

**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

Resultatet av litteraturgjennomgang i EFFIN,WP4; "Hva er nytten av brukersentrerte metoder og aktiviteter?"

FORFATTER(E)

Jan Heim, Linn Anette Solberg, Asbjørn Følstad

OPPDRAGSGIVER(E)

EFFIN-prosjektet

RAPPORTNR. STF90 A04092	GRADERING Åpen	OPPDRAGSGIVERS REF.	
GRADER. DENNE SIDE Åpen	ISBN ISBN-82-14-03642-9	PROSJEKTNR. 403318.04	ANTALL SIDER OG BILAG 20
ELEKTRONISK ARKIVKODE eRoom/EFFIN/WP-04/Litteraturgjennomgang		PROSJEKTLEDER (NAVN, SIGN.) John Krogstie	VERIFISERT AV (NAVN, SIGN.) Petter Bae Brandtzæg
ARKIVKODE	DATO 2004-12-07	GODKJENT AV (NAVN, STILLING, SIGN.) Bjørn Skjellaug, Forskningsjef	

SAMMENDRAG

EFFIN-prosjektet skal utvikle brukersentrerte metoder og verktøy for å introdusere ny teknologi i offentlig sektor. En vesentlig del av dette arbeidet er å utarbeide hensiktsmessige måter å måle effekten, eller nytten, av brukersentrerte aktiviteter i en utviklingsprosess. Denne rapporten beskriver resultatet av en gjennomgang av relevant "Human-Computer Interaction" (HCI) litteratur på området.

Rapporten beskriver fire ulike tilnærminger til nytten av brukersentrerte aktiviteter, identifisert i litteraturen. Denne firedelingen kjenner vi ikke fra tidligere.

1) *Rettfærdiggjøringstilnærmingen*, der det argumenteres for nytten av brukersentrert utvikling i henhold til et omfattende sett kriterier, men man diskuterer ikke hvorvidt brukersentrerte aktiviteter og metodene faktisk kan knyttes direkte til nytten som fremheves.

2) *Den reduksjonistiske tilnærmingen*, der man sammenligner forskjellige brukersentrerte metoder ut fra deres umiddelbare output, isolert fra konteksten de blir brukt i. Man vurderer bl.a. ikke hvordan "output" fra metodene passer inn i den generelle utviklingsprosessen. I denne tilnærmingen fokuserer man på hvordan evaluere effektiviteten til metodene og hvilke kriterier som bør brukes i vurderingen av metodens "output".

3) *Den prosessorienterte tilnærmingen*, der det reduksjonistiske perspektivet kritiseres for å være for enkel når den gjelder effektmåling, og at forskningen så langt har vært til liten nytte for praktikerne. I denne tilnærmingen føyer man til flere aspekter som også bidrar til metodenes effektivitet og gevinst i utviklingsprosessen som bl.a. om en metode fører til designendringer som faktisk fikser problemet.

4) *"Spør ekspertene"-tilnærmingen* bidrar med et subjektivt perspektiv på nytte, idet man rett og slett mener at man får kunnskap om metodens nytte ved å spørre de som bruker metoden selv – ekspertene.

Rapportering av et krav/brukerproblemer/designforslag på grunnlag av en brukersentrert aktivitet kaller vi et *råd*. En viktig konklusjon er at nytten av råd basert på brukersentrerte metoder bør vurderes i henhold til følgende kriterier:

- Riktighet. (Correctness)
- Viktighet. (Importance, severity)
- Kompletthet. (Thoroughness)
- Endringspotensial – hvor langt rådet går i å gi utvikleren en gjennomførbar løsning
- Påvirkningskraft. (Impact, persuasive power).
- Kompetanseheving blant deltagere i utviklingsprosessen
- Forankring blant brukerne av løsningen

STIKKORD	NORSK	ENGELSK
GRUPPE 1	IKT	ICT
GRUPPE 2	Menneske-maskin interaksjon	HCI
EGENVALGTE	Litteraturgjennomgang	Literature review
	Metoder	Methods

INNHALDSFORTEGNELSE

1	Hvordan måle nytten av brukersentrerte metoder og aktiviteter?.....	3
1.1	Innledning.....	3
1.2	Fokus og avgrensning.....	3
2	Rettferdiggjøringstilnærmingen.....	5
2.1	Beskrivelse av tilnærmingen	5
2.2	Nytte i rettferdiggjøringstilnærmingen.....	5
2.3	Måling av nytte.....	6
2.4	Diskusjon.....	7
3	Den reduksjonistiske tilnærmingen.....	8
3.1	Beskrivelse av tilnærmingen	8
3.2	Nytte i den reduksjonistiske tilnærmingen.....	9
3.3	Måling av nytten.....	10
3.4	Diskusjon.....	11
4	Den prosessorienterte tilnærmingen	11
4.1	Beskrivelse av tilnærmingen	11
4.2	Nytte i den prosessorienterte tilnærmingen.....	12
4.3	Måling av nytten.....	15
4.4	Diskusjon.....	15
5	”Spør eksperten”-tilnærmingen	15
5.1	Beskrivelse av tilnærmingen	15
5.2	Nytte i ”spør eksperten”-tilnærmingen.....	16
5.3	Måling av nytten.....	16
5.4	Diskusjon.....	17
6	Generell diskusjon	17
6.1	Oppsummering	17
6.2	Nytte – en siste gang	17
6.3	Hvordan vurdere nytten av råd?	18
6.4	Nytte i forhold til deltagelse i prosessen	18
6.5	Relasjoner mellom begrepene	19
7	Referanseliste.....	19

1 Hvordan måle nytten av brukersentrerte metoder og aktiviteter?

1.1 Innledning

I EFFIN-prosjektet har vi en ambisjon om å måle kostnads- og nytteeffekten av å benytte brukersentrerte metoder i en utviklingsprosess. Hensikten med dette er å gjøre oss i stand til å vurdere hvorvidt en brukersentrert tilnærming er hensiktsmessig og lønnsom når man skal introdusere og utvikle offentlige tjenester i offentlig sektor. Fokus i denne rapporten vil være å kartlegge indikatorer relatert til nytten av brukersentrert utvikling, og måling av omfanget av dets effekt. Økt effektivitet av IT-tjenester som utvikles, vil også bli evaluert.

Før vi forsøker å måle gevinsten av pågående brukersentrerte aktiviteter, har vi gjennomført en litteraturgjennomgang for å fastslå "state-of-the-art", eller nåværende kunnskap om vanlige effektmåling av brukersentrerte utviklingsmetoder. I etterkant av denne gjennomgangen vil relevante alternative indikatorer av effekter av brukersentrert arbeid, bli gransket i samarbeid med prosjektpartnerne.

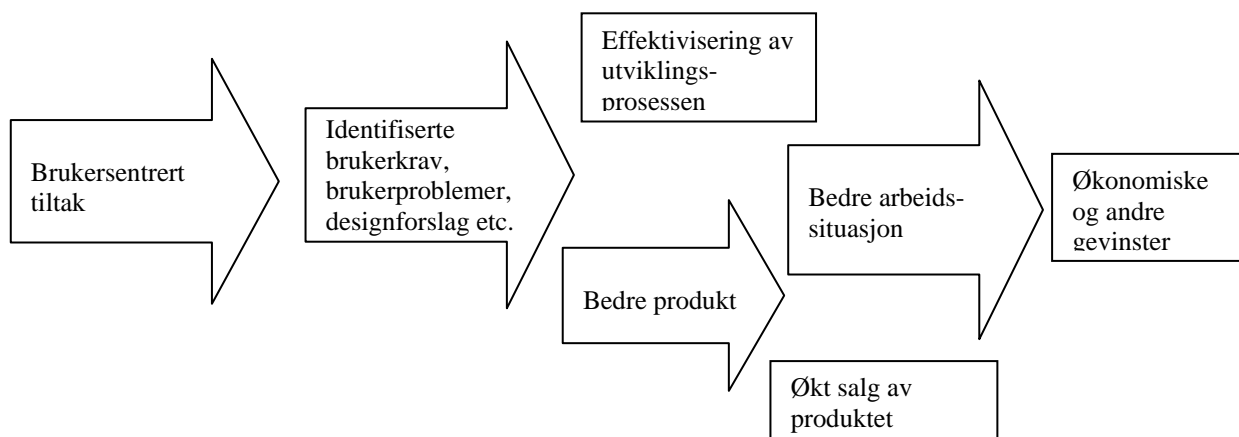
1.2 Fokus og avgrensning

Fokus i litteraturgjennomgangen har vært på effektivitet av brukersentrert utvikling, og hvordan målet omfanget av dets effekt. Vi har tatt utgangspunkt i formelen for effekt som brukes i kost/nytte analyser. Formelen er:

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

Kostnad er knyttet til investeringen gjort gjennom et tiltak. Nytte er knyttet til tiltakets resultat, og kan beskrives som differansen mellom produktet eller tjenestens verdi med tiltaket og dets verdi uten tiltaket.

