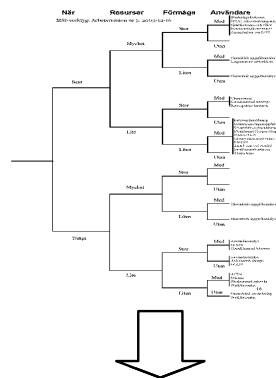




Jens Alfredson, Per-Anders Oskarsson, Martin Castor, Jonathan Svensson

Metodvalsverktyg

Ett hjälpmedel vid planering av MSI-utvärdering



Metod	METODEN																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Metod 1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Metod 20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

© 2004, FOI. Alla rättigheter förbehållna. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se

Metod	Metodens beskrivning
Metod 1	Metod 1: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 2	Metod 2: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 3	Metod 3: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 4	Metod 4: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 5	Metod 5: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 6	Metod 6: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 7	Metod 7: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 8	Metod 8: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 9	Metod 9: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 10	Metod 10: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 11	Metod 11: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 12	Metod 12: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 13	Metod 13: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 14	Metod 14: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 15	Metod 15: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 16	Metod 16: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 17	Metod 17: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 18	Metod 18: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 19	Metod 19: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se
Metod 20	Metod 20: Beskrivning av metodens syfte och mål. Detta dokument är ett arbetsredskap och ska inte användas utan tillstånd från FOI. För mer information se FOI:s webbplats www.foi.se

TOTALFÖRSVARETS FORSKNINGSPENSTITUT

Ledningssystem
581 11 Linköping

FOI-R--1295--SE

Augusti 2004

ISSN 1650-1942

Metodrapport

Jens Alfredson, Per-Anders Oskarsson, Martin Castor, Jonathan Svensson

Metodvalsverktyg

Ett hjälpmedel vid planering av MSI-utvärdering

Utgivare Totalförsvarets Forskningsinstitut - FOI Ledningssystem 581 11 Linköping	Rapportnummer, ISRN FOI-R--1295--SE	Klassificering Metodrapport
	Forskningsområde 8. Människa och teknik	
	Månad, år Augusti 2004	Projektnummer E70501, E7096
	Verksamhetsgren 5. Uppdragsfinansierad verksamhet	
	Delområde 81 MSI med fysiologi	
Författare/redaktör Jens Alfredson Per-Anders Oskarsson Martin Castor Jonathan Svensson	Projektledare Helena Holmström	
	Godkänd av Erland Svensson	
	Uppdragsgivare/kundbeteckning Försvarets materielverk och Försvarmakten	
	Tekniskt och/eller vetenskapligt ansvarig Jens Alfredson	
Rapportens titel Metodvalsverktyg - Ett hjälpmedel vid planering av MSI-utvärdering		
Sammanfattning (högst 200 ord) <p>Vid utveckling av nya försvarssystem är anpassning avseende människa system interaktion (MSI) av stor vikt. En prototyp till ett verktyg för att underlätta val av metoder vid MSI-utvärderingar har därför utvecklats. Syftet med metodvalsverktyget är i första hand att visa hur ett hjälpmedel kan utformas för FMV:s handläggare vid kontakter med exempelvis forskningsorganisationer, högskolor och industri vid upphandling av MSI-utvärderingar. Verktyget omfattar 26 metoder. Den som använder verktyget besvarar fyra frågor om utvärderingen. Om den utförs tidigt eller sent under systemutvecklingsprocessen? Om det finns små eller stora resurser tillgängliga? Om det är tillräckligt att de som genomför utvärderingen har allmän MSI-kompetens, eller om specialistkompetens krävs? Om utvärderingen kommer att göras med eller utan användare? Svaren på dessa frågor leder användaren genom ett trädformat diagram. Diagrammet avslutas vid ett löv där det anges vilka metoder som kan vara lämpliga för utvärderingen i fråga. En tillhörande tabell anger dessa fyra och ytterligare tio egenskaper för varje metod, vilket ger god överblick över olika metoders för- och nackdelar. För mer detaljerad information finns för varje metod en tabellformad redogörelse på 1-3 sidor, där en utförlig beskrivning av varje specifik egenskap ges. Metodvalsverktyget har provats vid en workshop. Det allmänna intrycket var att det har potential att bli ett värdefullt hjälpmedel för handläggare vid upphandling av MSI-utvärderingar.</p>		
Nyckelord utvärdering, metod, metodik, MSI		
Övriga bibliografiska uppgifter	Språk Svenska	
ISSN 1650-1942	Antal sidor: 116 s.	
Distribution enligt missiv	Pris: Enligt prislista	

Issuing organization FOI – Swedish Defence Research Agency Command and Control Systems SE-581 11 Linköping	Report number, ISRN FOI-R--1295--SE	Report type Methodology report
	Programme Areas 8. Human Systems	
	Month year August 2004	Project no. E70501, E7096
	General Research Areas 5. Commissioned Research	
	Subcategories 81 Human Factors and Physiology	
Author/s (editor/s) Jens Alfredson Per-Anders Oskarsson Martin Castor Jonathan Svensson	Project manager Helena Holmström	
	Approved by Erland Svensson	
	Sponsoring agency Swedish Defence Materiel Administration and Swedish Armed Forces	
	Scientifically and technically responsible Jens Alfredson	
Report title (In translation) Instrument for choice of methods - A means of assistance in planning of MSI-evaluation		
Abstract (not more than 200 words) <p>When new defence systems are developed, adaptation concerning MSI (Man System Interaction) is of great importance. Therefore, a prototype of an instrument to facilitate the choice of MSI-evaluation methods has been developed. The main purpose of the instrument is to demonstrate a model of an aid for FMV (Swedish Defence Materiel Administration) administrators to support them in their contact with research organizations, universities, and industry, when they are purchasing MSI-evaluations. The instrument includes 26 methods. The user of the instrument answers four questions about the evaluation. If it is performed early or late in the process of systems development? If the amount of available resources is large or small? If it is sufficient that contracted evaluators have general MSI-competence or if specific specialist competence is necessary? If users will be included in the evaluation or not? The answers to these questions guide the user through a tree shaped diagram. The diagram terminates in a leaf, where methods that may be suitable for the present evaluation are given. An adherent table gives the values of all methods for these four and further ten properties, which gives a good overview advantages and disadvantages of the different methods. For more detailed information of each method, each method is separately described in a table of 1-3 sides, where a detailed description of each specific property is given. The instrument has been tested in a workshop. The general view was that it has the potential to become a valuable aid for administrators in their processes of purchasing MSI-evaluations.</p>		
Keywords evaluation, method, methodology, MSI, Human Factors		
Further bibliographic information	Language Swedish	
ISSN 1650-1942	Pages 116 p.	
	Price acc. to pricelist	

Innehållsförteckning

INLEDNING.....	1
1 BAKGRUND.....	2
1.1 TIDIGARE ANSATSER.....	2
1.2 UTVÄRDERINGSMETODIK.....	3
1.2.1 Designmodeller.....	3
1.2.2 Olika typer av studier.....	5
1.2.2.1 Kvantitativa respektive kvalitativa metoder?.....	6
1.2.2.2 Fältförsök.....	6
1.2.2.3 Laboratoriestudier.....	6
1.2.2.4 Simuleringar.....	7
1.2.3 Statistiska metoder för datareducering och modellering.....	7
1.2.4 Regler för arbete med användargrupper.....	8
1.2.4.1 Inom- eller mellangrupsdesign.....	8
1.2.4.2 Kategori av försökspersoner.....	9
1.2.4.3 Etik.....	9
2 LÄSANVISNINGAR.....	11
2.1 FYRA FRÅGOR.....	11
2.2 OMVÄNDA CUP-VERKTYGET.....	11
2.3 METODERNAS ORDNING.....	11
2.4 STORA MATRISEN.....	12
2.5 METODBESKRIVNINGAR.....	12
3 OMVÄNDA CUP-VERKTYGET.....	14
4 STORA MATRISEN.....	16
5 METODBESKRIVNINGAR.....	17
5.1 BLINKNINGSFREKVENNS.....	17
5.2 EPOG OCH AVSÖKNINGSMÖNSTER.....	20
5.3 HJÄRTFREKVENNS OCH HRV.....	23
5.4 KOMMUNIKATIONSANALYS.....	26
5.5 ÖGONRÖRELSER VIA EOG.....	29
5.6 HIERARKISK UPPGIFTSANALYS INKLUSIVE ”DECOMPOSITION METHODS”.....	32
5.7 LOGGNING AV INTERAKTION.....	36
5.8 OBSERVATION.....	39
5.9 OSTRUKTURERAD INTERVJU.....	42
5.10 RETROSPEKTIV TESTNING.....	45
5.11 ENKÄTUNDERSÖKNING.....	48
5.12 GEMENSAM LÄGESUPPFATTNING.....	52
5.13 KVANTITATIVA PRESTATIONSMAÅTT.....	56
5.14 MODIFIERAD COOPER-HARPERSKALA.....	59
5.15 NASA-TLX.....	62
5.16 OBSERVATION MED VIDEO.....	65
5.17 SAGAT.....	68
5.18 SART OCH CC-SART.....	71
5.19 STRUKTURERAD INTERVJU.....	74
5.20 TÄNKA HÖGT.....	77
5.21 ANVÄNDARANALYS.....	81
5.22 GOMS.....	84
5.23 AXIOMATISK DESIGN.....	88
5.24 WALKTHROUGHS.....	91
5.25 ACTA.....	94
5.26 HEURISTISK UTVÄRDERING.....	97
6 REFERENSER.....	101
APPENDIX. ORDLISTA OCH BEGREPPSFÖRKLARINGAR.....	107

Inledning

Det är viktigt att beakta MSI-frågeställningar på ett strukturerat sätt för att säkerställa en kostnadseffektiv och rationell utveckling av framtida system. En förutsättning för detta är att MSI-frågeställningar entydigt beaktas tidigt i beställningsdialogen. Det är endast om man explicit fokuserar på dessa frågeställningar som de kan få möjlighet att bidra till systemutveckling på ett effektivt sätt. Detta dokument har därför föregåtts av inventering, selektering och strukturering av lämpliga metoder för att ta fram ett MSI-verktyg för utvecklingsprocessen. Dokumentet är till för att ge en översikt över metoder för MSI-analys, för att fungera som ett MSI-verktyg, exempelvis vid en beställningsdialog. De valda metoderna utgör exempel på metoder som kan ingå i en sådan beställningsdialog och sammanställningen gör således inga som helst anspråk på att vara en komplett samling av metoder inom området. Metodbeskrivningarna är skrivna i syfte att orientera läsaren om metodens egenskaper, fördelar och nackdelar. Metodbeskrivningarna är därmed inte avsedda som en lathund för den som skall genomföra utvärderingar med metoderna, utan snarare som en "informationskarta" för att underlätta för en handläggare/beställare av teknisk systemutveckling att hitta rätt utvärderingsmetoder. De metoder som är sammanställda här ska inte heller ses som rekommenderade metoder, framför metoder som inte är med i sammanställningen, utan endast som exempel på metoder.

Arbetet har avgränsats mot MSI-utvärderingsmetoder som studerar prestation, perception, kognitiva processer, mental arbetsbelastning, situationsmedvetande och kommunikation, med avseende på utveckling av lättanvända system med låg belastning på operatören. Hänsyn till dessa faktorer är nödvändig för utveckling av säkra system, med minimerad risk för felgrepp och missförstånd som försvårar arbetet eller orsakar olyckor.

För att system ska vara lätta att använda är även frågor inom ergonomi väsentliga, exempelvis att operatörens arbetsställning inte ger ryggbesvär eller andra fysiska problem. Även andra arbetsmiljörelaterade problem, som buller och kemiska hälsorisker har betydelse för användbarhet. Att även inkludera dessa faktorer i MSI-verktyget är knappast görligt, eftersom verktyget då skulle bli alltför omfattande och svårarbetat. Att parallellt utveckla ett motsvarande verktyg för val av utvärderingsmetod avseende ergonomi eller arbetsmiljö vore dock en möjlig väg.

Syftet är inte att beskriva samtliga förekommande MSI-metoder, utan att exemplifiera utvärderingsmetoder, eller kategorier av utvärderingsmetoder. Det finns exempelvis ett stort antal varianter av uppgiftsanalyser. Kirwan och Ainsworth (1992) beskriver t.ex. 41 typer av uppgiftsanalyser, varav de anger 25 som särskilt betydelsefulla. Att med hjälp av de fyra frågorna kunna bedöma vilken av dessa uppgiftsanalyser som är mest lämpad för en enskild utvärdering vore ogörligt. Hierarkisk uppgiftsanalys och ACTA har därför tagits med i MSI-verktyget som exempel på typiska metoder för uppgiftsanalys.

Då man kommit fram till att uppgiftsanalys bör genomföras, är nästa steg därför att bedöma vilken typ av uppgiftsanalys som är mest lämpad. Eftersom flertalet metoder för uppgiftsanalys är besläktade kan vilka metoder utvärderaren bäst behärskar vara avgörande för val av metod. Motsvarande förhållande gäller även ett flertal andra metoder, t.ex. kvantitativa prestationsmått, där det finns ett otal olika mätmetoder och variabler att mäta på.

1 Bakgrund

Anledningen till att arbetet med att ta fram ett MSI-verktyg initierades var att det fanns behov av stöd för att beakta MSI-frågeställningar strukturerat. Ett verktyg efterfrågades i beställningsdialogen för teknisk systemutveckling. Verktyget avsågs utgöra ett urval av MSI-metoder, utan ambition att vara komplett. Verktyget skulle därmed kunna användas för att underlätta val av metod såväl som vid värdering av ett projekts val av metod. På så vis kan beställningsdialogen lyftas från ett metodplan till behovsplan. Avsikten är därmed att undvika att metoder används av slentrian och att dialogen innehåller en analys av vilka behov som alternativa metoder kan tänkas fylla. I det att verktyget översiktligt beskriver de ingående metoderna kan det också fungera som stöd för handläggare som inte har kännedom om en viss metod, sedan tidigare.

Riktlinjer och checklistor kan vara till stor hjälp vid systemutveckling som stöd för en MSI-värderingsprocess. Det är dock nödvändigt att observera att MSI-problem inte kan lösas med enbart riktlinjer och checklistor. Enligt Sanders och McCormic (1993), existerar det inte någon checklista eller riktlinje som, om den så följdes till punkt och pricka, skulle kunna garantera en bra MSI-produkt. De menar att kompromisser, hänsyn till den specifika tillämpningen och vetenskapligt grundade bedömningar är exempel på sådant som inte kan uppfångas av riktlinjer eller checklistor.

1.1 Tidigare ansatser

Före andra världskriget bestod MSI-värderingar mestadels av undersökningar om huruvida en person var kapabel att sköta en maskin eller inte. Misslyckades personen testades en ny, osv. tills någon tillräcklig duglig hittades (Meister & O'Brien, 1996). I och med att maskiner blev mer och mer komplicerade ökade kraven på människan, vilket minskade möjligheten att genom enkla urvalsprocesser hitta tillräckligt många och tillräckligt dugliga operatörer. Därmed väcktes insikt om behovet av att börja anpassa maskiner efter människans förmåga.

En mycket stor del av mänskliga felhandlingar och olyckor med komplexa system har dålig systemdesign som underliggande orsak (Price, 1990). Det är dock inte ovanligt att denna typ av olyckor skylls på den mänskliga faktorn. Price menar att om olyckor kan orsakas genom dålig design, så kan olyckor också förebyggas på designstadiet, förutsatt att MSI frågor beaktas, så att systemets design utgår från människans förutsättningar och begränsningar.

Inom militär verksamhet kan minimering av skador på personal och materiel vara helt avgörande. MSI-forskning som syftar till att förbättra operatörens förmåga att snabbt tillgodogöra sig information från tekniska system, och därmed fatta riktiga beslut, har följaktligen stor relevans. Inom militära organisationer har därför olika ansträngningar gjorts för att få in MSI-aspekter i systemutvecklings- och anskaffningsfaser.

