteres av utviklingsteamets representant som én av følgende:

1. *Høy.* (Se Seksjon 2.2)
2. *Middels.* (Se Seksjon 2.2)
3. *Lav.* (Se Seksjon 2.2)
4. *Feil.* Det aktuelle problemet/forslaget vurderes av utviklingsteamet til ikke å innebære riktighet

Som et forslag til tommelfingerregel kan man gjøre følgende vurdering av påvirkningskraft på grunnlag av en slik prioritering:

- Mer enn halvparten av funnene gitt lav prioritert eller vurdert som feil indikerer lav påvirkningskraft
- Mer enn halvparten av funnene gitt høy prioritet eller vurdert som feil indikerer høy påvirkningskraft.

3.4 Vurdering av resultatets kvalitet etter implementering

Erfaringsinnsamling med brukere og brukerkontakter (support/helpdesk)

Som supplement til erfaringsinnsamling med utviklingsprosjektets kunde eller utviklingsteam kan det gjennomføres erfaringsinnsamling med representanter for brukere eller brukerkontakter.

Hensikten med denne erfaringsinnsamlingen er å få enkel, direkte tilgang til brukerproblemer tilknyttet neste versjon av løsningen (redesignet på grunnlag av resultatet fra den brukersentrerte aktiviteten). Svakheten ved denne formen for datainnsamling kan være at man ikke får verifisert det informantene sier i forhold til typisk brukeratferd.

Erfaringsinnsamlingen gjennomføres etter at neste versjon av løsningen er implementert. Datainnsamlingen foregår ved halvstrukturert intervju eller gruppeintervju, med en eller flere representanter for aktuelle brukergrupper eller brukerkontakter (support/helpdesk).

Tilbakemelding med relevans for resultatets riktighet og kompletthet er knyttet til:

- Brukerproblemer tilknyttet eksisterende funksjonalitet i løsningen
- Brukerproblemer tilknyttet manglende funksjonalitet i løsningen
- Brukerproblemer tilknyttet integrering av løsningen i brukerens kontekst

Brukertest av neste versjon

Brukertest av neste versjon innebærer at man undersøker resultatet fra en brukersentrert aktivitet i forhold til brukskvaliteten av neste versjon. Denne formen for evaluering er nyttig for å evaluere kvaliteten av design-innspill gjort på grunnlag av brukersentrerte aktiviteter. For å kunne evaluere kvaliteten på brukerkrav og brukerproblemer ved brukertest av neste versjon må man klare å skille mellom kvaliteten på krav/problemer og kvaliteten på design/implementering (f.eks. ved å undersøke i hvilken grad design/implementering er i henhold til føringene i de identifiserte kravene/problemene).

Evaluering av riktigheten av resultatet fra en brukersentrert aktivitet kan gjøres på grunnlag av en oppsummering av:

- Andelen funn som fjernet brukerproblemer (gruppert i henhold til viktighetsklassifisering)
- Andelen funn som ledet til nye brukerproblemer (gruppert i henhold til viktighetsklassifisering for funnet både før og etter redesign)
- Andel funn som ikke ble tatt til etterretning i den videre utviklingen

Evaluering av komplettheten av resultatet fra en brukersentrert aktivitet kan gjøres ved å oppsummere brukerproblemer som ikke kan knyttes til innspill fra forrige brukersentrerte aktivitet.

Feltundersøkelse av kjørende versjon

Feltundersøkelse av kjørende versjon innebærer at man undersøker resultatet fra en brukersentrert aktivitet i forhold til brukskvaliteten av en kjørende versjon.

Denne formen for evaluering gir et sikrere resultat enn brukertester, fordi den gjør det mulig å ta i betraktning alle forhold knyttet til løsningens brukskvalitet, ikke bare de som avdekkes i en kunstig evalueringssituasjon. Imidlertid er feltundersøkelser gjerne kostnadskreven, og kan også være vanskelig å gjennomføre i praksis.

Som for brukertesten, er evaluering av kvaliteten på brukerkrav og brukerproblemer avhengig av at man klare å skille mellom kvaliteten på krav/problemer og kvaliteten på design/ implementering.