Ulike tilnærminer til spørsmålet om kost/nytte varierer med hensyn til hvordan tiltaket og nytten er spesifisert. Tiltaket kan være spesifisert i relativt stor detalj slik at man kan beregne kostnadene ved tiltaket, eller den kan være beskrevet som at man i et prosjekt har en generell "Human factors"-tilnærming, med de problemer det medfører når det gjelder anslag av kostnad. Nyten kan være beskrevet i form av tiltakets resultat, f. eks. en prioritert oversikt over antatte brukerproblemer, eller som et anslag av verdien av tiltaket, ved at man beskriver kortere utviklingstid, økt effektivitet ved bruk av produktet/tjenesten, økt salg etc. Figur 1. illustrerer noen av de faktorer som har vært studert.



Figur 1. Mulige effekter av et brukersentrert tiltak

Som Figur 1 illustrerer kan et brukersentrert tiltak ha et sett med resultater som inngår i systemutviklingsprosessen. Disse resultatene kan bidra til å effektivisere utviklingsprosessen i seg selv og/eller gi et bedre produkt/tjeneste. Et bedre produkt/tjeneste kan føre til økt salg av eller bruk av produktet/tjenesten og/eller føre til en bedre arbeidssituasjon, som igjen kan ha økonomisk og andre gevinster.

En komplett kost/nytte studie må:

- 1) Spesifisere ekstra kostnader tilknyttet det brukersentrerte tiltaket
- 2) Spesifisere verdien med tiltaket
- 3) Spesifisere verdien uten tiltaket
- 4) Sannsynliggjøre at det er en nødvendig sammenheng mellom tiltaket og nytten

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 en blant mange andre aktiviteter i et utviklingsprosjekt, der 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.

Nyttebegrepet er sentralt for å vurdere effektiviteten av brukersentrert arbeid. Derfor har vi gjennomført en litteraturgjennomgang for å kartlegge hvilke faktorer som antas å bidra til nytten av brukersentrerte aktiviteter og metoder.

Hovedhensikten er å kartlegge de forskjellige eksisterende tilnærminger som finnes til "nytte" for brukersentrert utvikling. Hvordan defineres og operasjonaliseres nyttebegrepet innen de forskjellige tilnærminger til brukersentrert utvikling? I denne sammenhengen vil vi legge mindre vekt på hvordan man kan beregne kostnadene ved brukersentrerte aktiviteter.

Vi har valgt å strukturere resultatene fra litteraturgjennomgangen som fire ulike tilnærminger til "nytte". Denne firedelingen av litteratur tilknyttet kost/nytte av brukersentrerte metoder kjenner vi ikke fra andre steder. Inndelingen er ikke nødvendigvis basert på at forfattere tilknyttet de ulike tilnærmingene identifiserer seg med hverandres synspunkter. Inndelingen er snarere gjort i forhold til de ulike forfatterens overordnede synspunkt på hva nytte for brukersentrerte aktiviteter egentlig innebærer. De fire tilnærmingene er beskrevet i hvert sitt kapittel under overskriftene:

1. Rettferdiggjøringstilnærmingen
2. Den reduksjonistiske tilnærmingen
3. Den prosessorienterte tilnærmingen
4. "Spør eksperten"-tilnærmingen

I utgangspunktet var intensjonen å kartlegge alle typer brukersentrerte aktiviteter i denne rapporten. Det viste seg imidlertid at det som det var skrevet mest om var brukersentrerte evalueringsmetoder (Usability Evaluation Methods, UEM). Dette er hovedgrunnen til at metoder for kravspesifikasjon og innføringsprosesser nesten ikke er berørt i denne rapporten.

2 Rettferdiggjøringstiltaksnæringsen

2.1 Beskrivelse av tiltaksnæringsen

Utgangspunktet for rettferdiggjøringstiltaksnæringsen er knyttet til sent åtti- og tidlig nittitalles litteratur, der fokus er på å gi flest mulig gode grunner for at man bør inkludere brukersentrerte aktiviteter i systemutviklingsprosesser. Hele tiltaksnæringsens rasjonale er at brukersentrerte metoder og teknikker er for lite i bruk, og man bidrar med gode argumenter for å få gjennomslag for slike teknikker.

I tidlig litteratur innen denne tiltaksnæringsen fokuseres det i hovedsak på nytte i forhold til å gjøre systemer i en arbeidssituasjon mer lønnsomme i innføring, bruk og vedlikehold. Dette på grunn av den forholdsvis lave utbredelsen av applikasjoner for konsumentmarkedet. I senere litteratur vektlegger i tillegg dette perspektivet nytte i forhold til økt lønnsomhet i konsumentmarkedet f.eks. i forhold til produkters salgbarhet og kunders evne til å gjennomføre kjøp i nettbutikker. Generelt tilstreber litteratur som er klassifisert inn under rettferdiggjøringstiltaksnæringsen, å gi økonomiske begrunnelser for brukersentrering, selv om andre ikke-målbare-fordeler også ofte blir nevnt.

Bias and Mayhew's (1994) *Cost-Justifying Usability* er det mest sentrale bidraget innen denne tiltaksnæringsen. Her finnes artikler fra flere sentrale teoretikere bl.a. Nielsen, Mantei og Karat.

2.2 Nytt i rettferdiggjøringstiltaksnæringsen

Innen *Rettferdighetstiltaksnæringsen* listes det rutinemessig opp mange former for nytte ved bruk av brukersentrerte aktiviteter i utviklingen av nye produkter eller tjenester. Nytt som nevnes av Bias and Mayhew (1994) og Mantei and Teorey (1988) er blant annet:

- Økt brukskvalitet - målt gjennom brukertester.
- Reduksjon av brukerfeil
- Økt arbeidskvalitet og effektivitet
- Ergonomisk forbedring
- Redusert arbeidsfravær
- Mindre behov for support
- Mindre opplæringskostnader
- Reduserte vedlikeholdsutgifter
- Økt salg gjennom produktet (e-handel)

I følge Mantei og Teorey kan tiden det tar å lære et nytt system reduseres med $\frac{1}{4}$ hvis man bruker Human Factors design (brukersentrerte aktiviteter). Designendringer som blir gjort på prototypen og det ferdige systemet koster $\frac{1}{4}$ av hva det ville kostet å gjøre disse endringer i et system som allerede er i bruk.

Spesifikk nytte som gjelder for utviklerorganisasjonen:

- *Identifikasjon av brukskvalitetsproblemer*
- *Identifikasjon av feil i kravfasen*
- *Felles syn på produktutviklingen*
- *Økt kundetilfredshet*
- *Gjenbruk av analyse og designkunnskap*
- *Forflytte designendringer til en tidligere fase i produktets livsløp*

For produkter som skal ut på markedet nevnes ofte:

- Økt salg av produktet
- Kortere tid før produktet kommer i salg

Re-design kan være svært kostbart, og brukere bør derfor involveres tidlig i livssyklusen for å unngå å bomme på brukergrupper og brukerbehov. Når brukere ikke involveres tidlig, kan ikke designerne vite om de har bygget et funksjonelt ubrukelig system inntil sent i prosessen. Dette fører vanligvis til dyrt re-design. Når man bruker software prototyper, kan brukeren få en følelse av hvordan softwaren er å bruke før den endelige løsningen implementeres. På denne måten vil mer interaksjon og designendringer skje tidlig i livssyklusen. Mantei og Teorey nevner også flere ikke-målbare-fordeler med brukersentrerte metoder:

- *Unngå å utvikle unødvendige prosedyrer:* Brukerne bruker helst mindre kompliserte prosedyrer for å utføre en oppgave - selv om de er litt mindre effektive. Det er derfor lite trolig at mer avanserte prosedyrer som finnes i systemet, vil bli brukt.
- *Unngå sabotasjeproblemer:* Hvis det kreves at man skal bruke et system som er uegnet, vanskelig eller utilstrekkelig for en oppgave, kan dette føre til stor frustrasjon og sabotasje.
- *Motvirke at systemet krever så mye konsentrasjon at brukeren ikke har overskudd til kreativ problemløsning:* Hvis et softwaresystem krever mye konsentrasjon fra deltakerne for å utføre en oppgave, vil denne konsentrasjonen tar bort tilgjengelige mental kapasitet som kan brukes for å løse problemer generelt. Problemet kan løses, men ikke på en kreativ måte. Dette tapet av kreativitet som oppstår ved bruk av vanskelige systemer er vanskelig å måle, men tapene kan bli store

2.3 Måling av nytte

Innen det vi definerer som rettferdiggjøringstilnærmingen benyttes to interessante måter å måle nytte på. Den ene er måling av nytte gjennom estimer. Den andre er måling av nytte gjennom empirisk undersøkelse av utviklingsprosessens sluttprodukt.