Under mitten av 1980-talet började den amerikanska armén att utveckla MANPRINT (MANpower and Personnel INtegration). Syftet med MANPRINT är att försöka påverka systemdesign på ett sådant sätt att användning och underhåll av materiel och informationssystem sker så kostnadseffektivt som möjligt, men i överensstämmelse med människans förutsättningar och begränsningar (MANPRINT, 2000).

Inom NATO har en forskningsgrupp (RSG-24, Research Group-24), inom Forsknings- och Teknikorganisationen (RTO, Research and Technology Organisation), utarbetat riktlinjer avseende utvärderingsmetodik för MSI-frågor vid systemutveckling. Deras målsättning är att utarbeta standardiserade riktlinjer för genomförande av testning och utvärdering. Utarbetade riktlinjer från denna forskningsgrupp har publicerats av Geddie m.fl. (2001). Inom NATO har senare RTO forskningsgruppen TG-008 utarbetat riktlinjer för värdering av operatörens funk-

tionella tillstånd (Operator Functional State Assessment), publicerat av Wilson m.fl. (2003). Detta arbete ger en omfattande genomgång, dels av faktorer som negativt påverkar operatörens funktionella förmåga, men även av metodik för att värdera dessa faktorer påverkan. Exempel på behandlade faktorer är dels miljöfaktorer som buller, acceleration och hetta, men även interna fysiska och psykologiska faktorer som sjukdom, sömnbrist, störd dygnsrytm, droger och medicin, mental, kognitiv och fysisk belastning.

Inom GARTEUR (Group for Aeronautical Research and Technology in Europe, en ramverksorganisation för europeiska flygforskningsaktörer) har riktlinjer för bedömning av metoder för att mäta mental arbetsbelastning utarbetats inom gruppen GARTEUR FM AG13 Action Group on Mental Workload Measurement. Nio forskningsinstitut och flygindustrier har deltagit i gruppens arbete som pågått 1999-2003. FOI har ansvarat för ordförandeskapet i gruppen. Slutrapport från gruppens arbete har publicerats under 2003 (Castor m.fl., 2003). Idén med metodmatriser och inspiration till föreliggande rapports metodik för att välja metoder har hämtats från GARTEUR.

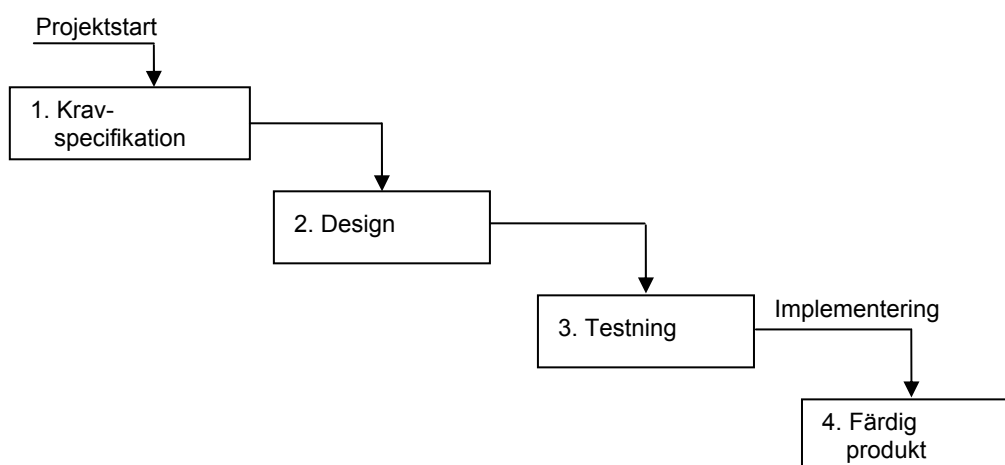
En likartad civil ansats är exempelvis Internationella Standardiseringsorganisationen (ISO, International Standardisation Organisation) som publicerat standardiserade riktlinjer för designprocessen vid systemutveckling utgående från människans behov och förutsättningar (ISO 13407, 1999).

1.2 Utvärderingsmetodik

1.2.1 Designmodeller

Traditionen inom mjukvaruutveckling har varit att använda ett sekventiellt designförlopp. Ofta har vattenfallsmodellen använts (Figur 1). Vattenfallsmodellen beskriver designförloppet som ett sekventiellt skede som börjar med att systemets tillämpningar beskrivs, kravspecifikationer görs, systemet designas och slutligen implementeras. Enligt Preece (1994) är ett stort problem med ett sekventiellt förlopp att det är nästintill omöjligt att helt förstå användarnas krav innan en stor del av designarbetet har genomförts.

I praktiken är vattenfallsmodellen dock inte helt linjär. Vissa aktiviteter kan vara överlappande. Iterationer och tillbakagång till tidigare utförda aktiviteter kan även förekomma (Preece, 1994).

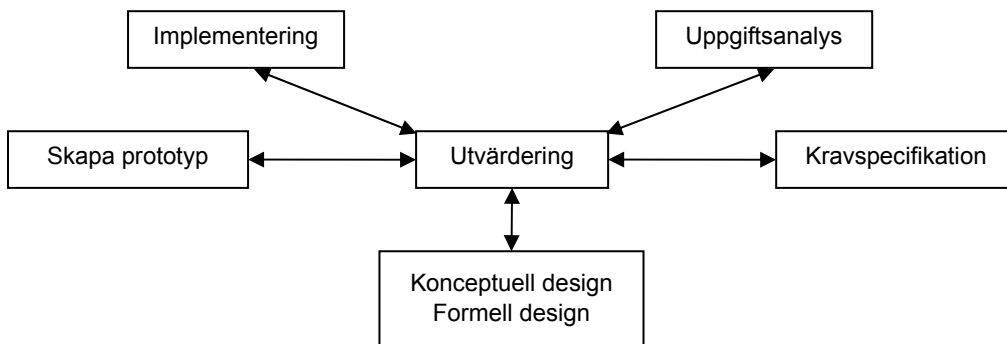


Figur 1. Vattenfallsmodellen, en sekventiell designmodell.

En tydligare betoning på utvärderingar som iterativa förlopp är dock önskvärt, där resultat från genomförda MSI-utvärderingar återkopplas till designers/konstruktörer som då gör om gränssnittet/systemet enligt utvärderingens resultat. Och så vidare, till gränssnittets/systemets prestanda motsvarar uppställda MSI-krav.

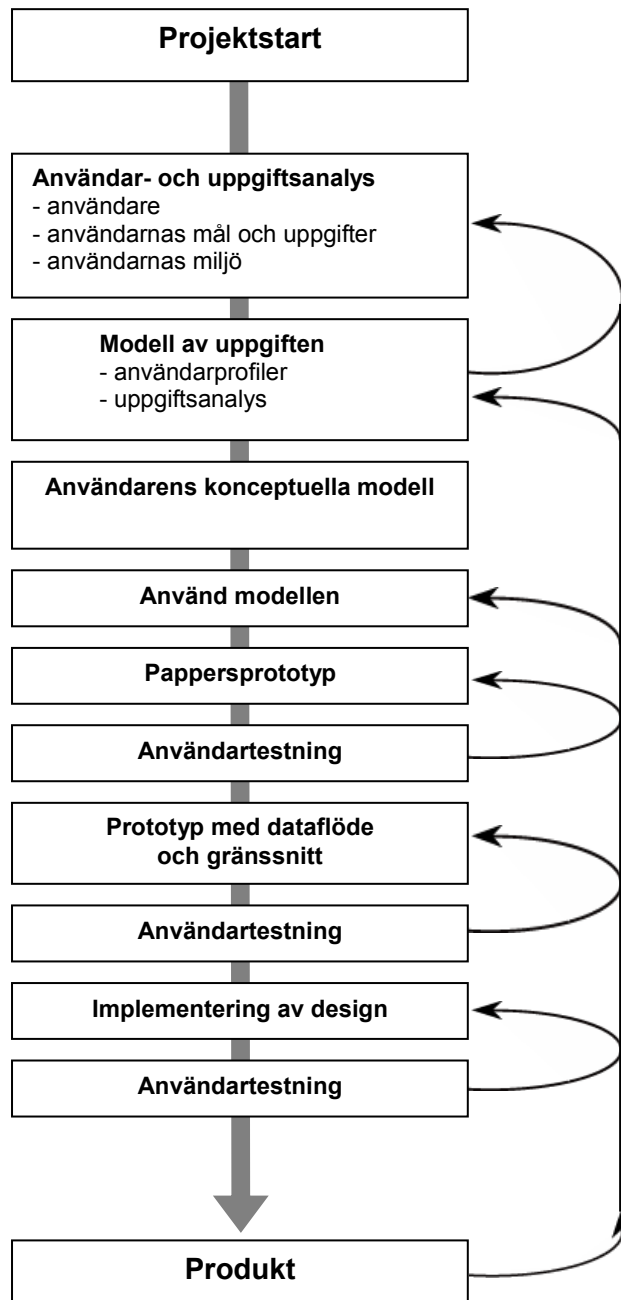
Inom mjukvarudesign används ibland spiralmodellen för att beskriva ett iterativt designförlopp. Enligt spiralmodellen sker utveckling/design i flera cykler, vilket kan beskrivas som att designprocessen börjar längst inne i en spiral, och blir mer och mer omfattande för varje cykel (varv i spiralen). Varje cykel består av olika analyser (kravspecifikationer), skapande av prototyp och utvärderingar av denna. I nästa cykel genomförs nya analyser, en mer förfinad prototyp skapas, nya utvärderingar genomförs, och så vidare tills tillräckligt många cykler genomlöpts för att ett funktionellt system skall kunna implementeras och testas (Oskarsson, 2002). Modellen bygger, enligt Preece (1994) m.fl., på att varje cykel i spiralen, där en prototyp utvecklas, syftar till att öka förståelse och krav inför nästa stadium i designcykeln där nästa prototyp utvecklas.

En alternativ beskrivning av en iterativ designmodell är "Star life cycle" (Figur 2). Dess mest centrala komponent är utvärderingar, vilka betraktas som relevanta under samtliga utvecklingsstadier, till skillnad mot vattenfallsmodellen där utvärderingar tenderar att begränsas till slutet av utvecklingskedjan (Preece, m.fl., 1994).



Figur 2. Schematisk beskrivning av "Star life cycle", en iterativ designmodell med utvärderingar i centrum (anpassad efter Preece m.fl., 1994, hämtad från Oskarsson, 2002).

Ett mer generellt iterativt designförlopp, anpassat från Hackos och Redish (1998), för hela utvecklingskedjan, med vikt på tidiga användar- och uppgiftsanalyser exemplifieras i figur 3. Eftersom det är mer kostnadseffektivt att modifiera prototyper än fullskaliga system, är iterativa designprocesser särskilt viktiga före implementering av färdiga system (Liu, 1997). Originalfiguren av Hackos och Redish (1998) har inte någon pil bakåt från den färdiga produkten. I figur 3 har en bakåtpil från den färdiga produkten lagts till för att åskådliggöra att även färdiga produkter kan utvärderas för att genomföra modifieringar, eller för att kontrollera att systemets prestation motsvarar uppställda rikt- och säkerhetsvärden.



Figur 3. Schematisk beskrivning av designprocessens iterativa livscykel, med användaranalys, uppgiftsanalys och utvärderingar (anpassad efter Hackos & Redish, 1998, hämtad från Oskarsson 2002).

1.2.2 Olika typer av studier

Det finns olika typer av metodik för att genomföra MSI-utvärderingar vid systemutveckling. Ett exempel är användning av kvantitativa eller kvalitativa metoder. Om man experimentellt vill studera isolerade delar av system, eller interaktionsprinciper används ofta laboratoriestudier. Om man vill testa system, eller komponenter av system, under naturliga arbetsförhål-

landen med riktiga användare, genomförs vanligen fältförsök. Vid simuleringar kan fördelar med fältförsök och laboratoriestudier kombineras. Litteraturstudier är viktiga för teoretisk bakgrund av respektive ämnesområde. Exempelvis för information om redan genomförd försöksverksamhet och områdets teknikutveckling.

1.2.2.1 Kvantitativa respektive kvalitativa metoder?

Kvantitativa metoder avser vanligen automatisk insamling av prestationsmått, t.ex. genom loggning av antal fel, precision, hastighet eller tid. Kvantitativa prestationsmått förutsätter i regel att användaren/försökspersonen utför handlingar som ger någon typ av ”yttre” registrerbar respons. Enligt Meister (1996) anser vissa MSI-forskare att kvantitativa mått är tillräckligt för att studera mänskligt beteende. Meister menar dock att detta synsätt kan vara problematiskt, eftersom det tenderar att betrakta människan som en ”svart låda” där endast ”stimulus input” och ”respons output” studeras, medan vad som sker inne i den svarta lådan antingen ses som oväsentligt eller som korrelerat med stimulus och respons.

Kvalitativa metoder avser vanligen mänskligt insamlade mått, exempelvis genom observation eller genom insamling av användarnas egna synpunkter. Dessa metoder kan vara nödvändiga för att komma åt inre mentala processer hos användaren vid kognitiva arbetsuppgifter, exempelvis problemlösning. Även denna typ av metoder kan dock kritiseras. Enligt Meister (1996) är en invändning att användare kanske inte alltid vet vad de har tänkt på, och att processen att språkligt formulera sina handlingar kan påverka prestation och beteende.

Med psykofysiologiska mätningar, exempelvis av hjärtfrekvens, variabilitet i hjärtfrekvens eller blickregistrering går det få kvantitativa mått som avspeglar kognitiva processer som mental arbetsbelastning och situationsmedvetande.

Bäst kan vara att följa en medelväg där data med olika typer av metoder kombineras, vilket innebär att fördelar från samtliga typer av utvärderingsmetoder kan tillvaratas.

1.2.2.2 Fältförsök

Fältförsök avser studier i naturliga arbetsmiljöer, t.ex. på arbetsplatser. Inom försvaret kan detta avse systemets naturliga arbetsmiljö, exempelvis studier av besättningar på flygplan, fartyg eller stridsfordon som framförs under fältmässiga förhållanden, eller studier av ledningscentraler och staber under realistiska övningsförhållanden.

Vid fältförsök är kontroll av yttre faktorer är vanligen låg, dvs. låg reliabilitet. Fördelen är dock att systemet testas i en situation som är typisk för systemets riktiga användning, vilket ger studien god extern validitet. Detta är värdefullt eftersom vissa problem endast uppkommer vid naturlig användning av ett system.

Experimentella fältförsök, där vissa variabler kontrolleras, vilket ger viss kontroll över experimentsituationen, kan ses som ett mellanting mellan fältförsök och laboratoriestudier. Detta kan ibland vara nödvändigt för att det under fältmässiga förhållanden skall vara möjligt att studera vissa frågeställningar. Exempelvis kan de situationer man vill studera uppträda naturligt så sällan att de måste genereras av forskaren för att kunna studeras. Det kan även vara nödvändigt att eliminera vissa störvariabler.

1.2.2.3 Laboratoriestudier

För experimentella studier av isolerade systemkomponenter, eller generella interaktionsprinciper används vanligen laboratoriestudier. Man kan då konstanthålla, eller balansera ut, oberoende störvariabler som kan påverka de beroendevariabler man mäter på. Problemet är att denna kontroll kan ge en onaturlig arbetssituation, vilket är problematiskt för extern validitet. Väl genomförda laboratoriestudier har som regel god reliabilitet.

1.2.2.4 *Simuleringar*

Simuleringar omfattar ett flertal olika tekniker, från fullständigt datorbaserade simuleringar av hela system, till enkla pappersprototyper som endast avspeglar utvalda delar av ett system. Simuleringar kan användas för ett flertal ändamål, exempelvis för ergonomi och design av kontrollfunktioner, identifikation av möjliga felkällor, för utprovning av lämpliga arbetsmetoder vid systemanvändning, eller för träning (Kirwan & Ainsworth, 1992).

Simulatorer har först och främst utvecklats för träning och utbildning. De är därför viktiga instrument för utveckling och värdering, analys av träningseffekter, samt för uppbyggnad av träning och prestationsvärdering. Man skiljer ofta på tränings- och forskningssimulatorer, vilka blivit centrala forskningsplattformar, där dagens flygsystem i huvudsak utvecklas (Angelborg-Thanderz, 1990).