Evaluering av riktighet og kompletthet gjøres på samme måte som for brukertest av neste versjon (se "Brukertest av neste versjon" over)

3.5 Vurdering av sluttproduktets brukskvalitet

Sluttproduktets brukskvalitet vurderes gjerne i forhold til faktorene måloppnåelse, effektivitet og tilfredshet. Dersom et utviklingsløp har inneholdt de riktige brukersentrerte aktivitetene, og resultatene fra aktivitetene har blitt tatt til etterretning gjennom utviklingsløpet, kan man forvente at sluttproduktet holder høy brukskvalitet.

For å evaluere egen praksis vil det for en HCI-praktiker antagelig være nyttigere å undersøke kvaliteten og påvirkningskraften til resultatene av egne brukersentrerte aktiviteter, snarere enn å undersøke brukskvaliteten til sluttproduktet. Særlig fordi man da unngår problemet med å differensiere mellom resultater fra de brukersentrerte aktivitetene og andre aktiviteter i utviklingsløpet.

Ikke desto mindre er det viktig for HCI-praktikeren å skaffe seg informasjon om sluttproduktets brukskvalitet, siden dette er den overordnede motivasjonen for de fleste brukersentrerte aktiviteter. I tillegg kan informasjon om sluttproduktets brukskvalitet utgjøre et nyttig korrektiv informasjon om kvaliteten og påvirkningskraften til resultatene av de brukersentrerte aktivitetene.

Vurdering av måloppnåelse og effektivitet

Faktorene måloppnåelse og effektivitet refererer til hvorvidt de identifiserte brukergruppene til løsningen klarer å løse de oppgavene de skal (måloppnåelse) uten overdreven bruk av tid eller andre ressurser (effektivitet). Forhold i brukergrensesnittet som medfører redusert måloppnåelse eller effektivitet betegnes brukerproblemer.

Måloppnåelse og effektivitet vurderes gjerne ved hjelp av feltundersøkelser, brukertester, eller ulike former for analytiske evalueringsmetoder. Det vil føre for langt å presentere disse i denne rapporten, særlig siden metoder for vurdering av sluttproduktets brukskvalitet som regel krever store ekstra ressurser, som gjør dem uaktuelle for HCI-praktikeren som ønsker å evaluere egen brukersentrerte praksis.

Ved en evaluering av sluttproduktets brukskvalitet bør det, som for erfaringsinnsamling med brukere og brukerkontakter, inngå en oppsummering av følgende:

- Brukerproblemer tilknyttet eksisterende funksjonalitet i løsningen
- Brukerproblemer tilknyttet manglende funksjonalitet i løsningen
- Brukerproblemer tilknyttet integrering av løsningen i brukerens kontekst

Brukerproblemene bør grupperes i henhold til klassifisering av viktighet. Løsningen bør ikke inneholde noen kritiske brukerproblemer, og kun få alvorlige brukerproblemer. Særlig gjelder dette i den grad eventuelle brukerproblemer er tilknyttet sentral funksjonalitet.

Vurdering av tilfredshet

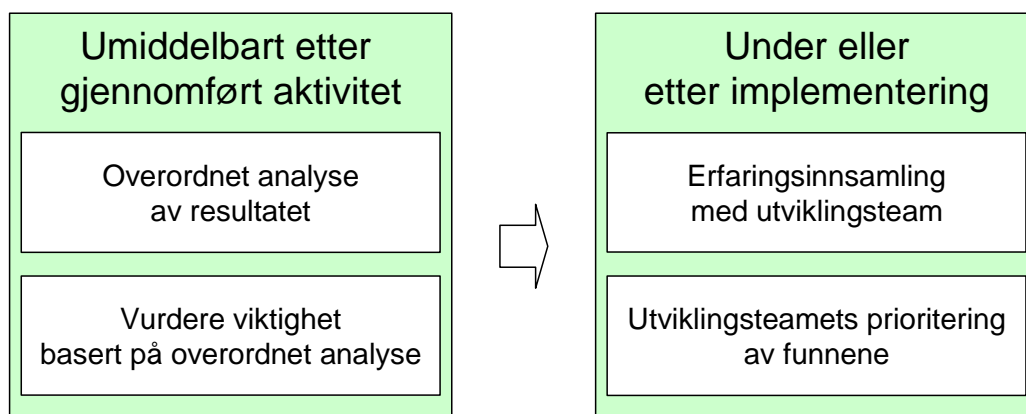
Brukertilfredshet vurderes som regel gjennom survey-undersøkelser, der respondentene utgjør et representativt utvalg av en eller flere viktige brukergrupper for løsningen. Det finnes enkelte standardiserte spørreskjema på markedet. Det mest kjente av disse er WAMMI (se <http://www.wammi.com>). Det er også vanlig å benytte seg av enkle egenutviklede spørreskjema til dette formålet. Måling av kundetilfredshet ved et enkelt målepunkt er mest hensiktsmessig dersom man har tilgang til en "baseline" å se sitt eget resultat i forhold til. Gjentatte målinger med det samme spørreskjemaet er nyttig for å se trender i brukertilfredshet.