Måling av nytte gjennom estimer

Når dette perspektivet argumenterer for hvorvidt de ulike formene for nytte er substansielle, gjøres dette ofte på grunnlag av *estimer* av en eller annen gevinst, og hvordan den kapitaliseres over systemets levetid.

Det er ikke uten videre enkelt å estimere en direkte økonomisk gevinst ved å ta i bruk brukersentrering i utviklingsprosessen. Dette skyldes at mange av de formene for nytte som er beskrevet kan være vanskelige å måle i praksis. F.eks. vil en mulig nytte som ”økt arbeidskvalitet og effektivitet”, som det hyppig argumenteres for innen rettferdiggjøringstilnærmingen, være problematisk å måle. Dels fordi man da må tallfeste hvor mye tid som spares i form av økt arbeidskvalitet og effektivitet, og fordi man må forutsette en bestemt timepris for det arbeidet som spares. Et relevant eksempel på måling av nytte ved bruk av estimer er Jacob Nielsens undersøkelse fra 1994.

Nielsen (1994) gjennomførte en eksperthevaluering av hvor mye tid som spares i opplæring og bruk av et bestemt monitoreringsprogram i en telebedrift. Elleve evaluatorene rapporterte hvor mange dager man sparte i opplæring og hvor mange prosent raskere en ekspert ville arbeide dersom man hadde eliminert alle 44 brukerproblemer som var identifisert i en heuristisk evaluering. Ved å gjøre en del antagelser om vurderingene og brukskonteksten kan det beregnes et beløp for hva man sparer per eliminert problem. Estimerte reduserte utgifter for et system som blir brukt av 250 brukere blir i Nielsens studie:

<u>Timelønn</u>	<u>\$15</u>	<u>\$40</u>
Reduksjon av...		
...opplæring av brukere	15 000	40 000
...brukerfeil	48 375	129 000
...utgifter til vedlikehold	24 000	24 000
<u>sum i \$</u>	<u>87 375</u>	<u>193 000</u>

Vi ser at Nielsen utleder svært konkrete gevinster basert på sine estimater av besparelser knyttet til redesign. Gevinstene er store - langt større enn kostnaden knyttet til evalueringen. Denne måten å estimere gevinst på er imidlertid problematisk, og vil diskuteres nedenfor i Kapittel 2.4.

Måling av nytte gjennom empirisk undersøkelse av sluttproduktet

Innen rettferdiggjøringstilnærmingen er det også vanlig å måle nytte ved å underkaste utviklingsprosessens sluttprodukt for empirisk undersøkelse. Gjerne i form av før-etter-målinger. Problemer ved denne måten å måle nytte på vil diskuteres i større dybde nedenfor, men det kan her nevnes at det faktisk er denne typen målinger ikke tar hensyn til hvorvidt et brukersentrert tiltak er skyld i endringen, og at man derfor er avskåret fra å vite om det var brukersentrerte aktiviteter eller andre forhold som eventuelt bidro til en målt nytte. I verste fall ender man i en situasjon der brukersentrerte aktiviteter blir gitt æren for forbedringer de ikke er direkte årsak til.

Nielsen & Gilutz, (2003) undersøkte empirisk gevinsten ved å gjøre et brukergrensesnitt mer brukervennlig. Trettifem web-steder ble undersøkt før og etter re-design, totalt ble det gjort 42 målinger. Forskjellige typer måleparametere ga noe forskjellig resultat. Nedenfor vises hovedresultatene som gjennomsnittlig økning i måleparametere før og etter re-design (noen ekstreme "outliers" har blitt fjernet i denne oversikten):

	Antall prosjekter	Økning
Salg	(N=14)	100%
Trafikk	(N=8)	150%
Produktivitet (Tid på å løse oppgave) (N=8)		161%
Bruk av enkelte funksjoner	(N=6)	202%

Undersøkelsen til Nielsen og Gilutz levner ingen tvil om de positive resultatene tilknyttet at en løsning gjøres brukervennlig. Imidlertid er det umulig å si noe sikkert om hvilke av aktivitetene som ble gjennomført under redesignet som er medvirkende til de positive resultatene. Det kan kanskje være fristende å ta denne typen resultater til inntekt for at det lønner seg å benytte brukersentrerte metoder og teknikker i utviklingsprosessen, men det er det altså ikke dekning for på grunnlag av denne typen undersøkelser.

2.4 Diskusjon

Et fellestrekk for litteratur innen rettferdiggjøringstilnærmingen er at den springer ut fra et ønske om å bevise en hel yrkesgruppes eksistensberettigelse. Dette legges det jo heller ikke skjul på når tradisjonen navngir seg selv: "Cost-justifying Usability". Kostnadene ved arbeid med

brukskvalitet skal rettferdiggjøres. Det er altså ikke spørsmål om dette er en lønnsom virksomhet (for det er det!), men om hvordan man kan anslå hvor lønnsomt det er.

Innenfor denne tradisjonen tallfestes nytte og kostnader ofte på en slik måte at selv beregninger basert på rene estimater kan få de et skinn av å være objektive og basert på faktiske regnskaper. Vanligvis er anslagene basert på vurderinger og "best guess", men også delvis på empiriske målinger. For beregninger av denne typen gjelder det imidlertid at de empiriske størrelsene som benyttes (f. eks. 2 minutter spart på å gjennomføre en oppgave) inngår i større beregninger hvor det gjøres anslag av antall brukere, bruksfrekvens, produktivitet, timelønn, redusert opplæringsstid og forrentning for å nevne de vanligste.

Det største problemet i denne forbindelse er nok å estimere produktivitet etter å ha gjennomført en endring som resulterer i økt effektivitet for en bestemt oppgave. Ett eksempel: Om en kontorarbeider gjør syv oppslag om dagen i en database og hvert oppslag tar 2 minutter har han altså en samlet tid i løpet av dagen på 14 minutter på dette arbeidet. Om så hvert oppslag bare tar ett minutt, kan vi da anta at han vil gjøre 14 oppslag? Eller kan vi anta at hver arbeider sparer 7 minutter per dag, i 235 dager per år altså ca 27 timer i året? Denne type gevinst forutsetter imidlertid at arbeidskonteksten og brukerens arbeidskapasitet perfekt kan tilpasses endringene og at spart tid uten videre kan konverteres til annet nyttig arbeid.

Det er påpekt at for å få et korrekt bilde av nytten av å benytte en bestemt brukersentrert metode må det utvikles to versjoner av løsningen, en utviklet med bruk av metoden og en uten. Begge versjoner må så være i stabilt bruk i lengre tid for å få et realistisk bilde av foredler og ulemper (Nielsen, 1994; Gray et. al 1992). Dette er et strengt krav som sjelden er oppfylt, men "før – etter studier" av prosjekter med og uten bruk av brukersentrerte metoder vil tilnærme seg dette kravet.

For en problemstilling som fokuserer på nytten av en bestemt type metode, er det ikke tilstrekkelig å vise at en gevinst kommer som følge av et re-design. Fordi det er mange faktorer som spiller inn vedrørende et produkts suksess, må man også kontrollere for en rekke andre faktorer. For det først er det mange kilder til god design som ikke uten videre kan tilskrives brukerinvolvering. Effektiv prosjektstyring, gode utviklingsverktøy, engasjerte og kreative medarbeidere er faktorer som ikke forutsetter brukerinvolvering. Dernest er det mange faktorer som virker inn ved lansering eller relansering av et produkt som ikke kan tilskrives produktet, men utenforliggende faktorer som type markedsføring, presseomtale, forventninger hos brukerne, bedre hjelp- og støttetjenester for å nevne noe. Det er derfor viktig å påpeke betydningen av å kontrollere for andre faktorer når en skal vurdere kost-nytte ved en bestemt metode. Det er ikke nok at to versjoner av samme programvare sammenliknes etter lang tids bruk, man må også sørge for at alle andre faktorer holdes konstant. Det er vel unødvendig å påpeke at dette antagelig er et praktisk talt umulig prosjekt; vi har ikke funnet noen studier som påberoper seg denne typen streng kontroll.