Simulatorer medger ofta testning av ett system i en naturlig arbetsmiljö, men med bibehållen kontroll. Bland de första tillämpningarna av VR-teknik (Virtual Reality) var flygsimulatorer, för träning av både trafikflygare och militära piloter (Bullinger, Bauer & Braun, 1997). För flygforskning finns därför ofta befintliga simulatorer att tillgå.

Vid systemutveckling kan simulatorer användas, exempelvis för utprovning av ny instrumentering och utrustning (ex., Derfeldt, m.fl., 1999; Nilsson & Nåbo, 1996). Simulatorer ger också goda möjligheter att genomföra loggning, psykofysiologiska och subjektiva mätningar, exempelvis för analys av hur en flyguppgift påverkar pilotens arbetsbelastning och situationsmedvetande i förhållande till prestation (ex., Berggren, 2000; Svensson m.fl., 1997b; Svensson, Angelborg-Thanderz & Wilson, 1999).

En särskild typ av simulering är ”Wizard of Oz”, vilket innebär att användaren arbetar med vad han/hon tror är ett helt datoriserat system. I själva verket är det en människa som från en annan plats/terminal simulerar systemet, eller delar av det, genom att mata in relevanta svar till användaren. Enligt Faulkner (2000) är detta bl.a. fördelaktigt eftersom systemet inte behöver konstrueras för att testas, vilket är särskilt användbart vid utveckling av system som skall arbeta med naturligt språk.

Mycket enkla simuleringar kan genomföras med prototyper som endast simulerar vissa delar av färdiga system. Avsikten kan vara att demonstrera systemets funktioner eller att utföra en tidig utvärdering före implementering.

Prototyper som endast modellerar vissa utvalda delar av ett system med fullständig funktionalitet, kallas *vertikala prototyper*. Prototyper som modellerar hela systemet med reducerad funktionalitet kallas *horisontella prototyper* (Nielsen, 1993). Fördelen med vertikala prototyper är att modellerade delar av systemet kan testas med riktiga användaruppgifter. Med horisontella prototyper testas i stället hela system, men med lägre realism eftersom användaren inte kan utföra riktiga arbetsuppgifter. Nielsen menar dock att horisontella prototyper kan ge en bra bild av hur systemet ”hänger ihop” och hur det uppfattas som en helhet.

Vissa enklare utvärderingar kan även göras med pappersprototyper. Dessa är främst användbara för utvärdering av skärmbilder vid människa-dator-interaktion. Detta är en enkel och billig metod för att tidigt bedöma ett flertal olikartade designförslag. Att tidigt arbeta med flera alternativ är fördelaktigt, eftersom det kan finnas en benägenhet att hålla fast vid den designidé som valts vid projektets start (Hackos & Redish, 1998).

1.2.3 *Statistiska metoder för datareducering och modellering*

Korrelationsanalys, Faktoranalys och ”andra generationens” Multivariata analyser, vilka inbegriper Strukturell ekvationsmodellering och Multidimensionell skalning, är viktiga metodologiska verktyg inom beteendevetenskaplig forskning.

Multivariata analyser används för analys av multipla relationer och för att möjliggöra experimentell design under tillämpade förhållanden, t.ex. fältstudier, och är nödvändiga verktyg för modellering av mänskligt beteende. Multivariata analyser baseras på korrelationsanalys, och visar graden av samband mellan studerade variabler som mängd förklarad varians.

Faktoranalys (FA) är, utan jämförelse, den mest använda metoden för datareducering, och utgör grund för besläktade metoder som Klusteranalys, Multidimensionell skalning, Strukturell ekvationsmodellering och Linjär strukturell ekvationsmodellering (LISREL). Faktoranalys är en analytisk metod som möjliggör reducering av ett stort antal relaterade ”tydliga” och mätbara variabler (manifest variables) till ett mindre antal ”dolda” variabler (latent variables). Syftet med Faktoranalys kan sägas vara att påvisa de faktorer som ligger bakom den gemensamma variansen mellan mätbara variabler. Klassisk Faktoranalys omfattar en explorativ och en hypotestestande procedur.

Linjär strukturell ekvationsmodellering (LISREL) är en användbar metod för att bekräfta och dra generella slutsatser om strukturer och kausalitet i Faktoranalys. LISREL kan användas för både explorativ och bekräftande Faktoranalys. Metoden kännetecknas av två grundläggande beståndsdelar: en strukturell modell och en mätmodell. Den strukturella modellen är en ”path” modell, vilken relaterar oberoende variabler till beroende variabler. Mätmodellen ger den maximala sannolikheten för att Faktoranalysen anger relationerna mellan ”tydliga” (manifest) och ”dolda” (latent) variabler.

Multidimensionell skalning (MDS) är en metod för att passa in en mängd variabler i en rymd, eller i ett plan, så att avståndet mellan variablerna så väl som möjligt avspeglar givna olikheter och skillnader mellan dem. Vanligtvis ger Multidimensionell skalning en modell med färre dimensioner än Faktoranalys.

För vidare läsning om datareducering och modellering, se Svensson (2003) vilket legat till grund för ovanstående beskrivning.

1.2.4 Regler för arbete med användargrupper

När testning genomförs är det viktigt att nödvändiga och gällande säkerhetsföreskrifter upprättas och följs. Den som planerar genomförandet av ett test bör därför upprätta en checklista avseende säkerhetsföreskrifter vid testning (O’Brien, 1996). Detta är synnerligen viktigt vid fältmässiga utvärderingar av militära aktiviteter, vilka kan medföra risk både för deltagare och forskare som genomför utvärderingar.

1.2.4.1 Inom- eller mellangrupsdesign

Inomgruppsdesign innebär att samtliga försökspersoner/användare genomför samtliga moment av ett test, t.ex. testar samtliga versioner av en prototyp. Detta är fördelaktigt eftersom man undviker problemet att eventuella skillnader i resultatet kan bero på skillnader mellan testgrupperna. Problematiskt är att försökspersonerna ofta lär sig något av det första testet som påverkar resultatet på nästa test, och så vidare. Detta problem kan dock vanligen kontrolleras genom balansering, vilket innebär att olika försökspersoner genomför testmomenten i olika ordning.

Mellangrupsdesign innebär att det är olika försökspersoner för olika delar av ett test. I vissa typer av studier kan det vara omöjligt att använda inomgruppsdesign (samma försökspersoner i olika moment). Exempelvis om påverkan av att ha genomgått ett testmoment är så stor att resultatet av nästa moment påverkas allt för mycket av det första. Fördelen med mellangrupsdesign är att varje försöksperson är ”naiv” vid försökets början. Ofta finns dock stora individuella skillnader mellan olika försökspersoner, varför det vid mellangrupsdesign är

nödvändigt att randomisera (slumpmässigt skapa) försöksgrupperna så att de blir så lika som möjligt. Det behövs också fler försökspersoner än vid inomgruppsdesign.

Mellangrupsdesign måste givetvis användas när man studerar skillnader mellan olika typer av användargrupper (t.ex. nybörjare – experter). Vid fältstudier, exempelvis stabsövningar, kan man vara intresserad av skillnader mellan olika typer av befattningar. I dessa fall är det skillnaderna mellan grupperna som studeras, varför man av naturen arbetar med olikartade grupper.

1.2.4.2 Kategori av försökspersoner.

Genom användaranalys bedöms vilken kategori av människor som kommer att använda systemet. En huvudregel vid systemutvärdering är att försökspersoner bör vara så representativa för slutanvändarna som möjligt (Nielsen, 1993). Om företaget eller organisationen där ett system skall användas tillhandahåller försökspersoner finns dock risk att de, för att göra ett så positivt intryck som möjligt, väljer ut de skickligaste eller snabbaste användarna.

Om utvärderingar görs med studenter som försökspersoner bör man ställa sig frågan om systemet även kommer att användas av äldre människor (Nielsen, 1993). Om så är fallet, bör även äldre inkluderas i utvärderingen.

Om försökspersonerna är ”nybörjare” och systemet som utvärderas använder nya interaktionstekniker som de saknar erfarenhet av, kan det vara nödvändigt att innan ett försök genomförs ge dem träning. Detta visade sig bl.a. när Windows-baserade operativsystem var relativt nya och många saknade vana vid arbete med mus. Det hände att jämförelser mellan system med Dos- och Windows-gränssnitt, utföll till fördel för Dos-baserade system för otränade försökspersoner, och tvärt om för vana musanvändare (Nielsen, 1993).

1.2.4.3 Etik

När utvärderingar genomförs med försökspersoner eller användare är det försöksledarens ansvar att etiska regler följs. Ett problem är att användare kan känna sig illa till mods när deras fel observeras eller registreras. Vid vissa försök, och i synnerhet vid laboratoriestudier, kan uppgifternas svårighetsgrad vara hög, beroende på att det är felhandlingar och problem som man vill studera. Vid utvärderingar är det också vanligt att försökspersoner gör många fel, beroende på att systemets design är dålig. För att minimera risken att försökspersonen känner sig inkompetent bör därför försöksledaren alltid noggrant påpeka att det är systemet som utvärderas och inte användaren eller försökspersonens prestation.

Om försökspersoner på något sätt blir lurade under ett försök, som exempelvis vid utvärderingar med ”Wizard of Oz”, där en försöksledare imiterar vissa funktioner hos ett system, bör försöksledaren informera försökspersonerna om detta innan försöket inleds. Om detta inte kan göras, exempelvis om det finns starka skäl att tro att kunskap om att en människa utför vissa av systemets funktioner kan påverka försökspersonens prestation, skall försöksledaren informera försökspersonerna om detta efter försökets genomförande.

Nielsen (1993) ger bland annat följande riktlinjer för etiska hänsynstaganden under utvärdering:

Före testning:

- Betona att det är systemet som utvärderas och inte försökspersonen.
- Tala om att systemet är nytt och oprövat, och därför kan ha brister.
- Tala om att försökspersonerna har rätt att avsluta testet när de själva önskar.
- Förklara anledningen till varje inspelning, logging eller annan registrering som sker.
- Tala om för försökspersonen att testresultatet kommer att behandlas konfidentiellt (och sedan naturligtvis också behandla det konfidentiellt).

Under testning:

- Försök se till att försökspersonen lyckas med några uppgifter i början av testet.
- Se till att testet sker i en avslappnad atmosfär, bjud på kaffe eller se till att det finns möjlighet till pauser.
- Antyd aldrig på något sätt att försökspersonen gör fel eller är för långsam.
- Minimera antalet observatörer.
- Om testet sker på försökspersonens arbetsplats, tillåt inte att någon från företagsledningen kan observera testet.
- Avbryt testet om det av någon anledning blir för obehagligt för försökspersonen.

Efter testning:

- Avsluta med att framhålla att försökspersonen varit till hjälp, exempelvis genom att ha hittat saker som behöver förbättras.
- Rapportera aldrig resultaten så att individuella försökspersoner kan identifieras.
- Visa aldrig videoinspelningar utanför utvärderingsgruppen, såvida försökspersonen inte gett sitt tillstånd.

Etiska riktlinjer för försök inom psykologi har utarbetats av American Association of Psychology (1953, 2002) Den första versionen av dessa riktlinjer publicerades 1953 och har sedan utkommit i nya upplagor vid ett flertal tillfällen. Den senaste versionen, från 2002, finns för närvarande tillgänglig på American Association of Psychology's hemsida, även publicerad i *American Psychologist* (2002). Vissa av dessa riktlinjer är även relevanta vid MSI-utvärderingar som vägledning för arbete med användargrupper.

Det finns dock tillfällen då det kan vara problematiskt att följa uppställda etiska riktlinjer. Enligt ovanstående riktlinjer skall försöksledaren före ett försök alltid tala om vad studiens syfte är. Enligt Goodwin (1995) har exempelvis försök genomförts där bullerpåverkan på koncentrationsförmåga och prestation studerats. Försökspersoner som var ovetande om studiens syfte påverkades, enligt gällande teori, med sämre koncentrationsförmåga och prestation. Försökspersoner som på förhand informerades om studiens syfte påverkades dock inte negativt av buller. Den forskare som genomför en utvärdering bör därför alltid göra en bedömning av hur försökspersoner kan påverkas av information om försökets syfte.

Enligt American Association of Psychology (2002) skall även debriefing göras efter ett genomfört försök. Om information om försökets syfte undanhållits skall försöksledaren nu informera försökspersonerna om detta och beskriva försökets egentliga syfte. Vid debriefingen skall försöksledaren också bemöda sig om att ingen försöksperson känner sig illa till mods efter försöket. Vid systemutvärdering kan försökspersoner t.ex. känna sig nedstämda och förebrå sig själva om de tycker att de gjort många fel, eller presterat dåligt. Det kan därför inte nog poängteras att det i så fall är försöksledarens skyldighet att framhålla att det var systemet som testades och inte försökspersonen. Om studiens design var sådan att försökspersonerna skulle göra mycket fel, eller om felhandlingar berott på ett dåligt gränssnitt, skall detta också framhållas.

2 Läsanvisningar

De MSI-utvärderingsmetoder som beskrivs i rapporten är utvecklade för att möta olika typer av designproblem och för att användas i olika stadier av en designprocess. Metoderna skiljer sig också med avseende på vilken mängd av förkunskaper som behövs för att kunna använda sig av metoden.

2.1 Fyra frågor

Tanken bakom verktyget är att en handläggare med hjälp av fyra enkla frågor med binära svar, som han/hon oftast kan besvara redan tidigt i designprocessen, skall kunna begränsa antalet metoder som behöver studeras närmare. De fyra frågor som används för att navigera i ”omvända cup-verktyget” är:

1. *När?* När i designprocessen är det tänkt att MSI-utvärderingen skall genomföras? Tidigt innebär att det inte finns någon fungerande prototyp av det aktuella systemet. Sent innebär att det finns någon slags ”körbar” prototyp, simulering eller motsvarande att utvärdera.
2. *Resurser?* Frågan handlar om hur stora resurser som finns tillgängliga (alt. resurser som krävs) för att genomföra MSI-utvärderingen. Resurser ses i vid mening och omfattar både tid och utrustning.
3. *Förmåga?* Frågan handlar om hur stora förkunskaper, till exempel inom MSI-området, domänkunskap eller formell kompetens som den som utför värderingen behöver för att kunna använda metoden. Stor innebär att utföraren har någon typ av specialiserade kunskaper medan liten motsvarar allmän MSI-kompetens.
4. *Användare?* Frågan handlar om det behövs användare för att kunna genomföra en värdering. Med användare menas oftast personer tränade på/med det aktuella systemet (eller liknande system om det är ett helt nytt system som utreds).

2.2 Omvända cup-verktyget

”Omvända cup-verktyget” har utvecklats som ett stöd för att välja bland den mängd av MSI-värderingsmetoder som finns utvecklade.

Verktyget inklusive metodbeskrivningarna används i en trestegsprocess:

1. Handläggaren ställer sig "fyra frågor" (När?, Resurser?, Förmåga? och Användare?). Därigenom har antalet aktuella metoder begränsats avsevärt.
2. Handläggaren gör en grov bedömning av de metoder som ”omvända cup-verktyget” pekat ut som lämpliga genom att studera "stora matrisen".
3. Handläggaren skapar sig en än tydligare bild av utpekade metoders fördelar och nackdelar genom att läsa igenom metodbeskrivningarna.
4. (Tillval: Efter att ha identifierat en eller två lämpliga metoder kan dessa studeras vidare genom angivna referenser)

2.3 Metodernas ordning

Den ordning som metoderna beskrivs i ”Stora matrisen” motsvarar deras ordning i ”Omvända cup-verktyget”. Det vill säga metodernas ordning avspeglar inte någon traditionell vetenskaplig kategorisering, utan metodernas binära kategorisering för de fyra frågorna. Denna ordning

har valts utgående från antagandet att en handläggare som med ”Omvända cup-dokumentet” som utgångspunkt identifierat vissa metoder, sedan så lätt som möjligt skall hitta beskrivningarna för dessa metoder.