Måling av brukertilfredshet kan være interessante, men det er vanskelig å knytte disse slike målinger til en vurdering av brukersentrerte aktiviteter i utviklingsløpet. Ikke desto mindre kan brukertilfredshetsundersøkelser være nyttige som korrektiv til konklusjoner trukket på grunnlag av andre undersøkelser.

4 ”Quick and clean” evalueringsprosedyre

I Kapittel 3 er det beskrevet en rekke elementer som kan inngå i en prosedyre for evaluering av brukersentrerte aktiviteter. Kostnaden tilknyttet de enkelte elementene er svært varierende. I avsnittene nedenfor presenteres en basisprosedyre som utelukkende består av de minst kostnadskrevene elementene fra Kapittel 3. Basisprosedyren er særlig rettet mot HCI-praktikere som ønsker en evaluering av egen praksis. Figur 3 gir en grafisk oversikt over rekkefølgen og elementene i evalueringsprosedyren.

Det er særlig **resultatets påvirkningskraft** som det er mulig å gi en sikker vurdering av uten bruk av vesentlige ressurser. Sikker kunnskap om **resultatets kvalitet** og **sluttproduktets brukskvalitet** krever mer ressurser. Basisprosedyren gir derfor sikrest kunnskap om påvirkningskraften til den brukersentrerte aktiviteten som evalueres.



Figur 3: ”Quick and clean” evalueringsprosedyre

Til bruk for den enkeltstående HCI-praktiker antas prosedyren å være enkel og tilstrekkelig sikker, derav betegnelsen ”Quick and clean”. Prosedyren antas også å være tilstrekkelig sikker for omfattende metodeevaluering med fokus på undersøkelse av resultatets påvirkningskraft. Dersom man ønsker å gjøre omfattende metodeevaluering med fokus på resultatets riktighet og kompletthet, er **ikke** prosedyren beskrevet i dette kapittelet tilstrekkelig; den må suppleres med andre elementer beskrevet i Kapittel 3.

Evalueringsprosedyren er tidligere brukt for evaluering av en evalueringsmetode utarbeidet med utgangspunkt i ”cognitive walkthrough”. Denne evalueringen vil brukes som eksempel-case gjennom presentasjonen av evalueringsprosedyren.

4.1 Eksempel-case: Evalueringsprosedyren brukt på brukersentrert evaluering av mobile IKT løsninger

Det har blitt gjennomført en evaluering av nytten av en analytisk evalueringsmetode for strukturert gjennomgang av brukergrensesnitt med grupper av eksperter. Evalueringen ble gjennomført ved hjelp av evalueringsprosessen utarbeidet i EFFIN (beskrevet i Figur 3).

Eksempel-casets problemstilling: Evalueringen ble gjennomført for å undersøke hvorvidt det gir mest nytte å bruke domene-eksperter eller usability-eksperter som evaluatorene ved analytisk evaluering av mobile IKT-løsninger.

Nytte kan, som det fremgår av Kapittel 2, forstås både i forhold til (1) resultatets kvalitet og (2) påvirkningskraft i form av resultatets utnyttelse i det videre utviklingsløpet. I eksempel-caset ble det i hovedsak fokusert på nytte som utnyttelse av resultatet i den videre utviklingsprosessen.

Analytisk evaluering er samlebetegnelsen på en rekke metoder for evaluering av løsnings brukergrensesnitt, brukt innen fagområdet HCI. En analytisk evaluering foregår ved at en eller flere evaluatorene går gjennom løsningen som skal evalueres, i henhold til en fastlagt prosess. Dette i motsetning til empirisk evaluering der evalueringen gjennomføres ved observasjon av brukers faktiske bruk av løsningen, enten i et brukskvalitetslaboratorium eller i felt.