3 Den reduksjonistiske tilnærmingen

3.1 Beskrivelse av tilnærmingen

I denne tilnærmingen vurderer man brukersentrerte metoder og aktiviteter isolert fra konteksten de brukes i (derfor uttrykket reduksjonistisk). Fokus er på den umiddelbare "output" av metodene, som f.eks. antall identifiserte brukerproblemer, og ikke på hva "output"en" brukes til. Eller om "output"en" bidrar til at sluttproduktet blir mer brukervennlig. I det alt vesentlige er det forskjellige evalueringsmetoder som vurderes innen denne tilnærmingen, eller som man gjerne betegner det innen tradisjonen: UEM (usability evalueringsmetoder).

3.2 Nytte i den reduksjonistiske tilnærmingen

Det overordnede spørsmålet for kvaliteten av en UEM, innen litteratur vi har plassert i kategorien ”den reduksjonistiske tilnærmingen”, er i hvilken grad UEM”en hjelper en å oppdage reelle brukskvalitetsproblemer (usability-problemer). Et brukskvalitetsproblem er reelt dersom dette er et problem som brukere vil møte i reell bruk av løsningen som evalueres, og som har betydning for brukerens evne til å bruke løsningen.

Andre og Hartson

Andre (2000) og Hartson (2001) fokuserer i stor grad på hvordan evaluere effektiviteten til metodene, i tillegg fremhever de viktigheten av å kunne avgjøre om brukskvalitetsproblemene man finner er reelle. De vektlegger at det er vanskelig å sammenligne UEM”er, fordi det ikke finnes standarder for hvordan evaluere en UEM. De har derfor prøvd å definere og forklare en UEMs effektivitet.

Hva sammenligner man etter? Når er en metode god? Kriterier som presenteres av Andre og Hartson inkluderer:

- *Reliability* – Vil en få liknende resultat ved gjentatte undersøkelser gjennomført under like forhold?
- *Thoroughness* - Antall reelle problemer funnet i relasjon til antall eksisterende problemer
- *Validity* – Finner man utelukkende viktige/reelle problemer?

I tillegg til disse tre grunnleggende kriteriene for nytte opererer Andre og Hartson med to sammensatte kriterier: Måloppnåelse og kostnadseffektivitet.

- *Måloppnåelsen (effectiveness)* til en UEM defineres av Andre som en kombinasjon av ”thoroughness” og ”validity”. Høy ”thoroughness” alene kan føre til at problemer som ikke er reelle blir tatt med, og høy ”validity” alene kan føre til at reelle problemer blir utelatt. Ved høy ”thoroughness” unngår man dermed Type II feil (reelle problemer ikke funnet), og ved høy ”validity” unngår man Type I feil (”falske” problemer funnet).
- *Kostnadseffektivitet* er definert av Andre som kombinasjonen av måloppnåelse og kostnad. Dette betraktes som det ultimate evalueringskriteriet for UEM.

Andre og Hartson viser også til ”downstream utility” som et mulig evalueringskriterium for UEM”er. ”Downstream utility” hevdes imidlertid av Andre og Hartson å ikke ha noe med evalueringsmetoden som sådan å gjøre, men med metodens applikasjon i en utviklingsprosess. Hvorvidt ”downstream utility” er en del av evalueringsmetoden eller ikke er det som vi vil se i neste kapittel, delte meninger om. (I det kapittelet presenteres den prosessorienterte tilnærmingen, der den enkelte metodes evne til å inngå om en hensiktsmessig del av en utviklingsprosess betraktes som et sentralt kriterium for metodens nytte.)

Andre og Hartsons kriterier går igjen i senere sammenlignende UEM-studier, som f.eks. Law og Hvannberg (2002, 2004a, 2004b).

Andre tilnæringer til UEM evalueringskriterier

Et alternativt til Andre og Hartsons kriterier for UEM finner vi i Dutt, Johnson og Johnsons (1994) empiriske sammenligning av heuristisk evaluering (HE) og "cognitive walkthrough" (CW). I denne ble følgende vektlagt:

- *Problemenes alvorlighetsgrad*
- *Forekomst/frekvens av problem*
- *Evne til å generere krav til re-design*
- *Hvor lett er UEM å lære?*

De to første kriteriene tilsvarer Andres kriterium thoroughness, mens det tredje kriteriet (evne til å generere krav til re-design) minner om en mer prosessorientert tilnærming. Det siste kriteriet (hvor lett er UEM å lære) kan til en viss grad inngå i Andres kriterium kostnadseffektivitet.

Nielsen og Landauers (1993) teori om antall feil funnet i heuristisk evaluering og brukertester i forhold til antall deltagere (eksperter eller brukere) kan også sies å falle inn under den reduksjonistiske tilnærmingen. Vedrørende spørsmålet om "thoroughness" er det nettopp dette som adresseres i teorien, ved at man estimerer hvor mange problemer som metoden faktisk identifiserer i forhold til antall problemer som teoretisk finnes. I forhold til spørsmålet om "validity" anbefaler Nielsen og Phillips (1993) "severity ratings" som vanligvis innebærer at noen usability-eksperter angir på en skala hvor alvorlig et problem er. "Alvorlighet" er da definert ved hvor mange brukere som kan forventes å møte problemet, og hvor ødeleggende problemet er for oppgaveløsningen. Begge deler er i prinsippet empiriske størrelser, men man har lite empirisk materiale om dette.

Den reduksjonistiske tradisjon har bidratt til å kritisk vurdere gjeldende kunnskap i forhold til nytten av vanlige brukerinvolveringsmetoder. Gray & Salzman (1998) gjennomgikk fem studier som alle tok sikte på å sammenligne forskjellige UEM (Jeffries, Wharton, & Uyeda, 1991; Karat Campell, & Fiegel, 1992; Nielsen, 1992; Desurive, Kondziela, & Atwood, 1992; Nielsen & Phillips 1993). Sett under ett har disse fem studiene vurdert følgende nyttevariabler:

- Identifikasjon av potensielle brukerproblemer
- Sannsynlighet for å finne et potensielt brukerproblem
- Brukerproblemenes alvorlighetsgrad
- Gjennomføringstid

Videre har studiene sett på forhold vedrørende selve gjennomføringen av metodene:

- Hvordan metoden brukes mest effektivt (individuell eller i team)
- Kompetansen hos ekspertene (Utviklere, UI - eksperter, Studenter, eller "dobbelt - eksperter")

Gray & Salzman (1998) vurderer disse studiene på et sett med validitetskriterier, og finner at alle studiene har alvorlige problemer på alle, eller nesten alle, av disse kriteriene. Ikke uventet har dette resultert i en omfattende debatt, i alt 12 artikler i *Human-Computer interaction* 1998, volume 13, diskuterer Gray & Salzmans artikkel.

3.3 Måling av nytten

Måling av nytten i den reduksjonistiske tilnærmingen er mest interessant i forhold til thoroughness og validitet. Disse kriteriene forutsetter kjennskap til antall brukerproblemer tilknyttet løsningen som evalueres, eller i det minste en oversikt over hvilke av de identifiserte brukerproblemene som er reelle. Dersom man vet dette er det en smal sak å telle seg frem til hvor grundig eller valid evalueringsmetoden i bruk har vært.

Hvordan avgjør man så om et problem er reelt? I følge Andre (2000) er det to måter:

- Ekspertene vurderer om hvert enkelt problem er virkelig eller ikke.
- Man sammenligner problemene man har kommet frem til med en liste over problemer som man vet er virkelige brukskvalitetsproblemer (gjerne som resultat av en brukertest).

Problemet med denne formen for bestemmelse av hvilke brukerproblemer som er reelle kan imidlertid være at kriteriet for å evaluere en UEM vil være resultatet fra en annen UEM.

3.4 Diskusjon

Innen den reduksjonistiske tilnærmingen er ønsket og viljen til å kritisk analysere og revurdere nytten av de forskjellige metoder absolutt til stede. Det som tradisjonelt antas som nyttige resultater ved å anvende UEM er identifikasjon av potensielle brukerproblemer, antall og alvorlighetsgrad, og tid brukt på oppgaveløsning.