2.4 Stora matrisen

När handläggaren fått ett mindre antal metoder utpekade görs en första utvärdering av metoderna genom att studera "Stora matrisen". Under ett antal rubriker (som förklaras under avsnitt 2.5 Metodbeskrivningar), anges olika egenskaper för respektive metod. Handläggaren kan nu göra en första utvärdering av metoderna.

2.5 Metodbeskrivningar

För metoder som fortfarande är intressanta efter det att matrisen gått igenom finns mer information om metoderna, exempelvis deras användning och vilka typer av fel de identifierar att hämta ur metodbeskrivningarna.

I Metodbeskrivningar, vilket är rapportens huvuddel, ges kortfattade beskrivningar av de olika utvärderingsmetoderna och deras egenskaper.

Varje metodbeskrivning inleds med en kort redogörelse av vad det är för typ av metod och vad den används till. Detta följs av beskrivningar i tabellform som förklarar olika egenskaper hos metoden och krav för genomförande av utvärdering med den. Tabellen innehåller samma rubriker som ”Stora matrisen”. Metodens angivna värden i denna tabell motsvaras alltså av metodens kolumn i ”Stora matrisen” och beskriver mer ingående där angivna egenskaper. Här följer en kort beskrivning av vad respektive rubrik i tabellen står för:

När i utvecklingsprocessen. När i utvecklingsprocessen metoden kan/bör användas.

Resurser. Resurser som krävs för att utvärdering med metoden skall kunna genomföras. Exempelvis ekonomiska, personella (tid och antal människor) och/eller materiella resurser (svåranskaffad, kostnadskrävande utrustning).

Förmåga. Kompetens som krävs för att förbereda respektive genomföra utvärderingar med metoden.

Med/utan användare. Om användare deltar eller inte när utvärdering med metoden genomförs.

Teoretisk bakgrund. Metodens teoretiska bakgrund.

Mognadsgrad. Om metoden är ny respektive beprövad och har använts under lång tid.

Reliabilitet. Om metoden är stabil. Att metoden ger samma resultat vid upprepade mätningar under liknande förhållanden.

Validitet. Om metoden mäter det som den avser att mäta. Exempelvis att ett användbarhetstest är relevant för användbarhet av verkliga produkter för verkliga användare utanför testmiljön.

Sensitivitet/känslighet. Metodens förmåga att detektera skillnader.

Diagnosticitet. Metodens förmåga att härleda orsaker bakåt.

Typ av fel som hittas. Vilka typer av fel respektive användarproblem som identifieras med metoden.

Resultatanalys. Om analyserna är omfattande, eller kräver stor kompetens. Vilka typer av analyser som kan/bör göras av data insamlade med metoden, exempelvis kvantitativa eller kvalitativa analyser. Precisering av krav, för- och nackdelar som kan finnas för angivna analysmetoder.

Andra metoder som ger motsvarande information. Här anges andra metoder som ger ungefär samma typ av information. Det kan också vara preciseringar, exempelvis under vilka förhållanden andra metoder kan ge samma typ av information. Observera dock att två metoder inte ger exakt samma information, men att de övergripande kan ge information som pekar på samma typ av fel eller problem.

Kompletterande metoder. Andra metoder som kan användas för att samla in information som man inte kommer åt med metoden ifråga.

Tillämpbarhet. Krav som kan förekomma för att kunna genomföra utvärderingar med metoden. Förberedelser som krävs för att genomföra utvärdering med metoden.

Störande. Om metoden upplevs störande eller påträngande av användarna/försökspersonerna.

Användares acceptans. Om användare brukar ha invändningar mot att delta i utvärderingar med metoden.

Referenser. Litteraturförteckning avseende författare som refererats till i metodbeskrivningen, annan relevant litteratur och studier på FOI där utvärdering med metoden har genomförts. Motsvarande referenser finns även angivna i rapportens avslutande referenslista.

3 Omvända cup-verktyget

På följande sida finns det "omvända cup-verktyget". Här får läsaren en översikt över de behandlade metoderna.

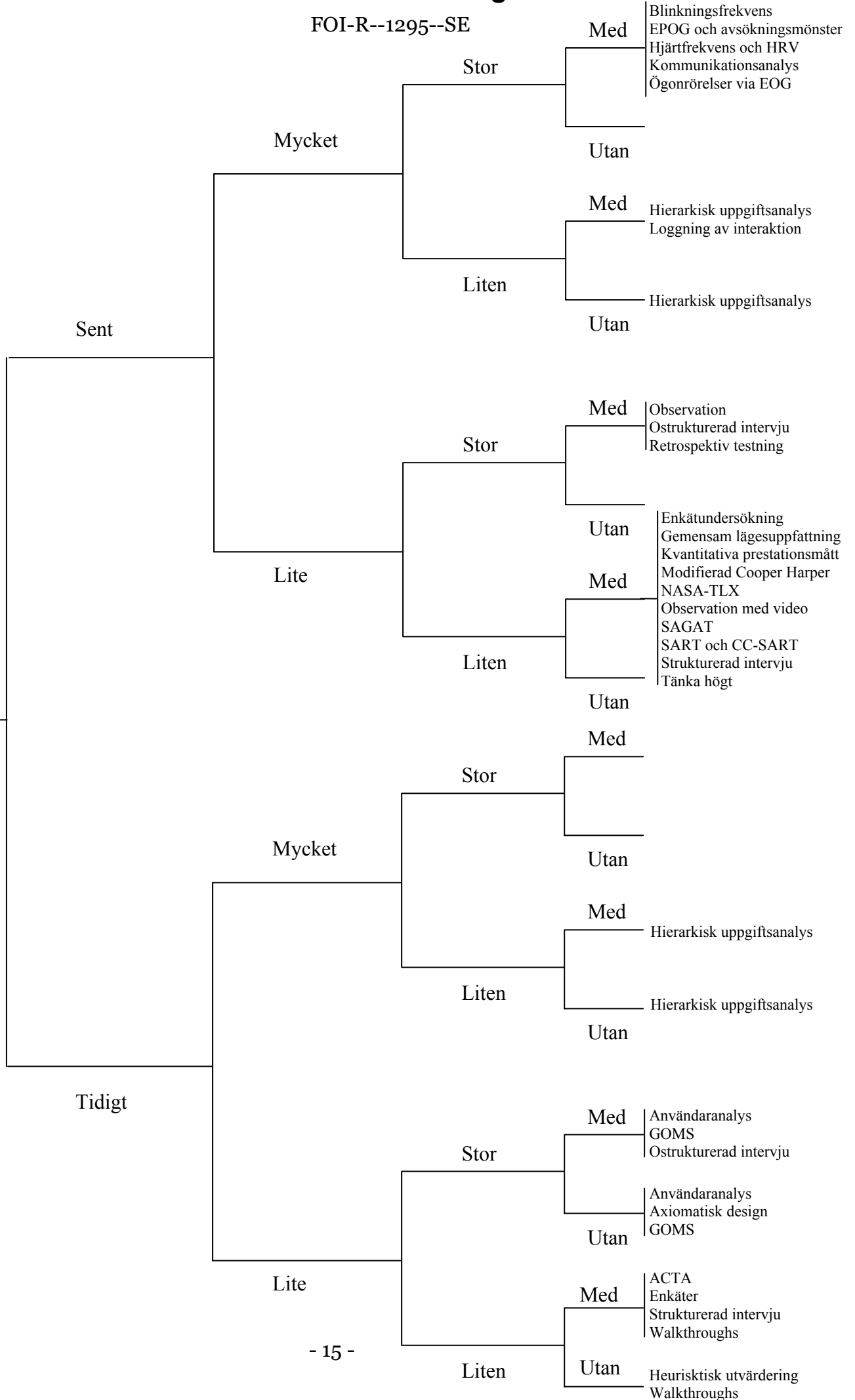
När

Resurser

Förmåga

Användare

FOI-R--1295--SE



4 Stora matrisen

Stora matrisen åskådliggör översiktligt de olika metodernas egenskaper.

METODER																										
Beskrivningar och krav	1. Blinkningsfrekvens	2. EPOG och avsökningmönster	3. Hjärtfrekvens och HRV	4. Kommunikationsanalys	5. Ögonrörelser via EOG	6. Hierarkisk uppgiftsanalys	7. Loggning av interaktion	8. Observation	9. Ostrukturerad intervju	10. Retrospektiv testning	11. Enkäter	12. Gemensam lägesuppfattning	13. Kvantitativa prestationsmått	14. Modifierad Cooper-Harper	15. NASA – TLX	16. Observation med video	17. SAGAT	18. SART, CC-SART	19. Strukturerad intervju	20. Tänka högt	21. Användaranalys	22. GOMS	23. Axiomatisk design	24. ACTA	25. Walkthroughs	26. Heuristisk utvärdering
När i utvecklingsprocessen	1	1	1	1	1	1/0	1	1	1/0	1	1/0	1	1	1	1	1	1	1	1/0	1	0	0	0	0	0	0
Resurser	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Förmåga	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
Med/utan användare	1	1	1	1	1	1/0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1/0	1/0	0	1	1/0	0
Teoretisk bakgrund	H	H	H	H	H	H	L	H	H	M	H	M	M	H	H	M	H	H	H	H	M	H	H	H	H	H
Mognadsgrad	H	H	H	H	H	H	M	H	H	M	H	L	M	H	H	H	M	M	H	H	M	H	M	M	M	M
Reliabilitet	M	M	M	L	M	M	H	L	M	M	M	M	H	H	H	M	M	M	M	M	L	M	M	M	M	M
Validitet	M	M	M	M	M	H	L	H	H	M	M	M	L	H	M	H	M	M	M	M	M	M	M	M	M	L
Sensitivitet/Känslighet	L	M	M	L	M	H	H	M	H	M	M	M	H	M	M	M	L	M	M	M	L	H	M	M	L	M
Diagnosticitet	L	M	L	M	L	H	L	H	H	M	M	L	L	L	M	H	L	H	M	H	L	M	H	M	L	L
Typ av fel som hittas																										
Resultatanalys	M	H	H	H	H	L	M	H	H	H	M	M	L	L	M	H	M	M	M	H	L	M	M	M	L	L
Andra metoder som ger motsvarande information																										
Kompletterande metoder																										
Tillämpbarhet	M	M	H	H	H	M	M	H	H	H	H	M	M	H	H	H	M	H	H	M	H	M	H	H	H	H
Störande	L	L	L	L	L	L	L	M	M	M	L	L	L	M	M	M	H	M	M	H	L	L	-	L	L	-
Användares acceptans	H	M	M	H	H	H	H	M	H	M	M	H	H	H	M	M	M	M	H	M	H	H	-	H	H	-

0 = Tidigt, Lite resurser, Liten förmåga, Utan användare
 1 = Sent, Mycket resurser, Stor förmåga, med användare
 H = Högt, M = Medium, L = Låg

5 Metodbeskrivningar

5.1 *Blinkningsfrekvens*

Blinkningsfrekvens kan användas för värdering av visuell arbetsbelastning, mental arbetsbelastning och trötthet/utmattning. Komponenter i den endogena blinkningen (dvs. en ögonblinkning som inte uppstår som följd av ett externt stimuli) har studerats med avseende på mental arbetsbelastning (Kramer, 1991; Stein, 1992). De komponenter som bedömts relevanta har framför allt varit blinkningsfrekvens, fördröjning av blinkning efter stimuli ("latency") och varaktighet. Generellt sett har färre blinkningar och blinkningar med kortare varaktighet kopplats till ökad belastning vid uppgifter som bilkörning i stad, läsning och vapenleverans i flygvapnet (Brookings & Wilson, 1994; Krebs, Wingert & Cunningham, 1977).

Ett problem är att en avsökning, eller en mental uppgift, ibland avslutas med en blinkning.

Blinkningsfrekvens	
Krav	Beskrivning
När i utvecklingsprocessen	Vanligtvis i senare faser i designprocessen, eftersom det är ett analytiskt verktyg som används för t.ex. trötthets- och uppmärksamhetsstudier, användbarhetstestning och utvärdering.
Resurser	Kan samlas in m.h.a. utrustning för EOG eller EPOG.
Förmåga	Bör vara hög. Bearbetning och korrigering av mätfel kräver expertis.
Med/utan användare	Med användare.
Teoretisk bakgrund	Stern (1993) fann att man via blinkningsfrekvens kunde upptäcka trötthet hos militära piloter. Blinkningsfördröjningstid ("latency") och blinkduration verkar vara lovande som mått för mätning av arbetsbelastning (Kramer, 1991). Fördröjningstiden har funnits öka med ökade krav på minnet och utförande av handlingar. Kramer (1991) anser att både fördröjningstid och blinkduration ökar med stigande visuella krav i uppgiften. Stein har arbetat med okulära mätningar vid flygtrafikledning, där han bl.a. funnit att sackduration minskar med ökad uppgiftsbelastning (Stein, 1992).
Mognadsgrad	Metoden bedöms vara mogen.
Reliabilitet	Individskillnader mellan försökspersonerna är vanligt förekommande. Frekvensen för ögonblinkningar kan öka vid snabb avsökning av exempelvis en instrumentpanel, beroende på att en sekvens av informationsinhämtning ofta avslutas med en ögonblinkning (Tsang & Wilson, 1997).
Validitet	Kopplingen mellan mental arbetsbelastning och blinkningar anses av flera forskare vara något svag (Krebs, Wingert & Cunningham, 1977; Casali & Wierwille, 1983) trots att empiriska bevis finns för att blinkningsfrekvensen minskar vid förhöjd uppgiftsbelastning. Troligen är blinkningsfrekvens i större utsträckning användbart vid mätning av trötthet och utmattning.
Sensitivitet/Känslighet	Måttet är endast känsligt för uppgifter som varar längre än ungefär tre minuter.
Diagnosticitet	Användbart för att kvantifiera visuell arbetsbelastning och trötthet/utmattning. Analys av blinkningsfrekvens kan inte finna/särskilja tecken på mindre förändringar som inte bidrar till förändrade krav på uppgiftsprocessande (t.ex. förändringar i färgkodning eller strategiskiftet).
Typ av fel som hittas	Arbetsuppgifter där försökspersonen blir visuellt belastad, mentalt belastad eller trött/utmattad.
Resultatanalys	Görs främst med dator och specialprogramvara. Relativt krävande att göra manuellt via videoinspelningar.
Andra metoder som ger motsvarande information	För uppskattning av visuell belastning kan även ögonrörelser via EOG och avsökningsmönster via EPOG användas. För uppskattning av mental arbetsbelastning kan även Hjärtfrekvens (HR), Variabilitet i hjärtfrekvens (HRV) och EPOG användas. Metoder för mätning/skattning av trötthet/utmattning, vilket är mått som blinkningsfrekvens är starkt relaterat till. Intervjuer/enkäter för att få ytterligare information hur arbetsuppgiften/systemet upplevts.
Kompletterande metoder	Subjektiva skattningar av mental arbetsbelastning (t.ex. Modifierad Cooper-Harper, NASA-TLX) och prestationsmått.
Tillämpbarhet/Implementeringskrav	Efterhandsanalys och korrigering av mätfel kräver expertkunskap.