Det ble gjennomført analytisk evaluering av tre ulike mobile løsninger fra tre utviklingsbedrifter. Den ene løsningen var en mobil applikasjon for pasientdata til bruk på sengeposter ved sykehus, den andre løsningen var en mobil applikasjon for parkeringsvakter, og den tredje løsningen var en mobil applikasjon for å få tilgang til nyhetsklipp fra et Intranett, bla. tenkt brukt mot Stortingets intranett. Alle løsningene var kjørende på tidspunktet for evalueringen. To av løsningene var i bruk i ulike virksomheter.

Den analytiske evalueringen ble gjennomført i henhold til en metode for strukturert gjennomgang av brukergrensesnitt med gruppe av eksperter, basert på såkalt "Cognitive Walkthrough". Metoden er beskrevet i SINTEF-notatet "Evaluering av mobile brukergrensesnitt: Strukturert gjennomgang med gruppe av eksperter" (Haugset et al, 2005). En evaluering gjennomført etter denne metoden gir en strukturert oversikt over brukerproblemer, forslag til designforbedringene, og gode løsninger i den foreliggende løsningen. I eksempel-caset var det 3-5 eksperter i hver gruppe under evalueringene.

Bruk av evalueringsprosedyren i eksempel-caset viste følgende:

- Usability-eksperter gjorde langt flere funn enn domene-eksperter, men domene-eksperter gjorde forholdsvis flere funn med høy viktighets-klassifisering
- Over halvparten av domene-eksperterens funn ble gitt høy prioritet av utviklingsbedriften, mens en drøy fjerdedel av usability-eksperterens funn ble gitt høy prioritet
- Domene-eksperterens funn ble vurdert som nyttigere enn usability-eksperterens funn. Dette skyldes dels at domene-eksperterens kunnskap ble vurdert som mer hensiktsmessig, dels at usability-eksperterene hadde en tendens til å fokusere på funn med lav viktighet.

Evalueringen i eksempel-caset er beskrevet i SINTEF-rapporten "Analytisk usability-evaluering i gruppe: Sammenligning av usability-eksperter og domene-eksperter som evaluatorene." (Følstad, 2005).

4.2 Overordnet analyse av resultatet

Resultatet av en brukersentrert kravinnhenting eller evaluering er et sett funn, klassifisert i henhold til viktighet. Den overordnede analysen av resultatet utarbeides i form av en tabell. Tabell 2 viser den overordnede analysen av resultatet i eksempel-caset.

Evaluatorenes klassifisering av viktighet	Usability-eksperter		Domene-eksperter	
	Antall	Andel	Antall	Andel
Kritisk	28	17 %	25	27 %
Alvorlig	45	27 %	33	36 %
Kosmetisk/ikke klassifisert	92	56 %	34	37 %
Sum	165	100 %	92	100 %

Tabell 2: Overordnet analyse av resultatet fra eksempel-caset.

Siden fokus i eksempel-caset var å undersøke om det gir mest nytte å bruke domene-eksperter eller usability-eksperter ved evaluering av mobile IKT-løsninger til klassifisering og gruppe evaluatorene, skilles det mellom disse to gruppene i oversikten.

4.3 Vurdere viktighet

Resultatets viktighet vurderes med utgangspunkt i tabellen med den overordnede analysen.

Resultatet fra eksempel-caset gir grunnlag for følgende vurderinger:

- Over 50 prosent av usability-eksperternes funn er kosmetiske eller ikke klassifisert. Dette indikerer at for mye av den brukersentrerte aktiviteten gikk med til å fokusere på mindre viktige forhold.
- Begge grupper frembrakte en vesentlig andel funn med høy viktighet. Dette indikerer at det var nyttig å gjennomføre undersøkelsen uavhengig av gruppe, forutsatt at resultatet ble tatt til etterretning (undersøkes under påvirkningskraft) og holder høy kvalitet (undersøkes i begrenset grad under riktighet, kompletthet og påvirkningskraft).

Vær oppmerksom på at denne enkleste formen for viktighetsvurdering gir ikke grunnlag for kontroll av klassifiseringen som har blitt gjort i løpet av den brukersentrerte aktiviteten. Selv om man klassifiserer i henhold til et fast begrepsapparat (f.eks. ved å klassifisere brukerproblemer i henhold til de definerte kategoriene kritisk, alvorlig og kosmetisk), kan den klassifiseringen man gjør være preget av subjektiv vilkårlighet. For de fleste HCI-praktikere vil det være en god investering å bruke litt ressurser på etterprøving av egen klassifisering (se Kapittel 3.1), for å motvirke vilkårlighet i egen resultatrapportering.