I forhold til den reduksjonistiske tilnærmingen er det interessant å se på arbeidet gjort på CUE (Comparative Usability Evaluation) i regi av DialogDesign (drevet av Molichs). Undersøkelser gjort på CUE viser:

- Ulike team finner ulike og i liten grad overlappende usability-problemer ved usability-tester (Molich et al., 2004)
- Ulike usabilityekspertene finner ulike og i liten grad overlappende usability-problemer ved heuristisk evaluering (Hertzum et al., 2002)

I begge undersøkelsene nevnt over var det utarbeidet en forholdsvis klar formulering av overordnede oppgaver og funksjonalitet som skulle inngå i evalueringen av de aktuelle nettstedene, henholdsvis www.hotmail.com og www.avis.com. I begge undersøkelser fant deltagerne over 300 brukerproblemer for det enkelte nettsted. Andelen identifiserte brukerproblemer klassifisert som alvorlige var langt lavere enn 300, men ingen av brukerproblemene som ble klassifisert som alvorlige ble funnet av mer enn halvparten av teamene og ekspertene. De aller fleste av de rapporterte alvorlige brukerproblemene ble funnet av færre enn 20 prosent av teamene og ekspertene.

Arbeidet tilknyttet CUE antyder at antall brukerproblemer knyttet til et produkt kan være uoverstigelig stort, og at problemene som identifiseres sterkt avhenger av hvem som gjennomfører undersøkelsen. Dette skaper problemer for den reduksjonistiske tradisjonens argumentasjon om (1) at nytten av en evaluering avhenger av størst mulig grad av "thoroughness" (dette er neppe hensiktsmessig siden antall problemer synes å kunne være svært høyt) og (2) at usability-testing kan være et godt validitetskriterium.

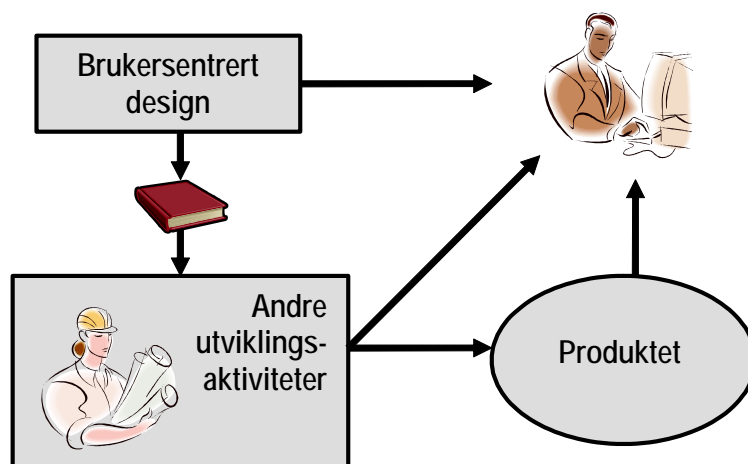
Gray & Salzman påpeker at studier hvor man prøver å sammenligne forskjellige metoder med hensyn på denne type nytte møter store metodiske problemer. Det mest åpenbare er problemene med den eksterne validiteten, nemlig hvorvidt de problemer som identifiseres ved bruk av en metode, faktisk vil være de samme problemene en bruker vil møte ved bruk av produktet i "den virkelige verden".

4 Den prosessorienterte tilnærmingen

4.1 Beskrivelse av tilnærmingen

Med den prosessorienterte tilnærmingen mener vi litteratur som dekker nytten av brukersentrerte aktiviteter, der "nytte" forstås som verdien av de brukersentrerte aktivitetenes leveranser, eller verdien av aktiviteten i seg selv. Nyttan av f.eks. en prototypeevaluering i henholdt til denne tilnærmingen kan være knyttet til (1) i hvilken grad leveransen av evalueringen (identifiserte

brugerproblemer og forslag til designforbedringer) er årsak til at utviklingsteamet implementerer endringer, (2) at endringene forårsaket av brukersentrerte aktiviteter faktisk innebærer en forbedring av produktets brukskvalitet, og (3) at den brukersentrerte aktiviteten bidrar til økt medvirkning og eierskap til løsningen. Aktiviteter tilknyttet brukersentrert design kan følgelig ha nytte på ulike nivå i utviklingsprosessen, men det er som Figur 2 skal illustrere, også andre deler av utviklingsprosessen som påvirker produktets brukskvalitet og brukerens opplevelse av å være medvirkende i utviklingsprosessen



Figur 2. Illustrasjon av den prosessorienterte tilnærmingen

Den prosessorienterte tilnærmingen kan ses på som en reaksjon på den reduksjonistiske tilnærmingen. Mens man i den reduksjonistiske tilnærmingen er tilfreds med å undersøke hvilke metoder som gir den riktige listen av brukerproblemer som "output", er man i den prosessorienterte tilnærmingen mer opptatt av hvilken effekt ulike metoder har på utviklingsprosessen og dens sluttprodukt.

Vi har bare funnet én artikkel som eksplisitt faller under denne tilnærmingen (Wixon, 2003). Vi er imidlertid rimelig sikre på at innvendningene mot den reduksjonistiske tilnærmingen og argumentet om å vurdere metoder som del av en større sammenheng (ikke isolert) deles av mange som praktiserer inne feltet.

Wixon (2003) kritiserer den reduksjonistiske tilnærmingen, og mener den er for enkel når det gjelder måling av effektivitet. Han mener at så langt har litteraturen om evaluering av brukskvalitetsmetoder vært til liten nytte for praktiserende, og at en må bruke måleparametere som er viktige for bedrifter. I praksis er målet i bedrifter å produsere, på kortest mulig tid, et suksessfullt produkt som møter spesifikasjonen med færrest mulig ressurser og med minimum av risiko. Dette krever altså, at for å utvikle en forståelse av metodologien for *anvendt usability*, må vi ta i bruk forretningskriterier, slik som hvor mye vi klarer å forbrede produktet og hvordan produktet klarer seg i markedet. Vi ser at argumentet er beslektet med Gray & Salzmans (1998) krav om ekstern validitet. John og Marks (1997) vektlegging av hvordan resultatene fra UEM inngår i utviklingsprosessen er også i tråd med denne kritikken.

4.2 Nytt i den prosessorienterte tilnærmingen

Wixons (2003) kritikk av den reduksjonistiske tradisjonen er at det å oppdage problemer er kun først steg i å forbedre et produkt. Det å isolere en metode fra dets kontekst fører til at en fjerner viktige vurderinger som hvordan teamet er, tilgjengelig ressurser, hvor lett det er å gjøre endringer

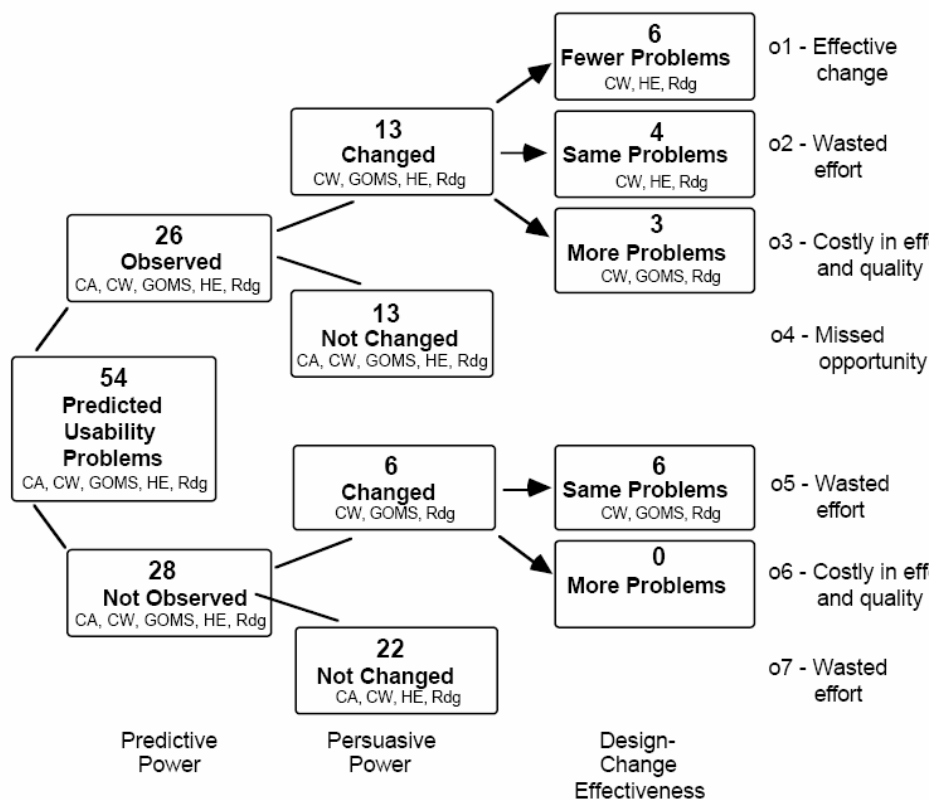
etc. En vitenskapelig tilnærming for å validere metoder er i uoverensstemmelse med den underliggende filosofi og kontekst i produkt utvikling, som har en teknisk tilnærming. I tråd med dette sier Molich (2004): "Some people have argued that usability professionals and their methods should ideally be measured on how effectively they introduce usability improvements into the product. We agree". (s. 72)