Störande/Påträngande	Beroende på utrustning som används vid datainsamling. Att använda EOG och portabel bandspelare för att mäta blinkningsfrekvens är mindre påträngande än att använda huvudmonterad blickregistreringsutrustning (Castor m.fl., 2003). Via användning av enbart videoinspelning får man en minimal störning, men en krävande och trubbig analys.
Användares acceptans	Beroende på i hur hög grad utrustningen upplevs som påträngande av operatören.
Referenser	<ul style="list-style-type: none"> • Castor, M., Hanson, E., Svensson, E., Nählinder, S., LeBlaye, P., MacLeod, I., Wright, N., Alfredson, J., Ågren, L., Berggren, P., Juppet, V., Hilburn, B., & Ohlsson, K. (2003). <i>GARTEUR Handbook of mental workload measurement</i> (GARTEUR FM AG13 Final Report - GARTEUR TP 145). • Brookings, J. B. & Wilson, G. F. (1994). Physiological and workload changes during a simulated air traffic control task. In <i>Proceedings of the Human Factors and Ergonomics Society, 38th Annual Meeting</i>. • Casali, J. & Wierwille, W. (1983). A comparison of rating scale, secondary task, physiological, and primary task workload estimation techniques in a simulated flight task emphasizing communications load. <i>Human Factors, 25</i>, 623-642. • Kramer, A. F. (1991). Physiological metrics of mental workload: a review of recent progress. I D. L. Damos (Ed.), <i>Multiple Task Performance</i>. London: Taylor & Francis. • Krebs, M. J., Wingert, J. W., & Cunningham, T. (1977). <i>Exploration of an Oculometer-Based Model of Pilot Workload</i>. NASA technical report CR-145153. Minneapolis, Minnesota: Honeywell Systems & Research Center. • Stein, E. S. (1992). <i>Air Traffic Control Visual Scanning</i>. FAA Technical Report DOT/FAA/CT-TN92/16. US Department of Transportation. • Stern, J. A. (1993). The Eyes: Reflector of attentional processes (synopsis by J. J. Kelly). In <i>CSERIAC Gateway</i>, 4 (4), 7-12. • Tsang, P., & Wilson, G. F. (1997). Mental workload. In G. Salvendy (Ed.), <i>Handbook of human factors and ergonomics</i> (2nd ed., pp. 417-449). New York: Wiley.

5.2 EPOG och avsökningsmönster

Eye-Point-of-Gaze (EPOG) innebär mer än enbart mätning av ögonens rörelser. Det inkluderar även den punkt (på vilken yta) man riktar blicken mot. Att se var och hur någon tittar kan relateras till mental arbetsbelastning. Vid sidan av EPOG och avsökningsmönster kan man även använda mått på andra ögonrelaterade parametrar, av primär och sekundär sort, för att samla in indikationer på mental arbetsbelastning.

Stein (1992) fann att flygledare tenderade att fixera blicken mindre frekvent då uppgiftsbelastningen ökade. Erfarna trafikledare tenderade dock att söka av sin skärm oftare. Det sistnämnda resultatet påvisar möjligen en existerande skillnad i avsökningsstrategi mellan erfarna och oerfarna flygledare.

EPOG och avsökningsmönster	
Krav	Beskrivning
När i utvecklingsprocessen	P.g.a. sin kontextkänslighet används studier av EPOG och avsökningsmönster lämpligen senare i utvecklingsprocessen.
Resurser	Relativt krävande metod med dyr specialutrustning för insamling av ögonrörelser, eventuellt i kombination med huvudrörelser.
Förmåga	Bör vara hög.
Med/utan användare	Med användare.
Teoretisk bakgrund	<p>De två främsta ögonrörelseparametrarna är sackader och fixeringar. Det är allmänt accepterat att sackader görs för att förflytta uppmärksamheten till viktig information, och fixeringar är nödvändiga för att bearbeta informationen. Man är relativt överens om att det inte sker någon perceptuell informationsbearbetning under sackader. Nutida forskning har dock visat på att viss bearbetning på låg kognitiv nivå faktiskt sker. Högre fixeringsfrekvens kan indikera hög arbetsbelastning (Svensson, Angelborg-Thanderz & Wilson, 1999). Kortare sackader kan indikera en ökad mental arbetsbelastning (May m.fl., 1990). Svensson m.fl. (1997a) fann att antalet korta fixeringstider head-up och antalet längre fixeringstider head-down ökade med ökad informationsbelastning på den taktiska indikatorn i ett militärt flygplan.</p> <p>Sekundära ögonrörelseparametrar är de parametrar som fås från primära parametrar genom filtrering, integrering, beräkning av medelvärden, eller via andra matematiska beräkningar. De viktigaste sekundära ögonrörelseparametrarna är entropi (rutiner i avsökningsmönstret och avvikelser från detta), fixeringstid (kvardröjning av blicken vid en punkt), avsökningsväg, perceptuellt och visuellt intervall (avsökningsområde) och kopplingen mellan blinkningar och sackader (blinkning i samband med vissa förflyttningar av blicken) (Castor m.fl., 2003).</p>
Mognadsgrad	Metoderna är relativt mogna och används både vid laboratorieförsök och i tillämpade miljöer.
Reliabilitet	Variation mellan individer i visuellt beteende finns. Även problem med kalibrering och återkalibrering av utrustningen kan orsaka reliabilitetsproblem.
Validitet	EPOG och avsökningsmönster är båda högst uppgifts- och individberoende. Pupilldiameter är känslig för t.ex. belysningsförhållanden.
Sensitivitet/Känslighet	Känsligheten kommer upp på en acceptabel nivå om en kombination av primära och sekundära ögonbaserade mått används för att uppskatta mental arbetsbelastning.
Diagnosticitet	Metoderna ger indikation på mental arbetsbelastning, påverkad av såväl kognitivt som visuellt krävande uppgifter.
Typ av fel som hittas	Kan identifiera arbetsuppgifter där försökspersonen har hög mental arbetsbelastning, eller hög visuell belastning.
Resultatanalys	Analysen kan göras i nära realtid, såväl som på inspelade data i efterhand.
Andra metoder som ger motsvarande information	EOG.
Kompletterande metoder	<p>Subjektiva skattningar av mental arbetsbelastning (t.ex. Modifierad Cooper-Harper, NASA-TLX) och prestationsmått.</p> <p>Intervjuer/enkäter för att få ytterligare information hur arbetsuppgiften/systemet upplevts.</p>
Tillämpbarhet/Implementeringskrav	EPOG kan användas för dynamisk mätning av mental arbetsbelastning, även i tillämpningar nära realtid.

Störande/Påträngande	Huvudmonterade system kan vara direkt eller indirekt störande. Nya EPOG-system med lägre störningsgrad är under utveckling.
Användares acceptans	Generellt hög acceptans, men vissa system stör operatören med avseende på rörelseförmåga och tyngd.
Referenser	<ul style="list-style-type: none"> • Castor, M., Hanson, E., Svensson, E., Nählinder, S., LeBlaye, P., MacLeod, I., Wright, N., Alfredson, J., Ågren, L., Berggren, P., Juppet, V., Hilburn, B., & Ohlsson, K. (2003). <i>GARTEUR Handbook of mental workload measurement</i> (GARTEUR FM AG13 Final Report - GARTEUR TP 145). • May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., & Brannan, J. R., (1990). Eye Movements indices of Mental Workload, <i>Acta Psychologica</i>, 75, 75-89. • Stein, E. S. (1992). <i>Air Traffic Control Visual Scanning</i>. FAA Technical Report DOT/FAA/CT-TN92/16. US Department of Transportation. • Svensson, E., Angelborg-Thanderz, M., Sjöberg, L., & Olsson, S. (1997a). Information complexity – mental workload and performance in combat aircraft. <i>Ergonomics</i>, 40, 362-380. • Svensson, E., Angelborg-Thanderz, M., & Wilson, G. F., (1999). <i>Models of pilot performance for systems and mission evaluation, psychological and psychophysical aspects</i>. AFRL-HE-WP-Tr-1999-0215.

5.3 Hjärtfrekvens och HRV

Mätning av hjärtfrekvens (Heart Rate, HR) och hjärtfrekvensvariabilitet (Heart Rate Variability, HRV) har visat sig fungera som relativa indikationer på mental arbetsbelastning.

Att utföra en mental arbetsuppgift kräver uppmärksamhet, informationsprocessande och mental ansträngning. Det har visat sig dels att hjärtfrekvensen stiger, dels att HRV sjunker vid ökad mental arbetsbelastning.

Både hjärtfrekvens och HRV tas ut med hjälp av kardiovaskulära tidsseriemätningar, framför allt EKG (elektrokardiogram), men även genom exempelvis dynamisk blodtrycksmätning eller pulsoxymetri.

Hjärtfrekvens och HRV	
Krav	Beskrivning
När i utvecklingsprocessen	Vanligtvis sent i utvecklingsprocessen, för att analysera effekter av systemets inverkan på operatörens mentala arbetsbelastning.
Resurser	Idag finns mindre, bärbar utrustning för EKG-registrering, men kostnaden är fortfarande relativt hög. Att analysera data är tidsödande även om automatiserad programvara finns.
Förmåga	Ställer krav vad gäller såväl insamling som tolkning av data.
Med/utan användare	Med användare.
Teoretisk bakgrund	<p>Hjärtfrekvens (Heart Rate, HR) och hjärtfrekvensvariabilitet (Heart Rate Variability, HRV) är två besläktade metoder som mäter psyko-fysiologiska fenomen som visat sig fungera som relativa indikationer på operatörsstatus, närmare bestämt mental arbetsbelastning (GARTERUR, 2002).</p> <p>Hjärtfrekvens kan variera beroende på mental belastning, mellan ca 50-150 slag per minut (BPM). Hjärtfrekvensen kan dock även ökas av fysisk aktivitet och andningsfrekvens.</p> <p>HRV är ett mått på hur mycket tiden mellan varje hjärtslag varierar, alltså hur ”stadig” hjärtfrekvensen är. Genom statistisk analys fås därmed ett mått på hur mycket hjärtfrekvensen varierar på mikronivå. Observera att man kan ha hög eller låg HRV både när hjärtfrekvensen är hög och låg. Med tillgång till HRV-värden kan man även räkna ut hjärtfrekvens (HR).</p> <p>Måtten är relativa, en baseline-mätning krävs initialt för att kunna uttala sig om ökad mental arbetsbelastning föreligger vid senare tillfällen. Jämförelse sker alltså på ordinalskala och endast med svårighet i jämförelse med andra operatörer.</p> <p>För vidare läsning om Hjärtfrekvens se t.ex. Magnusson (2002), Jorna (1993), Roscoe (1993), Hankins och Wilson (1998)</p>
Mognadsgrad	Relativt beprövad metod som använts vid många studier som mått på mental arbetsbelastning.
Reliabilitet	Båda måtten mäter endast relativa indikationer på mental arbetsbelastning. Reliabiliteten kan därför sägas vara medelhög. Andningsfrekvens måste kompenseras för, samt fysisk ansträngning. Risk för störning av elektrisk apparatur.
Validitet	Interna och externa faktorer och initialt tillstånd hos användaren påverkar resultaten, som kan räknas som relativt valida om dessa faktorer konstanthålls. Hur pass väl mätresultaten återger mental arbetsbelastning är en omdiskuterad fråga.
Sensitivitet/Känslighet	Medelhög. Metoden medger precision ner till en millisekunds nivå, men för att få pålitliga mått bör man ta in data under tidsperioder på minst 30 sekunder till fem minuter, så att initial puls vid applicering av utrustningen kan sjunka till normal nivå.
Diagnosticitet	Låg, eftersom metoden inte förklarar bakomliggande orsaker till förändrad mental arbetsbelastning.
Typ av fel som hittas	Fel som ställer ökade krav på operatörens mentala processande, uppmärksamhet, problemlösning etc.
Resultatanalys	Krävande arbete med stora datamängder.
Andra metoder som ger motsvarande information	Med andra psykofysiologiska metoder kan liknande information fås.
Kompletterande metoder	Subjektiva skattningar av mental arbetsbelastning, t.ex. Modifierad Cooper-Harper, NASA-TLX och prestationsmått.

Tillämpbarhet/Implementeringskrav	Hög, metoden går att använda i de flesta miljöer. Tillämpbarheten är dock begränsad om försökspersonen rör sig mycket under arbetet, eller blir fysiskt belastad.
Störande/Påträngande	Låg störning, elektroderna orsakar vanligen inga besvär för användarna.
Användares acceptans	Relativt hög, dock initialt motstånd mot fysiologisk mätning ibland.
Referenser	<ul style="list-style-type: none"> • Castor, M., Hanson, E., Svensson, E., Nählinder, S., LeBlaye, P., MacLeod, I., Wright, N., Alfredson, J., Ågren, L., Berggren, P., Juppet, V., Hilburn, B., & Ohlsson, K. (2003). <i>GARTEUR Handbook of mental workload measurement</i> (GARTEUR FM AG13 Final Report - GARTEUR TP 145). • Magnusson, S. (2002). Similarities and Differences in Psychophysiological Reactions Between Simulated and Real Air-to-Ground Missions. <i>The International Journal of Aviation Psychology</i>. Vol. 12(1), 49-61. • Jorna, P. G. A. M. (1993), Heart rate and workload variations in actual and simulated flight, <i>Ergonomics</i>, vol. 36, 1043-1054. • Roscoe, A. H. (1993), Heart rate as a psychophysiological measure for in-flight workload assessment. <i>Ergonomics</i>, vol. 36, 1055-1062. • Hankins, T. C., & Wilson, G. F. (1998), A Comparison of Heart Rate, Eye Activity, EEG and Subjective Measures of Pilot Mental Workload During Flight. <i>Aviation Space and Environmental Medicine</i>, vol. 69, 360-367.

5.4 Kommunikationsanalys

Operatörer som samarbetar, inom team, mellan enheter eller team, behöver koordinera sina handlingar. Detta görs till en stor del genom muntlig, gestbaserad samt skriftlig kommunikation. I det nätverksbaserade försvaret är behovet av att analysera kommunikation mellan enheter uppenbart. Analys av operatörer som samarbetar, dels inom team, dels mellan olika enheter eller team, kan innefatta någon form av kommunikationsanalys. Bland de frågor som kan besvaras med kommunikationsanalys är:

- Vem talar med vem?
- Vad är innehållet i kommunikationen (Innehållsanalys, ”content analysis”)?
- Vilka kommunikationskanaler/medier används, och i vilken utsträckning?
- Hur mycket kommuniceras det?
- Uppstår problem i kommunikationen?
- Får kommunikationsproblem allvarliga konsekvenser för prestationen?

Kommunikationsanalys	
Krav	Beskrivning
När i utvecklingsprocessen	Kommunikationsanalys kan användas såväl före, som under och även efter utvecklingen av ett nytt system. De respektive syftena med kommunikationsanalysen är då olika. Datainsamlingen kräver dock någon form av simulerat eller verkligt system.
Resurser	Relativt lite resurser för själva inspelningen, men ett tidskrävande arbete att analysera kommunikationen i efterhand. Analysen underlättas stort av domänförståelse, varvid experter bör finnas att rådfråga. Innehållsanalys kan underlättas genom att använda i förväg utformade protokoll.
Förmåga	En del förkunskaper samt erfarenhet i transkription krävs för att göra analysen effektivt. Bästa resultat fås genom att använda flera oberoende analyser av olika personer, i samråd med experter.
Med/utan användare	Inga slutanvändare behövs under analysarbetet. Deras medverkan och medgivande till inspelning och genomförande av kommunikationsanalysen behövs dock.
Teoretisk bakgrund	Med grund i lingvistik och informationsteorier har idag kommunikationsanalyslitteraturen en egen plats (t.ex. Clark, 1996); Linell, (1994). Metoder för att analysera kommunikation på olika nivåer innefattar ofta en skriftlig transkription av talad kommunikation. Om i huvudsak elektronisk skriftlig kontakt sker mellan enheterna/operatörerna så kan denna kommunikation i regel med enkelhet loggas. En nackdel med kommunikationsanalys är att endast öppen kommunikation kan analyseras. I exempelvis erfarna team kan koordination ske genom dold kommunikation, t.ex. ögonkontakt, gester eller manövrar, som inte med lätthet kan fångas i en ljudinspelning. Koordination kan även ske implicit, genom förutseende av andras behov, eller genom planering och kända regler för genomförandet. Kommunikationsanalys kan ge ledtrådar om operatörernas enskilda situationsmedvetenhet i specifika ögonblick.
Mognadsgrad	Relativt beprövad metod, men tillämpningen mot värdering av operatörers mentala arbetsbelastning och situationsmedvetande är förhållandevis ny.
Reliabilitet	Låg, då ett stort mått av tolkning krävs. Flera oberoende analyser skall helst genomföras, eftersom innehållskodning, transkription och tolkning till sin natur är ett relativt subjektivt arbete.
Validitet	Medelhög, det är till stor del upp till analytikernas domänförståelse att en korrekt tolkning av kommunikationen kan göras.
Sensitivitet/Känslighet	Relativt låg, men beroende på vilken nivå man gör analysen, det vill säga hur noggrant analytikerna väljer att bryta ned den.
Diagnosticitet	Medelhög. Uppenbara problem i systemet påverkar operatörerna i sådan mån att det märks i deras samarbete och därmed i deras kommunikation. Det krävs däremot domänkunskap för att kunna härleda problemen till kommunikationsbrister och inte rena problem med systemen i sig.