4.4 Erfaringsinnsamling med utviklingsteam

”Erfaringsinnsamling med utviklingsteam” og ”utviklingsteamets prioritering” gjennomføres som et arbeidsmøte mellom ansvarlig for den brukersentrerte aktiviteten og en eller flere representanter for utviklingsteamet. Beregn ca. 2 timer på gjennomføring av arbeidsmøtet.

Erfaringsinnsamling med utviklingsteamet utgjør et grunnlag for å vurdere resultatets

- Riktighet
- Kompletthet

Riktighet vurderes på grunnlag av utviklingsteamets prioritering i tillegg til erfaringsinnsamling med utviklingsteam. Kompletthet vurderes utelukkende på grunnlag av erfaringsinnsamling med utviklingsteam.

Erfaringsinnsamling med utviklingsteam gir, hvis ønskelig, mulighet for å samle informasjon om

- Overordnet vurdering av den brukersentrerte aktivitetens nytte
- Aktivitetens bidrag til kompetanseheving i utviklingsteamet
- Aktivitetens bidrag til forankring blant brukergrupper og interessenter.

Erfaringsinnsamlingen gjøres i henhold til en intervjuguide. På grunnlag av metodeevalueringen i eksempel-caset presenteres følgende forslag til spørsmål til intervjuguide for undersøkelse av brukersentrert evaluering:

(Overordnet)

1. I hvilken grad har evalueringen vært nyttig for dere?
2. Utdyp hvorfor/hvorfor ikke evalueringen har vært nyttig for dere.

(Riktighet)

3. Er resultatet på overordnet nivå i samsvar med den kunnskap dere allerede har?
4. I hvilken grad består resultatet av funn som er kjent for dere fra før?
Vis gjerne til konkrete funn eller deler av løsningen.

(Kompletthet)

5. Finnes det vesentlige vansker tilknyttet løsningens brukergrensesnitt som dere forventet at evalueringen ville fange opp, men som ikke var med i resultatet?
Utdyp for de forhold dette gjelder.

(Kompetanseheving)

6. Har du/dere hatt noen form for læringsutbytte fra gjennomføring av, og evt. deltagelse i evalueringen - ut over det å motta resultatet fra rapporten?
La den/de som intervjues selv definere hva som forstås med læringsutbytte
Be om utdyping av de ulike former for læringsutbytte som foreslås

(Forankring)

7. Brukersentrert evaluering kan i enkelte tilfeller bidra til økt forankring hos fremtidige brukere av løsningen, og således forenkle en innføringsprosess. Tror du/dere den evalueringen som har vært gjennomført her kan ha en slik forankringseffekt?
Be om utdyping av de ulike former for forankring som foreslås

Intervju-dataene gir grunnlag for kvalitativ vurdering av resultatets riktighet og kompletthet. Denne kvalitative gjennomgangen av riktighet suppleres med en kvantitativ gjennomgang av riktighet under utviklingsteamets prioritering.

4.5 Utviklingsteamets prioritering

Utviklingsteamets prioritering utarbeides ved at en eller flere representanter for utviklingsteamet går gjennom alle funn i tur og orden. Gjennomgangen gjøres i dialog med ansvarlig for evalueringsprosedyren.

Under gjennomgangen vil hvert funn inkludert i resultatet vurderes separat. Utviklingsteamets representanter tilordner det enkelte funn en prioritet (høy/middels/lav) eller kategoriserer funnet som "feil". Resultatet av utviklingsteamets prioritering presenteres som en tabell. Utviklingsteamets prioritering i eksempel-caset presenteres i Tabell 3.

Utviklingsteamets prioritering	Usability-eksperters funn		Domene-eksperters funn	
	Antall	Andel	Antall	Andel
Høy	44	27 %	50	54 %
Middels	34	21 %	10	11 %
Lav	60	36 %	15	16 %
Feil	27	16 %	17	18 %
	165	100 %	92	100 %