Mye av kritikken mot den reduksjonistiske orientering og vektlegging av feltstudier og metoder brukt "i kontekst" finner vi også innen den tradisjon som vekselvis kalles "The Scandinavian Approach" og "Participatory design" (se f.eks. Greenbaum & Kyng, 1991). Forskjellen fra det som ellers vil falle inn under den prosessorienterte tilnærmingen er vurderingen av hva som er "nytte". Innen denne tradisjonen legges et bredere perspektiv til grunn, ved at det ikke bare er organisasjonens forretningsmessige mål man ønsker å understøtte, men også de ansattes demokratiske rettigheter til medbestemmelse ved utformingen av arbeidsmiljøet. Utviklingen foregår gjennom diskusjoner og arbeidsmøter med de som er valgt til å representere brukerne, disse vil som regel ha en lagt mer aktiv rolle i utformingen av den endelige løsning enn det som er tilfelle ved andre tilnærminger. Innen den skandinaviske tradisjonen har altså brukerinvolvering en egenverdi ved at brukere og utviklere utvikler eierskapsfølelse til produktet og at det ivaretar deltageres rettigheter i forhold til medbestemmelse. Som en følge av dette vil også innføringsprosessene bli enklere. Nyttene ved brukerinvolvering blir da:

- Medbestemmelse
- Engasjement og eierskap
- Hensiktsmessige innføringsprosesser

John og Marks (1997) tar opp problemstillinger relevante for en prosessorientert vurdering av brukersentrerte aktiviteter, i det de ønsker å se på hvordan resultatene fra usability evalueringmetoder (UEM) blir tatt opp av den eller de som skal gjøre de foreslåtte endringer. I John og Marks' studie fikk seks hovedfagsstudenter i oppgave å sette seg inn i hver sin systemutviklingsmetode (Claims Analysis, Cognitive Walkthrough, GOMS, Heuristic Evaluation, Action Notation og Basline. Den sistnevnte innebar kun å lese kravspesifikasjonen) for så å evaluere samme applikasjon. Deretter ble en utvikler som hadde arbeidet med applikasjonen tidligere gitt resultatene av analysene. Utvikleren implementerte så de endringer som han fant naturlig for en ny og bedre versjon. Begge versjoner av applikasjonen ble brukertestet, og resultatene fra brukertesten tjente som fasit i forhold til hva som var reelle brukerproblemer og ikke.

Figuren nedenfor viser resultatene på en skjematisk måte. Vi ser at av de 54 problemene som ble predikert gjennom en eller flere av de seks metodene, var det bare 26 som ble faktisk observert i brukertesten. Av disse ble halvparten (13) endret i det nye designet. Seks problemer ble faktisk fjernet, 4 forble stadig et problem og 3 av endringene førte til nye problemer. Vi ser at re-designet netto bare hadde 3 færre problemer enn det opprinnelige designet og at mye arbeid ble lagt ned i å analysere og prøve å rette predikerte problemer som ikke var reelle. Forfatterne påpeker at resultatene fra denne studien kan være vanskelig å generalisere, siden det kun er benyttet en enkelt utvikler og kun en analysator per metode. Det synes likevel åpenbart at de har påpekt vesentlige forhold som bør tas inn ved vurdering av nytten av en UEM.



Figur 3. "Effectiveness tree" for påviste brukskvalitetsproblemer (John og Marks, 1997).

En av de tingene som gjør John og Marks' studie interessant er at de anerkjenner at det er utvikleren (eller utviklingsteamet) som bestemmer eksakt hva som skal endres. Og med mindre usability-eksperten selv er utvikleren eller har en sentral rolle i utviklingsteamet, er det andre enn usability-eksperten som avgjør skjebnen til de problemer og designforslag som spilles inn. Videre er det interessant at John og Marks understreker at endringene som gjøres kan enten kan føre til reduksjon i antall problemer brukerne opplever, at resultatet blir uendret eller at det blir flere problemer enn før. Design-enderingseffektiviteten ("design-change effectiveness") vurderes ved å teste resultatene av den nye versjonen av systemet i forhold til det gamle. Under denne synsvinkelen blir følgende beregninger som ligger til grunn for økonomiske nytte av metodene slik den ofte beregnes innen rettferdiggjøringsperspektivet helt feil. Her antas det at alle identifiserte problemer elimineres. I John og Marks' studie ble det identifisert 54 problemer, men det ble bare 3 problemer mindre i det nye designet.

John og Marks føyer flere vesentlige kriterier til listen over hvilke aspekter som er vesentlige å vurdere i forhold til nytten av en UEM.

- "Predictive power": Klarer en UEM å avdekke brukskvalitetsproblemer som brukere faktisk møter på?
- Overtalesesevne: Hvor mange av brukskvalitetsproblemerne man finner vha. metoden fører til endringer i implementeringskoden? En UEM må uttrykke problemene og føre bevis for dem på en måte som motiverer utviklere til å gjøre endringer.
- "Design change effectiveness": En UEM bør føre til designendringer som faktisk fikser problemet.

4.3 Måling av nytten

Hartson (2001) innvender i forhold til den reduksjonistiske tilnærming at metodene vurderes, uavhengig av den utviklingsprosessen metoden skal inngå i. Med andre ord vurderes metoden isolert (in vitro), og den undersøkte nytten av metoden (antall reelle og alvorlige brukerfeil funnet) er en annen enn nytten av prosessen (lage et brukervennlig produkt). Wixon (2003) anbefaler tilsvarende å evaluere brukersentrerte metoder in vivo – dvs. at man vurderer metodene på virkelige produkter og i virkelige situasjoner, og dermed studerer man metodene ikke isolert fra konteksten de brukes i.

Wixon argumenterer for at vi må ta i bruk casestudier heller enn eksperimentelle tilnærminger. En casestudietilnærming er den eneste praktiske måten å få frem kunnskap om anvendt usability, og også den mest effektive. Han viser til to kasusstudier hvorav den ene benytter RITE (Rapid Iterative Testing and Evaluation). Slik det blir fremlagt synes dette å være en vellykket anvendelse av "mainstream" usability fremgangsmåte. Man tester riktignok produktet i en form for brukertest, men prosessen er langt mer integrert i utviklingen av produktet enn hva som er vanlig i de reduksjonistiske studiene. Det andre kaset benytter en "etnografisk" tilnærming for å studere brukerne av et produkt, og påviser slik problemer som ikke nødvendigvis vil fremkomme i en laboratorietest. I disse studiene blir nytte implisitt forstått som utvikleres mulighet til å forbedre systemet gjennom aktiv deltagelse i evalueringsprosessene. Det blir bl.a. påpekt hvor viktig det er at de som skal gjøre eventuelle endringer i produktet selv er med og observerer hva som skjer i brukertestene.

4.4 Diskusjon

Problemet med den prosessorienterte tilnærmingen til undersøkelse av nytte er det samme som ved alle kasusstudier: Man har ingen uavhengig (objektiv) vurdering av hvor vellykket metoden har vært. Satt på spissen kan vi si at rapporteringen kan bli som følger: "I utviklingen av applikasjon X brukte vi metode A og resultatet ble veldig bra. (Vi fikk til og med en hederspris)."

Dette problemet skal ikke frata denne tilnærmingens rimelige og nødvendige påpekning av at det er metodens evne til effektivt å møte organisasjonens forretningsmessige mål som er det endelige uttrykk for metodens nytte. For at dette målet skal kunne oppnåes er det metodens egenskaper til å inngå i et integrert samspill med alle de andre systemutviklingsaktivitetene som er det vesentlige.

Også resultater oppnådd ved "Participatory design" unndrar seg til en viss grad muligheten for uavhengig vurdering. Selv om medvirkning eller demokratisk deltagelse er egenskaper ved utviklingsprosessen som kan kvantifiseres, er det vanskeligere å undersøke hvorvidt disse egenskapene ved prosessen fører til en tjeneste med høyere brukskvalitet på grunn av utfordringer knyttet til sammenligninger på tvers av prosjekter som har høy grad av egenart. Et tilleggsproblem er at det lett kan oppstå uenighet om hva som er målsetningen. Man kan lett tenke seg en situasjon der "de ansatte likte løsningen, men ledelsen mente den var lite effektiv".