Typ av fel som hittas	<p>Problem i kommunikationen, bristande länkar och kontakter mellan delar i systemet och mellan operatörer.</p> <p>Att analysera ett team av operatörer som använder ett befintligt system kan tjäna som komplement till uppgiftsanalys. Om man sedan upprepar kommunikationsanalysen då ett nytt system införts, kan skillnaderna i kommunikationsmönster tala om tendenser, till fördel eller nackdel för det nya systemet.</p> <p>Kan ge ledtrådar om brister i operatörers enskilda, och gemensamma, situationsmedvetenhet i specifika ögonblick.</p> <p>Kan påvisa mental arbetsbelastning hos operatörer.</p>
Resultatanalys	<p>Tidsödande, dock beroende på vilken nivå man analyserar.</p> <p>Beroende på när i utvecklingsfasen analysen sätts in, kommer resultaten att vara olika.</p> <p>De mått som används innefattar både kvantitativa och kvalitativa mått på mängd, problemhalt, efterföljning av befintliga mallar, samt kommunikationens innehåll.</p>
Andra metoder som ger motsvarande information	Ev. observation, loggning av interaktion samt intervju.
Kompletterande metoder	Uppgiftsanalys.
Tillämpbarhet/Implementeringskrav	Hög. Metoden kan användas när fler än en operatör finns, och någon form av språkligt samarbete förekommer.
Störande/Påträngande	Inte speciellt störande. En enkel analys kan ske on-line med protokoll, men vanligen krävs ljudinspelning och efterhandsanalys för att få en större förståelse för kommunikationsmönster och problem.
Användares acceptans	Torde vara stor, då inspelningsutrustning och mikrofoner kan placeras så de inte stör.
Referenser	<ul style="list-style-type: none"> • Linell, P. (1994) <i>Transkription av tal och samtal: Teori och praktik</i>. Linköping: Institutionen för TEMA Kommunikation, Linköpings Universitet. • Clark, H. H. (1996). <i>Using Language</i>. Cambridge: Cambridge University Press.

5.5 Ögonrörelser via EOG

Analys av ögonrörelser kan göras genom utrustning för elektro-okulografi (EOG), videoinspelning och huvudmonterad utrustning som använder optisk teknik. Ögonrörelser och avsökingsbeteenden används ofta som mått på visuell uppmärksamhet i en uppgift. EOG-signalen är relativt enkel att registrera, spela in och använda. För vissa arbetsuppgifter kan ögonrörelseanalys även användas för bedömning av mental arbetsbelastning.

Ögonrörelser via EOG	
Krav	
När i utvecklingsprocessen	Kan användas i tidiga stadier av utvecklingen, men dess främsta tillämpningen är vanligtvis under senare faser, vid testning och utvärdering.
Resurser	Kräver EOG-utrustning, som kan vara portabel, samt dator och programvara för analys av ögonparametrar.
Förmåga	Datainsamlingen är relativt enkel. Analysen kräver dock datorresurser, tid och områdeskunskap.
Med/utan användare	Med användare.
Teoretisk bakgrund	<p>Det finns en del skrivet om tillämpningar av metoden, t.ex. inom flygforskning (t.ex. Castor m.fl., 2003).</p> <p>EOG är ett mått på aktiviteten hos de muskler som kontrollerar ögonglobens rörelser, vilket i sin tur genererar ögonens aktivitet och blickriktning. Då de visuella kraven i en uppgift ökar förändras EOG-måttet därefter.</p> <p>Användbarheten av ögonrörelseanalys för att förutsäga mental arbetsbelastning beror på uppgiftens natur, i synnerhet vad gäller visuella krav. Ögonrörelserna återspeglar inhämtning av information, snarare än central bearbetning av den. Följande mått har visat sig korrelera med arbetsbelastning:</p> <ul style="list-style-type: none"> • <i>Avsökningsbeteende</i>: sackadhastighet (stiger med ökad arbetsbelastning), sackadintensitet och kvardröjningstid (minskar med ökad arbetsbelastning). • <i>Blinkningsaktivitet</i>: frekvens och duration (minskar med ökad arbetsbelastning), och undertryckande av blinkning. <p>Metoden används även ofta som ett mått på visuell uppmärksamhet i en arbetsuppgift.</p>
Mognadsgrad	EOG är en standardmetod för mätning av ögonrörelser.
Reliabilitet	EOG-signalen är relativt tydlig och lätt att mäta. Vad gäller sackadiska ögonrörelser så är individuella skillnader ett mindre problem, jämfört med andra psykofysiologiska metoder. Dock finns stora individuella skillnader i måtten på blinkbeteende, vilket ställer krav på mätning av "baseline" hos försökspersonerna.
Validitet	Ögonrörelser påverkas av trötthet, stress och ljusnivå, vilket gör det nödvändigt att utveckla och använda sig av alternativa sätt att bedöma dessa påverkansfaktorer (t.ex. mätning av hjärnaktivitet eller subjektiva skattningar).
Sensitivitet/Känslighet	Sackadiska ögonrörelser påverkas av visuella uppgiftskrav på en relativ skala, dvs. man kan i resultaten särskilja mellan låg, medelhög och hög kravnivå. Blinkningsparametrar är mer känsliga för höga uppgiftskrav. EOG-signalen är relativt känslig och värdena förändras över några sekunder. Ett stabilare mått fås dock genom beräkning av medelvärdet över ett intervall på cirka 30 sekunder.
Diagnosticitet	EOG kan användas för att registrera blickriktning, blinkningar och sackadiska rörelser hos ögat. Måttet kan dock inte tas som ett generellt mått på mental arbetsbelastning, utan bör användas i kombination med andra belastningsmått, såsom hjärtfrekvens, hjärtfrekvensvariabilitet och hjärnaktivitet.
Typ av fel som hittas	Kan påvisa visuell belastning och mental arbetsbelastning.
Resultatanalys	Analysen av EOG-data görs med dator, och kräver särskild programvara med förmåga att urskilja beteenden, avseende ögonrörelser.

Andra metoder som ger motsvarande information	EPOG
Kompletterande metoder	NASA-TLX, Modifierad Cooper-Harper, HR, HRV, Hjärnaktivitet, Intervjuer
Tillämpbarhet/Implementeringskrav	EOG kan användas för att jämföra olika designförslag, avseende gränssnitt, konstruktion etc. Eftersom ögonrörelser påverkas av många andra variabler är det troligen säkrast att använda metoden vid jämförelser mellan relativt små skillnader i design, hellre än att jämföra helt olika miljöer. EOG är tillämpligt i såväl laboratoriemiljö som i operationella situationer. Utrustningen kan vara relativt liten (t.ex. i storleksordningen bärbar kassetbandspelare) men kan ändå spela in data under ca ett dygn. Registrering av ögonrörelser kan exempelvis ske genom att fyra elektroder fästs runt ögonen, ovanför och under ett öga och vid båda tinningarna, ca 1 cm vid sidan om vardera ögat.
Störande/Påträngande	Elektroden runt ögonen anses inte vara särskilt påträngande när de väl har fästs på plats. Dock kan kablagen komma att vara störande om de inte fästs upp ordentligt ur vägen.
Användares acceptans	EOG-mätning är väl accepterat bland flygbesättningar, och utrustningen tar bara ca 2-3 minuter att fästa på plats.
Referenser	<ul style="list-style-type: none"> • Castor, M., Hanson, E., Svensson, E., Nählinder, S., LeBlaye, P., MacLeod, I., Wright, N., Alfredson, J., Ågren, L., Berggren, P., Juppet, V., Hilburn, B., & Ohlsson, K. (2003). <i>GARTEUR Handbook of mental workload measurement</i> (GARTEUR FM AG13 Final Report - GARTEUR TP 145).

5.6 Hierarkisk uppgiftsanalys inklusive "decomposition methods"

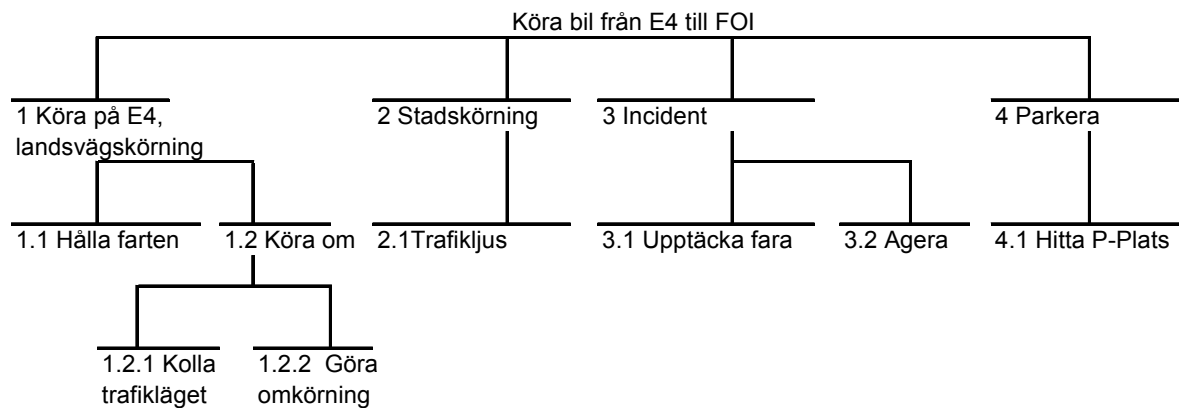
Hierarkisk uppgiftsanalys (HTA) tillsammans med "decomposition methods" måste anses vara de mest intuitiva och grundläggande uppgiftsanalysmetoderna. I en HTA beskrivs en operatörs handlande i termer av planer/mål och vilka operationer som behövs för att uppfylla planen. Dessa planer och operationer beskrivs i en hierarki där beskrivningen bryts ner till lämplig nivå.

"Decomposition methods" är samlingsnamnet på en typ av metoder som används för att ytterligare bryta ner eller beskriva en tidigare gjort hierarkisk uppgiftsanalys. Den typ av vidare nedbrytning av HTA som kallas "decomposition methods" beskrevs först av Miller (1953). Metoderna har också benämnts "tabular task analysis" då de vanligen resulterar i stora tabeller med smala kolumner.

Hierarkisk uppgiftsanalys	
Krav	Beskrivning
När i utvecklingsprocessen	Kan användas både tidigt och sent i designprocessen. Tidigt för att vara preskriptivt och sent för att vara deskriptivt.
Resurser	Analysen kan vara ganska tidskrävande. Enligt Kirwan och Ainsworth (1992) anses dock metoden som relativt billig eftersom det endast är nödvändigt att fokusera på väsentliga delar av uppgiften. Inga särskilda materiella resurser krävs,
Förmåga	För att genomföra analysen krävs viss kompetens och erfarenhet. Genom träning kan dock nödvändig kunskap förvärvas relativt snabbt (Kirwan & Ainsworth, 1992).
Med/utan användare	Expertstöd behövs om analytikern inte är insatt i domänen.
Teoretisk bakgrund	Utvecklades ursprungligen för träning, men har vidareutvecklats och tillämpats för ett stort antal andra sammanhang (Kirwan & Ainsworth, 1992). Det är följaktligen en generell metod för att analysera ett stort antal olika typer av uppgifter, inklusive uppgifter med kognitiva inslag.
Mognadsgrad	Mogen och mycket använd uppgiftsanalysmetod.
Reliabilitet	Medelhög, dock beroende på analytikerns kompetens.
Validitet	Hög.
Sensitivitet/Känslighet	Hög.
Diagnosticitet	En detaljerad uppgiftsanalys som bryts ner långt har hög diagnosticitet, framför allt om relevanta kategorier i ”decomposition methods” läggs till i analysen.
Typ av fel som hittas	Metoden lämpar sig att använda när man vill analysera en operatörs informationsbehov och kontrollmöjligheter. Man kan på förhand identifiera situationer där operatören tvingas titta på flera displayer samtidigt, eller där han/hon saknar kontrollmöjligheter.
Resultatanalys	Kräver viss kompetens.
Andra metoder som ger motsvarande information	I vissa fall kognitiv uppgiftsanalys.
Kompletterande metoder	Incidentanalys, Observation, Tänka högt, enkäter, intervjuer.
Tillämpbarhet/Implementeringskrav	Metoden lämpar sig bäst för att beskriva procedurella färdigheter med låg dynamik, dvs. för att beskriva i vilken ordning en operatör gör något som oftast ser likadant ut. Metoden blir snabbt osmidig när det finns mycket dynamik, och många iterationer görs.
Störande/Påträngande	Nej.
Användares acceptans	Hög, då resultaten är lättförståeliga
Referenser	<ul style="list-style-type: none"> • Kirwan, B. & Ainsworth, L., K. (1992). <i>A guide to task analysis</i>. London, (pp.95-104). Taylor & Francis. • Miller, R. B. (1953) <i>A Method for Man-machine Task Analysis</i>. Wright-Patterson Air Force Base, Ohio, USA: WADC Tech Rept. No 53, 137.

Ytterligare beskrivning av Hierarkisk uppgiftsanalys och decomposition methods

Ett enkelt exempel på hur resultatet av en HTA kan se ut återfinns i Figur 4. Här beskrivs de handlingar och mål man kan använda om man beskriver en person som tar av från E4 för att ta sig fram till FOI. I Tabell 1 ges exempel på tabeller som ”decomposition methods” resulterar i, när två av elementen från den Hierarkiska uppgiftsanalysen i Figur 4 brutits ner och beskrivs enligt åtta kategorier. I Tabell 2 ges en sammanställning av de kategorier för ”decomposition” som använts i tidigare forskning.



Figur 4. Ett exempel på utseendet hos en hierarkisk uppgiftsanalys (HTA). Här delvis nedbrutet i tre nivåer. Behovet avgör hur långt man bryter ner uppgifterna.

Tabell 1. Ett exempel på de tabeller som ”decomposition methods” resulterar i. Här har två av elementen från Figur 4 brutits ner och beskrivs enligt åtta kategorier.

Task Element	Purpose	Cue	Decision	Action
1.2 Köra om	Komma förbi bilen framför	Det går för långsamt Bil framför	Köra om eller ligga kvar	Blinka, byta fil gasa
Incident	Undvika att krocka	Ser något med perifera seendet	Bromsa eller svänga undan	Tuta, bromsa, gasa
Task Element	Displays	Controls	Feedback	Likely Errors
1.2.2 Göra om-körning	Hastighetsmätaren	Ratten, Blinkers	Ser bilen i backspegeln	Hinner inte om Kör in för tidigt
Incident		Ratten, bromsen	Faran är förbi	Hinner inte besluta i tid Upptäcker inte i tid

Tabell 2. Analysen styrs av vilka kategorier man väljer att använda i tabellen. En sammanställning av kategorier som använts i tidigare forskning återfinns i Kirwan och Ainsworth (1992).

Description of task

Description
Type of activity/behaviour
Task/Action verb
Function/Purpose
Sequence of activity

Requirements

Initiating cue/event
Information
Skills/training required
Personnel requirements

Hardware features

Location
Controls used
Displays used
Critical values
Job aids required

Performance on the task

Performance
Time taken
Required speed
Required accuracy
Criterion of response adequacy

Other activities

Subtasks
Communications
Co-ordination requirements

Nature of the task

Actions required
Decisions required
Responses required
Task complexity
Task difficulty
Task criticality

Outputs from the task

Output
Feedback

Consequences/problems

Likely/typical errors
Errors made/problems
Error consequences
Adverse conditions

5.7 Loggning av interaktion

Loggning av interaktion innebär att systemet antingen registrerar vissa utvalda, eller samtliga av användarens interaktioner med systemet. Eftersom loggning kan ske automatiskt är det inte nödvändigt att observatörer är närvarande vid inspelningstillfället. Loggning anses därför allmänt som den minst påträngande observationsmetoden.