Tabell 3: Utviklingsteamets prioritering i eksempel-caset

Forskjeller mellom to ulike grupper eller metoder kan undersøkes gjennom statistisk hypotesetesting ved Chi-kvadrat analyse. Enkle beskrivelser av Chi-kvadrat analyse, samt verktøy for å gjennomføre analysen er tilgjengelig på Internett (se f.eks. "Calculation for the Chi-Square Test", tilgjengelig på: <http://www.unc.edu/~preacher/chisq/chisq.htm>). Vær oppmerksom på at Chi-kvadrat analyse forutsetter uavhengighet mellom hver enkelt observasjon av samme kategori som inngår i analysen. Dette innebærer at Chi-kvadrat analyse ikke er egnet i tilfeller der det er åpenbare avhengighetsforhold mellom observasjonene (f.eks. at alle brukerproblemene av en bestemt kategori ble behandlet annerledes enn de andre identifiserte brukerproblemene pga. andre forhold enn gruppeforskjeller eller forskjeller mellom metoder).

I henhold til forslaget til tommelfingerregel i Kapittel 3.2, kan resultatet fra eksempel-caset gi grunnlag for følgende vurderinger:

- Over halvparten av usability-ekspertenes funn ble gitt lav prioritert eller vurdert å være feil. Påvirkningskraften kan antagelig vurderes å være lav.
- Over halvparten av domene-ekspertenes funn ble gitt høy prioritet. Påvirkningskraften kan antagelig vurderes å være høy.

Oppsummeringen av antall feil gir mulighet for å supplere den kvalitative vurderingen av resultatets riktighet (Kapittel 4.4) med kvantitative data. Det er imidlertid viktig å huske på at denne evalueringsprosedyren ikke inkluderer sikre kvantitative data om feil; dette krever brukertest eller feltundersøkelse av den ferdige løsningen. Utviklingsteamets vurdering av resultatets riktighet må følgelig kun tas som en indikasjon, ikke som sikker kunnskap.

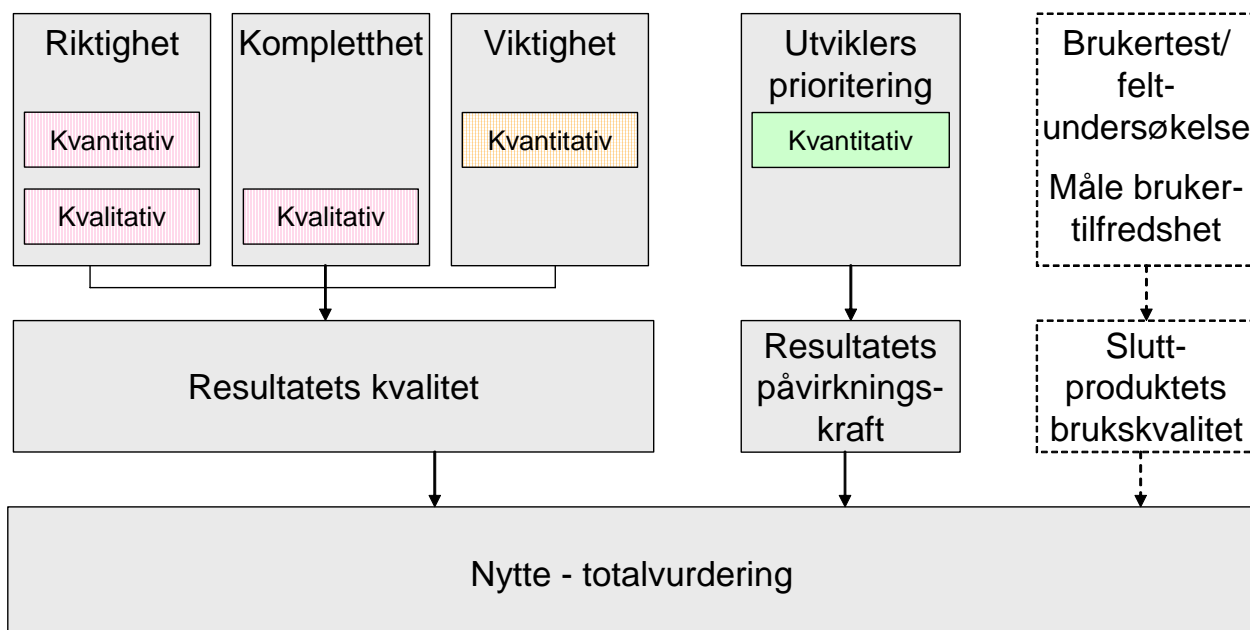
I eksempel-caset ser vi at snaut en femtedel av funnene ble vurdert som feil av utviklingsteamet. Dette kan synes å være noe høyt, samtidig som det antagelig ikke representerer en kritisk svakhet ved metoden. Det er vanskelig å sette opp forslag til grenseverdier for hvor stor andel feil man skal akseptere i resultatet fra en evaluering, særlig fordi konsekvensene av feil kan variere kraftig.