5 "Spør eksperten"-tilnærmingen

5.1 Beskrivelse av tilnærmingen

Et alternativ til måling av metodikkens nytte er å måle nytte gjennom subjektive kriterier, til forskjell fra objektive kriterier. Gjerne ved å spørre de som bruker metoden (ekspertene selv) hvorvidt de synes den gir de resultatene de ønsker eller ikke.

Innen det vi har valgt å kalle "spør eksperten"-tilnærmingen er det gjennomført en undersøkelse av Vredenburg et al.(2002) som ble presentert på konferanse CHI 2002, samt en undersøkelse gjennomført av Gulliksen et al (2004) presentert på konferansen NordiCHI 2004. I tillegg arbeides

det innen EFFIN-prosjektet med en undersøkelse av usability-eksperter i Norden, der det foretas en survey blant om lag 200 personer. Fokus i denne undersøkelsen er bl.a. på hvilke metoder som brukes, hvor nyttige metodene oppleves på ulike tidspunkt i prosessen, og hvordan man evaluerer sine egne brukersentrerte aktiviteter.

5.2 Nytte i ”spør eksperten”-tilnærmingen

I studiet gjennomført av Vredenburg et al. (2002) var det usability-eksperters subjektive vurdering av hva som er ”nyttig” som ble målt. Det ble i denne undersøkelsen ikke spurt om årsakene til at ulike brukersentrerte metoder ble vurdert til å være mer eller mindre nyttige (useful). Dette understreker forskjellen mellom denne tilnærmingen og de tre andre tilnærmingene vi har presentert, der målbarhet iht. andre kriterier enn brukerne av metodens egne vurderinger har vært i fokus.

Det er ofte svært forskjellige egenskaper både ved metoder og metodenes brukere som gjør at de vurderes som mer eller mindre nyttige. Sannsynligvis vil for eksempel de som synes feltstudier er viktig si dette fordi de vurderer bruk ”i kontekst” som mest sentralt, mens de som synes brukerkravanalyse er viktigst gjør det fordi de antar at innspill tidlig i utviklingssyklusen er avgjørende.

At denne subjektive tilnærmingen ikke nødvendigvis tar stilling til hva nytte faktisk innebærer, bare om nytten er til stede eller ikke, kan betraktes som en styrke i den forstand at man kan få førstehånds tilbakemelding fra brukerne av metoden, uten at denne tilbakemeldingen forstyrres av diskusjoner rundt hva det egentlig vil si at noe er nyttig eller ikke. Dette gjør det enkelt for praktikere å velge metode med større sikkerhet, ut fra hva andre praktikere synes er nyttig. Samtidig er det vanskelig å se for seg at denne formen for tilbakemelding kan være tilstrekkelig dersom man ønsker å videreutvikle det metodeapparatet vi har i dag, siden man ikke får tilbakemelding på hvilke egenskaper ved en metode som gjør den særskilt nyttig eller unyttig. Det kan også være vanskelig å tolke tilbakemeldingene i denne typen undersøkelser med mindre man også samler inn detaljert informasjon om utviklingsprosessene de ulike metodene man undersøker har blitt brukt i, og hva man har ønsket å få ut av dem.

5.3 Måling av nytten

Vredenburg, Mao, Smith & Carey (2002) spurte 103 deltagere på en HCI konferanse, alle med bred erfaring innen UCD-området, om hvilke metoder de benyttet og hvilke de så som mest nyttige. Nedenfor er det presentert en tabell som oppsummerer resultatene.

Metode	Frekvens	Vurdert nytte *
Brukertesting / iterativt design	65	2,15
Brukskvalitetsevaluering **	43	2,39
Oppgaveanalyse	34	2,61
Uformell ekspert vurdering	31	3,28
Feltstudier	28	2,00
Fokusgrupper	16	2,79
Formell Heuristisk Evaluering	15	2,86
Prototyping uten brukertest	15	3,07
Brukerintervjuer	11	3,00
Spørreskjema	9	3,17
Brukerkravanalyse	7	2,00
Deltagende design	7	3,40
Kortsortering	5	3,33

*Rangering av de fem mest nyttige (1 best, 5 dårligst)

**Dekker en rekke forskjellige evalueringsmetoder

Det er interessant å se at de metodene som vurderes som mest nyttige (Feltstudier og Brukerkravanalyse) ikke er de som benyttes mest.

5.4 Diskusjon

En mulig svakhet med ”spør ekspertene” survey-undersøkelser, kan være at resultatene fra undersøkelsen blir lite generaliserbar fordi vi ikke vet hva ekspertene legger til grunn for sitt nyttebegrep. Fremtidige undersøkelser av denne typen bør i tillegg til rangering av nytte også fokusere på årsaker til at ulike metoder rangeres som nyttigere enn andre. I tillegg bør man også samle inn informasjon om hva som kjennetegner utviklingsprosesser der de ulike metodene fungerer henholdsvis godt og dårlig i, og på hvilket stadium i utviklingsprosessen de ulike metodene har vært i bruk.

6 Generell diskusjon

6.1 Oppsummering

Innen rettferdiggjøringstilnærmingen blir nytten av brukerinvolvering estimert med utgangspunkt i størrelser som prinsipielt er empiriske (for eksempel redusert tid for å løse en oppgave), og som så inngår i beregninger som tar hensyn til brukssituasjonen, antall brukere osv. Således får man anslag av hvor mye brukerinvolveringen er verd – som regel uttrykt som en pengeverdi.

Innen den reduksjonistiske tilnærmingen er nytten av brukerinvolvering definert som ”det resultatet metoden gir”. Det antas at det er nyttig i seg selv, for eksempel, å få kjennskap til potensielle brukerproblemer.

Den prosessorienterte tilnærmingen vektlegger den enkelte organisasjons spesifikke forretningsmål, og kritiserer de andre tilnærmingene for å benytte UEM ut-av-kontekst. I den prosessorienterte tilnærmingen er nytte bestemt ut fra i hvilken grad metodens resultater opptas av prosjektdeltagerne og integreres i systemutviklingsprosessen. Dette kan måles ved deltagende observasjon, men prinsipielt (som hos John og Marks, 1997) er det mulig å etterspore de forskjellige resultater av metodebruk for å få et bilde av hvilken skjebne de enkelte resultater har fått i løpet av prosessen.

”Participatory design” deler mye av perspektivet som antas i den prosessorienterte tilnærmingen, idet det fokuseres på medbestemmelse, engasjement og hensiktsmessige innføringsprosesser. Å måle disse er selvfølgelig vanskelig, men ikke prinsipielt umulig. Gode spørreskjemaer applisert på forskjellig stadium i utviklingsprosessen burde kunne måle noe av dette.

”Spør ekspertene”-tilnærmingen måler nytte ved nettopp å gjøre det: Spørre ekspertene. Dette har blitt gjort ved bruk av spørreskjemaundersøkelser.

6.2 Nytte – en siste gang

De forskjellige tilnærminger har forskjellig fokus på hvor i utviklingsprosessen nytten av brukersentrerte aktiviteter gjør seg gjeldende.. Hensikten med denne litteraturgjennomgangen har vært å undersøke nytten av brukersentrerte aktiviteter, og å gi en oversikt over mulige måter å måle nytte på. Dette siste kapitlet oppsummerer hvordan man kan undersøke nytten av en brukersentrert aktivitet, ved det tidspunktet resultatene blir produsert.

Selv om litteraturgjennomgangen nesten utelukkende har vært konsentrert om evalueringsmetodikk, vil vi forsøke å oppsummere hvordan nytte kan vurderes i relasjon til å:

- identifisere krav til produktet/systemet
- identifisere brukerproblemer knyttet til en versjon av produktet
- foreslå designforbedringer

Resultatet av å bruke brukersentrerte metoder er dels den ”output” som metoden gir, så som kravspesifikasjoner, problemlister etc., m.a.o. de fremkomne råd som utviklerne kan forholde seg til. Dels er det den betydningen bruken av metoden har gjennom deltagernes involvering i å gjennomføre metoden. Nyten av brukersentrerte aktiviteter kan derfor diskuteres i relasjon til

1. Nyten av de krav/brukerproblemer/designforslag som identifiseres
2. Nyte i forhold til deltagelse i prosessen.

6.3 Hvordan vurdere nytten av råd?

Rapportering av et krav/brukerproblemer/designforslag på grunnlag av en brukersentrert aktivitet kaller vi et *råd*. ”Output” av en brukersentrert aktivitet vil følgelig være en rekke råd, og disse kan vurderes med hensyn på nytte.