Enligt Preece m.fl. (1994) finns det två huvudsakliga metoder för loggning:

1. *Tidsmärkta tangenttryckningar* – skapar en lista med exakt tid för varje tangenttryckning.
2. *Interaktionsloggning* – inspelningen görs i realtid, och kan spelas upp i realtid så att observatören kan se interaktionen mellan användaren och systemet.

Loggning av interaktion	
Krav	Beskrivning
När i utvecklingsprocessen	Vid sluttestning, uppföljningsstudier och vid fältstudier av system i drift (Nielsen, 1993). Loggning kan även användas vid användarstudier av prototyper.
Resurser	Låg kostnad för själva datainsamlingen, eftersom loggning inte kräver att någon observatör är närvarande. Kan vara kostsamt att implementera loggningsfunktioner i systemet. Stora datamängder kräver vanligen stora insatser för dataanalys. Detta arbete kan underlättas vid användning av automatiserade analysfunktioner.
Förmåga	Kräver vanligen viss teknisk kunskap om systemet. Dessutom programmeringskunskaper eller annan teknisk kunskap för att implementera loggningsfunktioner.
Med/utan användare	Med användare. Nielsen (Nielsen, 1993) rekommenderar minst 20 användare.
Teoretisk bakgrund	Loggning av interaktion har alltid varit populärt bland vissa forskare. Bl.a. eftersom forskare inte behöver vara på observationsplatsen under hela studien, och eftersom resultatanalysen ofta kan automatiseras (Preece, m.fl., 1994). Loggning av interaktion används vanligen för utvärdering av system i drift under fältstudier, men kan även användas som kompletterande metod under utvärdering av prototyper med användare (Nielsen, 1993).
Mognadsgrad	Utveckling av datorer och programvara har medfört att loggning kan göras mer selektivt. Jämfört med tidigare kan man lättare ange villkor för när loggning skall göras.
Reliabilitet	Hög. Samma handlingar ger samma mätdata.
Validitet	Loggning anses allmänt som den minst påträngande observationsmetoden. Trots detta finns alltid risk att vetskap om att loggning sker påverkar användarnas prestation. Eventuell påverkan kan dock antas störs i början av loggningen, medan användarna sannolikt efter en tid till och med glömmer bort att loggning sker (Faulkner, 2000), vilket då ger tillförlitligare data. Vid loggning kommer data ofta från enkla händelser på låg nivå, exempelvis en knapptryckning, medan knapptryckningen kan föregås av en lång och sammansatt process av perception och beslutsfattande. Det är därför vanligen inte antal knapptryckningar i sig som är det intressanta, utan knapptryckningarnas innebörd för användarens interaktion med systemet (Hennesy, 1990). Dvs. att tolka resultatets betydelse för systemets användbarhet kan vara problematiskt.
Sensitivitet/Känslighet	Hög känslighet för att visa hur frekvent olika delar av systemet, olika funktioner etc., har använts.
Diagnosticitet	Låg. Det krävs ofta kompletterande information för att få veta varför vissa delar av systemet eller funktioner använts mer eller mindre frekvent, och på vilket sätt de eventuellt varit problematiska.
Typ av fel som hittas	Identifiering av hur frekvent olika delar av systemet används (Nielsen, 1993). Loggfiler kan också visa var användaren gjort pauser eller på annat sätt ödslat tid, och vilken typ av fel som var mest frekventa (Nielsen, 1993)

Resultatanalys	<p>Loggning ger ofta mycket stora datamängder (Faulkner, 2000). Analys av stora datamängder kräver analysprogram (Nielsen, 1993). Analysprocessen kan dock vanligen automatiseras (Preece, 1994).</p> <p>Ett problem vid resultatanalys är att det ofta är svårt att bedöma vilka data som är intressanta, och vilka som kan ignoreras (Faulkner, 2000).</p> <p>Plottning, exempelvis av var på skärmen användarna klickat mest frekvent kan göras (Nielsen, 1993).</p>
Andra metoder som ger motsvarande information	Exakt samma information är svårt att få fram med andra metoder. Yttre observerbara interaktioner kan dock vara möjligt att registrera genom traditionell observationsstudie.
Kompletterande metoder	<p>Vid samtidig loggning, observation och videoinspelning av användarnas interaktioner med systemet kan en mycket god bild av användarens aktiviteter ges. Att synkronisera metoderna kräver dock ganska mycket förberedelser. Det är också synnerligen tidskrävande att analysera de stora datamängder som då genereras (Preece m.fl., 1994).</p> <p>Loggning visar endast vad användarna gjort, inte varför de gjorde det. Intervjuer kan därför efter avslutad loggning, eller vid pauser under arbetet, användas för att samla in kompletterande information. Vid behov kan man visa användarna delar av loggningen, eller insamlad data, för att belysa det man vill veta mer om (jämför retrospektiv testning).</p>
Tillämpbarhet	<p>Det kan vara kostsamt att implementera loggningsfunktioner i ett system.</p> <p>Går att använda i de flesta fall, när det finns ett system.</p>
Störande/Påträngande	<p>Loggning kan upplevas kränkande avseende användarens integritet (Nielsen, 1993).</p> <p>Om loggning görs på användarens arbetsplats och användarens fullständiga interaktioner loggas, så att det i efterhand går att se exakt vad användaren gjort, skall man garantera att ingen på företagets ledning får tillgång till loggningen, och helst möjliggöra för användaren att stänga av loggningen när känsliga uppgifter utförs (Nielsen, 1993).</p> <p>Om användaren garanteras att insamlade data behandlas konfidentiellt uppfattas loggning dock mindre påträngande jämfört med observation, eftersom forskaren inte behöver vara närvarande under deras arbete.</p>
Användares acceptans	Hög.
Referenser	<ul style="list-style-type: none"> • Faulkner, X. (2000). <i>Usability engineering</i>. Basingstoke: Macmillan Press Ltd. • Hennessy, R. T. (1990). Practical human performance testing and evaluation. In H. R. Booher (Ed.), <i>MANPRINT an approach to systems integration</i> (pp. 433-470). New York: Van Nostrand Reinhold. • Nielsen, J. (1993). <i>Usability engineering</i>. Boston: Academic Press, Inc. • Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). <i>Human-Computer Interaction</i>. Wokingham: Addison-Wesley.

5.8 Observation

Observation innebär att en eller flera personer observerar användare när dessa löser sina arbetsuppgifter. Dokumentation sker vanligen genom att observatören för anteckningar, spelar in användaren på video/ljudband, eller genom loggning av användarens interaktioner med systemet. Om dessa metoder kombineras kan de ge en mycket god bild av användarens aktiviteter. För att synkronisera metoderna krävs dock ganska mycket förberedelser. Det är också synnerligen tidskrävande att analysera de stora datamängder som genereras (Preece m.fl., 1994).

Anteckningar är det enklaste sättet att dokumentera observationer och kräver ingen annan utrustning än papper och penna. Motsvarande dokumentation kan dock även göras med bärbar- eller handdator. Vilken typ av anteckningar som görs är beroende av analysens nivå. Hackos och Redish (1998) skiljer mellan hög och låg analysnivå:

- *Analys på hög nivå* – av användare, uppgifter och omgivning där man försöker fånga generella intryck snarare än prestationens specifika detaljer. Kan innebära ett mindre formellt sätt att anteckna.
- *Analys på låg nivå* – av hur användare löser sina uppgifter. Kan kräva förberedda formulär där användarens handlingar och beslut prickas av.

Observation	
Krav	Beskrivning
När i utvecklingsprocessen	Vid uppgiftsanalys, uppföljningsstudier (Nielsen, 1993).
Resurser	Relativt stor tidsåtgång, i synnerhet för resultatanalys. Ett i förväg testat observationsschema underlättar resultatanalysen.
Förmåga	God observationsvana krävs, ofta även ämnesområdeskompetens avseende observerad verksamhet. Till viss del går det att kompensera mindre vana och ämnesområdeskompetens med tidsmässigt längre observationsstudier.
Med/utan användare	Med användare.
Teoretisk bakgrund	God teoretisk bakgrund. Metoden används främst vid fältstudier och innebär vanligtvis minimal påverkan på användarnas normala arbete.
Mognadsgrad	Har använts länge.
Reliabilitet	Låg experimentell kontroll (Nielsen, 1993). Hög belastning på observatören medför risk för reliabilitet. Om observationen spelas in på video kan problemet med observatörens belastning undvikas (se beskrivning av <i>Observation med video</i>). Granskning i efterhand kan då även göras tillsammans med användaren (se beskrivning av <i>Retrospektiva metoder</i>). Vid inspelning på video kan reliabilitet hos observationsdata dessutom ytterligare ökas genom att mer än en expert ges möjlighet att granska den videoinspelade observationen (Hennessy, 1990). Om ett i förväg testat observationsschema används minskar belastningen på observatören, vilket höjer reliabiliteten.
Validitet	God ekologisk validitet, visar användares verkliga uppgifter (Nielsen, 1993). Vid observation finns dock risken att observatörens närvaro medför att beteendet hos de människor som observeras omedvetet avviker från deras normala beteende, vilket vanligen kallas för Hawthorne-effekten (ex., Preece m.fl., 1994). Observatören bör därför vara så tyst och osynlig som möjligt, så att användarna arbetar med systemet som vanligt (Nielsen, 1993).
Sensitivitet/Känslighet	Känslig för kvalitativa skillnader mellan hur olika personer använder ett system. Mindre känsligt för trender, förändringar i tiden.
Diagnosticitet	God förmåga att upptäcka oförutsedda användarproblem som forskaren inte kunnat förutse, och därför inte skulle designat för i en traditionell laboratoriestudie.
Typ av fel som hittas	Användarproblem i verkliga arbetssituationer. Kan ge uppslag avseende funktioner och egenskaper hos systemet (Nielsen, 1993).
Resultatanalys	Huvudsakligen kvalitativ analys, vilket vanligen medför stor arbetsinsats. Kvantitativ analys är dock möjlig, exempelvis notering av hur frekvent viss funktion har använts.
Andra metoder som ger motsvarande information	Observation med video, retrospektiv testning. Insamling av observationsdata är mycket arbetskrävande. En person kan därför inte förväntas samla in stora mängder av olikartade data utan att observationens reliabilitet och precision försämras (Charlton & O'Brien, 1996). Hackos och Redish (1998) föreslår därför att observatören som komplement till anteckningar under pauser, medan det observerade fortfarande finns i minnet, exempelvis går till en avskild plats och talar in kommentarer på en bärbar bandspelare. Alternativt kan

Andra metoder som ger motsvarande information (forts. från föreg. sida)	observationen spelas in på video (se <i>Observation med video</i>). Ibland kan bandinspelning av enbart ljud vara tillräckligt, exempelvis när användarnas interaktioner huvudsakligen består av verbal kommunikation, eller när studerad interaktion sker med eller vid objekt som ej får filmas.
Kompletterande metoder	Vid samtidig inspelning på video och loggning av interaktionen, kan en mycket bra bild av användarnas aktiviteter erhållas. Intervjuer för klarläggande, exempelvis av handlingar som observatören inte uppfattat eller förstått. Intervjuer, enkäter för att samla in användarnas subjektiva värderingar.
Tillämpbarhet	Kan tillämpas då arbetsmiljön medger att en observatör finns närvarande utan att störa, och då det är möjligt att observera vad som faktiskt händer. Kan därför vara svårt att använda då arbetsuppgiften är huvudsakligen kognitiv.
Störande/Påträngande	Att bli iakttagen under arbete kan upplevas som störande, vilket också kan påverka användarnas prestation både negativt och positivt (Hawthorne effekten). Vid så kallade användarlaboratorier sitter ofta observatörer bakom ljudisolerade väggar med spegelglas, vilket gör deras närvaro mindre påträngande (Nielsen, 1993)
Användares acceptans	Att bli observerad kan upplevas som störande. Acceptansen blir i allmänhet större om studiens syfte förklaras.
Referenser	<ul style="list-style-type: none"> • Charlton, S. G., & O'Brien, T. G. (1996). The role of Human Factors testing and evaluation in systems development. In T. G. O'Brien & S. G. Charlton (Eds.), <i>Handbook of human factors testing and evaluation</i> (pp. 13-26). Mahwah, New Jersey: Lawrence Erlbaum Associates. • Hackos, J. T., & Redish, J. C. (1998). <i>User and task analysis for interface design</i>. New York: Wiley. • Hennessy, R. T. (1990). Practical human performance testing and evaluation. In H. R. Booher (Ed.), <i>MANPRINT an approach to systems integration</i> (pp. 433-470). New York: Van Nostrand Reinhold. • Nielsen, J. (1993). <i>Usability engineering</i>. Boston: Academic Press, Inc. • Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). <i>Human-Computer Interaction</i>. Wokingham: Addison-Wesley.

5.9 Ostrukturerad intervju

Ostrukturerade (flexibla) intervjuer har vanligen ett visst förutbestämt innehåll, men ingen bestämd ordningsföljd; intervjuaren har frihet att följa upp svaren och ta reda på användarens personliga attityder (Preece, m.fl., 1994). Intervjutekniken bygger till stor del på öppna frågor, det vill säga frågor som inte kan besvaras med exempelvis ”ja” eller ”nej”, utan kräver någon typ av förklaring.

Ostrukturerade intervjuer används i huvudsak då intervjuaren vill få information som inte går att komma åt med andra metoder. Metoden medger att användaren kan uttrycka sina tankar och utveckla resonemang och innehåll.

Ostrukturerad intervju	
Krav	Beskrivning
När i utvecklingsprocessen	Intervjuer kan användas under samtliga designstadier (Popovic, 1999). Ostrukturerade intervjuer är i allmänhet att föredra när utvärderare är i behov av mer kunskap om användarna, eller under tidiga designstadier för att utvinna mer allmän information om användarna, deras uppgifter och arbetsmiljö (Faulkner, 2000).
Resurser	Relativt stor tidsåtgång, både för datainsamling och analys. Ingen dyrbar teknisk utrustning. Vid ostrukturerade intervjuer kan det vara svårt att föra anteckningar, varför Faulkner (2000) föreslår kompletterande dokumentation genom ljudupptagning med bandspelare eller motsvarande.
Förmåga	För att bli skicklig på att genomföra intervjuer krävs kunskap och övning. Ostrukturerade intervjuer är dessutom den svåraste intervjumetoden och kräver extra stor erfarenhet av den som genomför intervjun.
Med/utan användare	Med användare.
Teoretisk bakgrund	Inom socialvetenskap har ett flertal olika intervjuformer utvecklats, som kan användas vid MSI-utvärderingar, bl.a. ostrukturerad intervju (Faulkner, 2000). Eftersom intervjuaren har möjlighet att anpassa frågorna efter situationen är ostrukturerade intervjuer väl lämpade för explorativa studier, där man inte är säker på vilka svar man söker (Nielsen, 1993).
Mognadsgrad	Har använts länge.
Reliabilitet	Beroende på intervjuarens förmåga. Intervjuarens närvaro kan påverka användaren, så att han/hon kan ge de svar som han/hon tror är de "rätta" (Popovic, 1999).
Validitet	Hög
Sensitivitet/Känslighet	Känslig för användares subjektiva uppfattningar. Följdfrågor ger betydligt bättre möjlighet att få en god bild av en frågeställning, jämfört med strukturerade intervjuer och enkäter.
Diagnosticitet	God förmåga att diagnosticera oförutsedda problem.
Typ av fel som hittas	Klargör idéer och problem som är viktiga för användarna (Popovic, 1999). Problem vid kritiska situationer (Nielsen, 1993)
Resultatanalys	Ostrukturerade intervjuer ger företrädesvis kvalitativ data. Resultatanalysen är därför vanligen krävande. I gengäld ger ostrukturerade intervjuer betydligt rikare information, jämfört med enkäter och strukturerade intervjuer (Preece, 1994), exempelvis information om hur och varför användarna upplever ett system/gränssnitt som problematiskt, och ger därmed ofta uppslag till hur gränssnittet/systemet bör förändras.
Andra metoder som ger motsvarande information	–
Kompletterande metoder	Beroende på när i utvecklingsprocessen ostrukturerade intervjuer används.
Tillämpbarhet	Kan användas i de flesta situationer. Används i huvudsak då intervjuaren vill få information som inte går att komma åt med andra metoder.