4.6 Totalvurdering av den brukersentrerte aktiviteten

En totalvurdering av den brukersentrerte aktiviteten kan settes opp på grunnlag av overordnet analyse, vurdering av viktighet, erfaringsinnsamling med utviklingsteam og utviklingsteamets prioritering. Denne vurderingen gir et uttrykk for nytte som det krever lite kostnad å utarbeide, men som kan gi god innsikt i egen brukersentrert praksis.

Vurdering av nytte

I Figur 4 gis det en oversikt over utkommet fra "Quick and clean" evalueringsprosedyre for brukersentrerte aktiviteter.



Figur 4: Oversikt over utkommet fra "Quick and clean" evalueringsprosedyre.

Innsamlede data er angitt som kvantitative eller kvalitative. Data med høy grad av sikkerhet er merket med grønt ensfarget bakgrunn, data med noe lavere sikkerhet er merket med gul rutete bakgrunn, og data med lav sikkerhet er merket med røde vertikale striper.

Data om sluttproduktets brukskvalitet inngår ikke i den enkleste formen for evalueringsprosedyre. Elementer tilknyttet dette er angitt som hvite bokser med stiplet kant.

Evalueringsprosedyren leder ikke frem til en endelig "karakter" av den brukersentrerte aktivitetens nytte, men fungerer som en hjelp til å vurdere faktorer av betydning for aktivitetens nytte. Imidlertid er det angitt tommelfingerregler for tolkning av enkelte kvantitative resultater.

Merk at evalueringsprosedyren i sin enkleste form kun henter data fra utviklingsteamet og HCI-praktikerens egen resultatrapportering. Dvs. at den "fasiten" som utarbeides gjennom evalueringsprosedyren er begrenset til utviklingsteamets og HCI-praktikerens eget beste skjønn. Evalueringsprosedyren vil være en nyttig hjelp for HCI-praktikeren til å utnytte både egne resultater og utviklingsteamets kunnskap til å forbedre sin egen praksis. Samtidig bør brukeren av evalueringsprosedyren være oppmerksom på at prosedyrens evne til å gi sikker kunnskap om evalueringresultatets kvalitet er begrenset, nettopp fordi man ikke har benyttet sikrere (og dyrere) metoder for å utarbeide "fasiten".

Vurdering av nytte i forhold til kostnad

Ved hjelp av evalueringsprosedyren vil den enkelte HCI-praktiker kunne gjøre vurdering av en brukersentrert aktivitetens nytte, for deretter å se dette i forhold til aktivitetens kostnad. Denne vurderingen vil være basert på skjønn, idet evalueringsprosedyren ikke gir noen endelig "karakter" av aktivitetens nytte.

Ved sammenligning av to eller flere brukersentrerte metoder eller aktiviteter vil det være nyttig å ta i betraktning kostnaden assosiert med metodene, i tillegg til nytten av metodenes resultater. Relevante kostnader er beskrevet i Kapittel 2.5. Dersom metodene ikke skiller seg vesentlig fra hverandre når det gjelder nytte, kan forskjeller i kostnad være avgjørende for hvilken metode man velger å benytte neste gang.

5 Vurdering og videreføring av evalueringsprosedyren

Den presenterte evalueringsprosedyren har blitt brukt for å sammenligne to brukersentrerte evalueringsmetoder med hensyn på nytte i utviklingsprosessen. Resultatet av denne evalueringen har blitt gjengitt i eksempel-caset i kapittel fire.

Utprøvingen i eksempel-caset viste at evalueringsprosedyren gir nyttig innsikt i fordeler og ulemper med ulike evalueringsmetoder i forhold til hverandre, og at utkommet av evalueringsprosedyren kan brukes som grunnlag for å velge en metode fremfor en annen ved senere utviklingsprosjekter.

Evalueringsprosedyren har ennå ikke blitt brukt for å evaluere enkeltstående HCI-praktikers egen brukersentrerte aktivitet. Senere i EFFIN-prosjektet vil vi innhente erfaringer med denne måten å bruke prosedyren på. Disse erfaringene vil danne grunnlag for en revidert versjon av prosedyren.