Riktighet. (Correctness). Hvor riktig er de råd som gies? Dette er en problemstilling som ofte behandles under begrepet validitet. Et råd er i denne sammenhengen ”riktig” om det fører et bedre produkt/tjeneste dersom rådet blir fulgt.

Viktighet. (Importance, severity). Et relatert spørsmål er hvor rådene er. Man kan tenke seg at et råd (f. eks. om fargen på en knapp i skjermbildet) er riktig i den forstand at det gjør produktet *litt* bedre. Det vil allikevel ikke være så viktig hvis forbedringen er marginal i forhold til et annet råd (f. eks. tilknyttet overordnede menystrukturer) som gjør applikasjonen *mye* bedre.

Kompletthet. (Thoroughness). De råd man gir kan ha større eller mindre grad av kompletthet, dvs. i hvilken grad rådene er dekkende for den applikasjonen man uttaler seg om. Om man kan formulere 10 relevante krav til produktet kan det være dekkende dersom det er alt som trengs, men om man identifiserer 10 problemer og det faktisk eksisterer 90 i tillegg er rådene man gir lite dekkende.

Endringspotensial. (Design change effectiveness). Et sett med råd som gies i forhold til en applikasjon kan ha større eller mindre endringspotensial. For eksempel kan det være slik at å påpeke et brukerproblem har begrenset endringspotensial dersom utvikleren ikke vet hva som skal til for å eliminere problemet. Om problembeskrivelsen følges opp med en begrunnelse om hvorfor det er et problem, og det samtidig gies forslag om hvordan problemet kan overkommes, blir endringspotensialet større.

Påvirkningskraft. (Impact, persuasive power). Et råd kan ha større eller mindre påvirkningskraft i utviklingsprosessen. Det vil først og fremst vise seg vet at rådet blir fulgt. Her er det mange faktorer som spiller inn; hvordan rådet er formulert og hvor enkelt/kostnadskrevende det er å følge, er de viktigste.

6.4 Nyte i forhold til deltagelse i prosessen

Kompetanseheving. En brukersentrert prosess kan være nyttig dersom den bidrar til økt forståelse hos utviklerne hva gjelder bruk, brukere og brukskontekst av applikasjonen. En vanlig iakttagelse er at utviklere synes de lærer mer av å observere brukere i en brukertest enn å lese rapporten med problemlistene i etterkant. Nyten av metoden er da å få en økt bevissthet rundt, og praktisk erfaring med brukersentrering i utviklingsprosessen.

Forankring. Å involvere brukere kan ha betydning for de interessenter som deltar i prosessene og de brukergrupper de representerer, ved at de opplever at deres synspunkter og problemstillinger blir tatt alvorlig. Dette kan igjen ha stor betydning når et system skal taes i bruk, ved at systemet er forankret i brukere og interessenter som opplever eierskap til systemet. Vi kaller dette for forankring.

6.5 Relasjoner mellom begrepene

Begrepene ovenfor har en del interessante relasjoner. Et råd kan ikke være *viktig* uten å være *riktig*. Et sett med råd som er *komplett* må inneholde alle *riktige* råd. Det vil være naturlig å kreve at det ikke skal inneholde noen gale råd, men det kan nok diskuteres om dette ligger i selve begrepet. Et råd med stort *endringspotensial* kan være *galt* eller *riktig*. Jo større endringspotensial, jo mer avgjørende er det at rådet er *riktig*. Det samme kan sies om *påvirkningskraft*; jo større sannsynligheten er for at et råd faktisk blir fulgt, jo større er gevinsten hvis rådet er *viktig*, og jo større katastrofe hvis rådet er *galt*.

Logisk sett er effekten av prosessen uavhengig av egenskapene ved rådene som gies. Det er imidlertid sannsynlig at en prosess som generelt fremskaffer nyttige råd, også vil virke positivt i retning av *kompetanseheving* og økt *forankring*.

7 Referanseliste

- Andre, T. S. (2000). "Determining the effectiveness of the usability problem inspector: A theory-based model and tool for finding usability problems." PhD Thesis. Blacksburg, VA: Virginia Polytechnic Institute and State University.
- Bias, RG & Mayhew, DJ (1994). "Cost-justifying usability." Boston: Academic Press.
- Cockton, G. & Woolrych, A. (2002) "Sale must End: Should Discount Methods be Cleared off HCI's Shelves?". *Interactions*, September – October, 13 – 18.
- Dutt, Johnson & Johnson (1994). "Evaluating evaluation methods". In Cockton, Draper & Weir (pp. 109-121). *People and computer IX*. Cambridge, UK: Cambridge University Press.
- Gray & Saltzman (1998). "Damaged merchandise? A review of experiments that compare usability evaluation methods." *Human-computer Interaction*. 13 (3): 203-261
- Gray, W. D., John, B. E., & Atwood, M. E (1992) "The précis of project Grace, or an overview of a validation of GOMS." *Proc. ACM CHI'92* (Monterey, CA, 3-7 May 1992), 307 – 312.
- Greenbaum, J. & Kyng, M., (Red), (1991) "Design at work: Cooperative Design of Computer Systems", Hillsdale NJ, Lawrence Erlbaum Associates.
- Gulliksen, J. et al. (2004) "Making a difference – a survey of the usability profession in Sweden." *Proceedings of the third Nordic conference on Human-computer interaction*, October 23-27, Tampere, Finland
- Hartson, H. R., Andre, T. S., Williges, R., C. (2001). "Ultimate criterion for UEM effectiveness - Finding real usability problems" *International Journal of Human-Computer Interaction*, December 2001, Vol. 13: 373-410
- Hertzum, M., Jacobsen, N. E. & Molich, (2002). "Usability Inspections by Groups of Specialists: Perceived Agreement in Spite of Disparate Observations". *CHI2002 Extended Abstracts*, page 662-663. ACM Press)
- John, B. E. & Marks, S. J. (1997). "Tracking the effectiveness of usability evaluation methods". *Behaviour & Information Technology*. 16 (4-5): 188-202.

- Karat, C-M. (1993) "Cost benefit and business case analysis of usability engineering." Tutorial presented at the ACM SIGCHI Conference on Human Factors in Computing Systems. New Orleans, LA, April 28-May 2. 1993..
- Karat, C-M. (1997). "Cost-justifying usability engineering in the software life cycle". In Helander, Landauer & Prabhu (Eds.). *Handbook of human-computer interaction*. Amsterdam: Elsevier Science.
- Lai-Chong Law, E., Hvannberg, E. T. (2002) "Complementarity and convergence of heuristic evaluation and usability test: a case study of universal brokerage platform". Proceedings of the second Nordic conference on Human-computer interaction, October 19-23, 2002, Aarhus, Denmark
- Lai-Chong Law, E., Hvannberg, E. T. (2004a) "Analysis of strategies for improving and estimating the effectiveness of heuristic evaluation". Proceedings of the third Nordic conference on Human-computer interaction, October 23-27, Tampere, Finland
- Lai-Chong Law, E., Hvannberg, E. T. (2004b) "Analysis of combinatorial user effect in international usability tests". Proceedings of the 2004 conference on Human factors in computing systems, p.9-16, April 24-29, 2004, Vienna, Austria
- Mantei, M. M. & Teorey, T. J. (1988). "Cost/benefit analysis for incorporating human factors in the software lifecycle", *ACM*, April 1988, 31, 4: 428-439.
- Molich, R., Ede, M. R., Kaasgaard, K. & Karyukin B. (2004). "Comparative usability evaluation". *Behaviour & Information Technology*, Vol 23, Number 1, January-February 2004, page 65-74.
- Nielsen, J. & Gilutz, S. (2003) "Usability Return on Investment", Nielsen Norman Group
- Nielsen, J. & Landauer, T. K. (1993). "A mathematical model of the finding of usability problems." *Proc. ACM INTERCHI'93 Conf.* (Amsterdam, The Netherlands, 24-29 April), 206 – 213.
- Nielsen, J. & Phillips, V., L. "Estimating the relative usability of two interfaces: Heuristic, formal, and empirical methods compared." (1993) *Proc. ACM INTERCHI'93 Conf.* (Amsterdam, The Netherlands, 24-29 April), 206 – 213.
- Nielsen, J. (1994) "Guerilla HCI: Using discount usability engineering to penetrate the intimidation barrier." In Bias, RG & Mayhew (eds) *Cost-justifying usability*. Boston: Academic Press.
- Vredenburg, Mao, Smith & Carey (2002) "A survey of User-Centered Design Practice". *Proc. CHI 2002*, (Minneapolis, Minnesota, USA. 20- 25 April) 471 – 478.
- Wixon, D. (2003). "Evaluating usability methods". *Interaction*, July-August: pp. 29-34.