Störande/Påträngande	Genomförs vanligen vid naturliga avbrott i arbetsuppgifter, eller efter arbetets slutförande, och uppfattas därför vanligen inte som speciellt störande. Vissa människor kan dock uppfatta alltför utforskande frågor som påträngande.
Användares acceptans	Vanligen hög, förutsatt att intervjuaren uppträder smidigt och professionellt.
Referenser	<ul style="list-style-type: none"> • Faulkner, X. (2000). <i>Usability engineering</i>. Basingstoke: Macmillan Press Ltd. • Nielsen, J. (1993). <i>Usability engineering</i>. Boston: Academic Press, Inc. • Popovic, V. (1999). Product evaluation methods and their importance in designing artifacts. In W. S. Green & P. W. Jordan (Eds.), <i>Human Factors in product design: Current practice and future trends</i> (pp. 26-35). New York: Taylor & Frances. • Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). <i>Human-Computer Interaction</i>. Wokingham: Addison-Wesley.

5.10 Retrospektiv testning

Retrospektiv testning innebär att en användare som interagerar med ett system spelas in på video (jämför, *Observation med video*). Efter genomförd interaktion tittar utvärderaren och användaren på filmen tillsammans (Nielsen, 1993). Enligt Nielsen får användaren under granskningen kommentera vad han/hon gjort under testet. Under tiden kan försöksledaren stoppa bandet och ställa kompletterande frågor.

Retrospektiv testning	
Krav	Beskrivning
När i utvecklingsprocessen	Kan användas för utvärdering av implementerade prototyper, och för utvärdering av färdiga system. Störst nytta antas dock vara för utvärdering av implementerade prototyper, som stöd för omdesign.
Resurser	Retrospektiv testning tar exempelvis dubbelt så lång tid att genomföra som "tänka högt" (Nielsen, 1993). Detta beror främst på att granskningen med användaren i efterhand vanligen tar minst lika lång tid som själva testet. Metoden är därför inte lämpad om användarna får hög ersättning eller har andra kritiska arbetsuppgifter att utföra (Nielsen, 1993). Låga materiella kostnader, endast videokamera krävs.
Förmåga	Domänkunskap krävs. Granskningen i efterhand kan ses som en intervjusituation, varför vana vid att genomföra intervjuer är lämpligt.
Med/utan användare	Med användare.
Teoretisk bakgrund	Retrospektiva metoder används huvudsakligen för att identifiera användares interaktionsproblem, under förhållanden när "tänka högt" ej kan användas. Enligt Hackos och Redish (1998) kan det exempelvis vara när: <ul style="list-style-type: none"> • Uppgiften innebär att användaren kommunicerar med andra människor. • Uppgiften kräver hög koncentration hos användaren. • Observatören mäter hur lång tid det tar att utföra uppgiften. • Användaren arbetar under tidspress. Kommentarer från användare när en videofilm granskas i efterhand är vanligen utförligare, jämfört med kommentarer som ges under tänka högt (Nielsen, 1993) Vid behov kan filmen stoppas, vilket ger observatören möjlighet att ställa följdfrågor.
Mognadsgrad	Har använts relativt länge.
Reliabilitet	Högre reliabilitet, jämfört med tänka högt.
Validitet	Retrospektiv testning förutsätter att användaren, efter testet när filmen granskas, språkligt kan beskriva sina perceptuella och kognitiva aktiviteter, vilket kan begränsas av stor vana med system där användarens handlingarna blivit automatiserade (Kanis, 1999) Jämfört med "tänka högt" är risken större för efterhandskonstruktioner.
Sensitivitet/Känslighet	Har fördelar jämförbara med både observation och intervjuer när det gäller känslighet för att identifiera oförutsedda problem.
Diagnosticitet	Relativt god förmåga att identifiera vad användaren upplever som problematiskt hos gränssnittet/systemet.
Typ av fel som hittas	Huvudsakligen för att identifiera användares interaktionsproblem.
Resultatanalys	Efter genomförd interaktion tittar utvärderaren och användaren på filmen tillsammans (Nielsen, 1993). Enligt Nielsen får användaren under granskningen kommentera vad han/hon gjort under testet. Under tiden kan försöksledaren stoppa bandet och ställa kompletterande frågor. Både kvantitativa och kvalitativa analyser kan göras.
Andra metoder som ger motsvarande information	<i>Tänka högt</i> , under förutsättning att arbetssituationen är sådan att metoden kan genomföras.

Kompletterande metoder	<p>Intervjuer och enkäter för att samla in data om användarnas subjektiva värderingar. Efteranalysen tillsammans med användarna kan dock ses som en typ av intervjusituation.</p> <p>Kvantitativa och psykofysiologiska mätningar.</p>
Tillämpbarhet	<p>Särskilt värdefullt vid testsituationer där det är svårt att hitta representativa användare, eftersom man utvinner väldigt mycket information ur varje enskild användare (Nielsen, 1993).</p> <p>Det är inte alltid det går att få tillstånd att filma ett system.</p>
Störande/Påträngande	<p>Jämfört med tänka högt, betydligt mindre påträngande då testet genomförs.</p> <p>Att bli filmad och iakttagen under arbete kan upplevas som störande, vilket också kan påverka prestationen både negativt och positivt (Hawthorne effekten). Denna risk kan dock rimligen minimeras om videokameran installeras flera dagar före filmning (Preece m.fl., 1994), så att användaren hinner vänja sig.</p>
Användares acceptans	<p>Vanligtvis relativt god, ger goda möjligheter för användaren att uttrycka sina synpunkter. Testningen tar dock relativt lång tid, varför tid för deltagande kan vara problematiskt.</p>
Referenser	<ul style="list-style-type: none"> • Hackos, J. T., & Redish, J. C. (1998). <i>User and task analysis for interface design</i>. New York: Wiley. • Kanis, H. (1999). Design centered research into user activities. In W. S. Green & P. W. Jordan (Eds.), <i>Human Factors in product design: Current practice and future trends</i> (pp. 36-46). New York: Taylor & Frances. • Nielsen, J. (1993). <i>Usability engineering</i>. Boston: Academic Press, Inc. • Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). <i>Human-Computer Interaction</i>. Wokingham: Addison-Wesley.

5.11 Enkätundersökning

Enkätundersökningar är relativt billiga att genomföra och kan distribueras till ett stort antal personer.

Öppna frågor innebär att det inte finns givna svarsalternativ, utan att den som svarar själv formulerar ett svar.

När användare besvarar enkätfrågor om ett system efter genomfört test skall frågorna besvaras innan eventuell diskussion om systemet genomförs. Detta för att undvika att försöksledarens kommentarer påverkar svaren (Nielsen, 1993).

Enkätundersökning	
Krav	Beskrivning
När i utvecklingsprocessen	Kan användas under samtliga stadier i utvecklingsprocesser. Enkäter används ofta tidigt, exempelvis vid användar- och uppgiftsanalyser. Enkäter används också ofta i samband med utvärderingar.
Resurser	Billig undersökningsmetod. Kan med låga kostnader distribueras till ett stort antal personer. Om elektroniska enkäter används krävs tillgång till datorer och utvecklingsverktyg för färdigställande.
Förmåga	För enklare undersökningar under utvärderingar krävs inte så stor förmåga. För välplanerade enkätundersökningar, som korrekt uppfångar intressanta frågeställningar och väl anpassar frågorna till efterföljande statistiska analyser, krävs dock goda kunskaper. För digitala enkäter kan även kunskap om specialanpassad programvara eller programmeringskunskap krävas. Kommersiella enkätprogram är dock ofta relativt lättanvända.
Med/utan användare	Med användare. Enkäter kan även delas ut, eller skickas till andra än slutanvändare, exempelvis beslutsfattare i användarnas organisation.
Teoretisk bakgrund	Används för att insamla användares/försökspersoners subjektiva uppfattningar. Subjektiva uppfattningar samvarierar inte alltid med prestation, men kan ge bra input till eventuella förändringar av systemet/gränssnittet. Ett flertal metoder för subjektiv värdering av situationsmedvetande (t.ex. SART, CC-SART) och mental arbetsbelastning (t.ex. NASA-TLX) kan sägas vara en specialform av strukturerade enkätundersökningar. Enkäter kan delas ut på papper, men också distribueras elektroniskt.
Mognadsgrad	Har använts länge.
Reliabilitet	Högre för digitala enkäter, eftersom kontroller kan läggas in som förhindrar att frågor hoppas över, eller besvaras felaktigt (t.ex. att någon gör mer än en markering på en skattningsskala).
Validitet	Risken finns att svar på frågor som kan uppfattas som känsliga inte är ärliga. Jämfört med intervjuer känner sig de som svarar dock vanligen mindre besvärade av känsliga frågor, när enkäter besvaras (Nielsen, 1993). Om möjligt bör anonyma enkäter användas för att reducera denna risk. När frågeformulär konstrueras finns risken att frågor eller alternativ formuleras otydligt, så att de blir svåra att tolka, vilket kan innebära att svaren blir oanvändbara. Innan frågeformulär används bör de därför, som all annan metodik för användbarhetsstudier, testas genom pilotstudier innan de används vid utvärderingar (Nielsen, 1993). Låg svarsfrekvens kan vara ett problem vid enkätundersökningar.
Sensitivitet/Känslighet	God förmåga att identifiera användarens subjektiva värderingar, exempelvis hur nöjd han/hon känner sig med ett visst gränssnitt.
Diagnosticitet	Beroende på hur många som besvarar enkäten, stor förmåga att hitta samband, trender etc. för studerade frågor.
Typ av fel som hittas	Beroende på användningsområde.

Resultatanalys	<p>Enkäter ger främst kvantitativ analys.</p> <p>Elektroniska enkäter kan ge möjlighet att presentera analyserade resultat direkt efter en undersökning, vilket kan vara till stor fördel vid studier av olika typer av arbetsgrupper, exempelvis staber (se t.ex. Modéer m.fl., 2001a, 2001b).</p> <p>Öppna frågor, och utrymme för egna kommentarer, ger möjlighet till kvalitativ analys. Öppna frågor är till störst nytta när det finns många svarsalternativ, eller när det är svårt att förutse lämpliga alternativ (Charlton, 1996). Det är dock vanligt att de som svarar hoppar över öppna frågor, eller att svaren blir kryptiska och svårtolkade (Nielsen, 1993), vilket medför att bearbetning och sammanställning av resultatet kan bli mycket tidskrävande.</p> <p>Vid konstruktion av enkäter skall endast frågor som behövs i resultatanalysen tas med.</p>
Andra metoder som ger motsvarande information	I vissa fall, strukturerade intervjuer.
Kompletterande metoder	<p>Intervju, exempelvis för att samla in ytterligare synpunkter om de svar som givits.</p> <p>Kvantitativa mätningar, observationer etc., för att identifiera hur användaren faktiskt interagerat med gränssnittet.</p>
Tillämpbarhet	<p>Kan tillämpas tillsammans med samtliga övriga metoder.</p> <p>Krävs ingen annan utrustning än dator och skrivare. För den som svarar krävs endast penna.</p> <p>För digitala enkäter krävs tillgång till dator för de som skall besvara enkäten.</p> <p>Lämpar sig bäst inomhus. Med digitala enkäter på handdator är enkätundersökningar även lämpade för fältstudier (se t.ex. Modéer, Söderberg & Elmquist, 2000).</p>
Störande/Påträngande	<p>Ej speciellt störande, eftersom enkäter vanligen ej behöver besvaras under pågående arbete.</p> <p>Vid frågor som kan uppfattas som känsliga känner sig de som svarar ofta mindre besvärade, jämfört med intervjuer (Nielsen, 1993).</p>
Användares acceptans	<p>Kan vara låg vid omfattande frågeformulär, ta därför inte med fler frågor än nödvändigt.</p> <p>Acceptansen är i allmänhet högre då digitala enkäter används.</p>
Referenser	<ul style="list-style-type: none"> • Charlton, S. G. (1996). Questionnaire techniques for test and evaluation. In T. G. O'Brien & S. G. Charlton (Eds.), <i>Handbook of human factors testing and evaluation</i> (pp. 81-99). Mahwah, New Jersey: Lawrence Erlbaum Associates. • Modéer, B., Söderberg, H., & Elmquist, K. (2000). <i>Datainsamling med elektroniska enkäter</i> (FOA-R--00-01614-505-SE). Linköping: Försvarets Forskningsanstalt. • Modéer, B., Wikberg, P., Oskarsson, P.-A., & Lundin, J. (2001a). <i>Nya metoder för datainsamling och värdering av stabsarbete (PM-ESÖ2)</i>. FOI Memo 01-1763/L). • Modéer, B., Wikberg, P., Oskarsson, P.-A., & Lundin, J. (2001b). <i>Mätning av samverkan vid experimentell simuleringsövning (PM-ESÖ3)</i>. FOI Memo 01-2716/L). • Nielsen, J. (1993). <i>Usability engineering</i>. Boston: Academic Press, Inc.

Ytterligare beskrivning av Enkätundersökning

Vanligtvis används slutna frågor med i förväg definierade alternativ. Vid den enklaste formen finns följande svarsalternativ:

Ja Nej Vet ej

Vid skattningsskalor med flera alternativ väljer den som svarar mellan olika alternativ (vanligen används fem till sju alternativ). Fem alternativ kan sägas underutnyttja människans förmåga att diskriminera. Sju alternativ anses därför ofta vara ett lämpligt antal, samtidigt som antalet inte är större än att den som svarar kan göra en rimlig värdering av varje alternativ. Vanligtvis är det mest intuitivt att ett högre värde på skalan innebär större samtycke med frågeställningen, exempelvis:

Var funktionen lätt att förstå? Nej inte alls 1 2 3 4 5 6 7 Ja mkt lätt

Detta sätt att formulera frågan medför dock risk att samtliga som tyckt funktionen varit svår att förstå svarar med en etta, eftersom uppfattningen om funktionen kan vara normalfördelad mellan mycket lätt och mycket svårt, varför det kan vara säkrare att formulera frågan enligt följande (jmf. semantiska differentieringsskalor):

Var funktionen lätt att förstå? Nej mkt svår 1 2 3 4 5 6 7 Ja mkt lätt

Skattningsskalor där varje alternativ är ett påstående kallas för Likertskalor. Exempelvis:

Funktionen var användbar:

_____	_____	_____	_____	_____	_____	_____
misstycker fullständigt	misstycker	misstycker delvis	ingen uppfattning	instämmer delvis	instämmer	instämmer fullständigt

Ibland utelämnas det neutrala mittvärdet, ”ingen uppfattning”, för att de som svarar skall tvingas till att ta ställning. Nackdelen med att utelägna mittkategorin är dock enligt Faulkner (2000) att vissa svar då kan avspegla artificiella uppfattningar om något som de som svarar inte har någon riktig uppfattning om.

Semantiska differentieringsskalor innebär att adjektiv används som motsatsord för att representera motsatta åsikter. Exempelvis:

Hur tycker du att funktionen var?

	Extremt	Mycket	Något	Neutral	Något	Mycket	Extremt	
Svår								Lätt
Otydlig								Tydlig

Vid rangordnade frågeformulär får den som svarar rangordna olika svarsalternativ.

Risken finns att svar på frågor som kan uppfattas som känsliga inte är ärliga. Jämfört med intervjuer känner sig de som svarar dock vanligen mindre besvärade av känsliga frågor, när enkäter besvaras (Nielsen, 1993). Anonyma enkäter kan dock användas för att reducera denna risk.

5.12 Gemensam lägesuppfattning

Metoden har utvecklats på FOI på institutionen för Människa Maskin Interaktion (MSI) i syfte att mäta vissa egenskaper av gemensam lägesuppfattning hos befattningshavare under stabsövningar. Stabsmedlemmar som deltar i en övning får besvara frågor i en digital enkät om vilka tre inspel (information om händelser, underrättelser etc. genererade av övningsledningen) de anser ha varit viktigast för stabens förmåga att lösa sin uppgift. De får också besvara frågor om vilka tre inspel de anser ha varit viktigast för var och en av de övriga befattningshavarna. De får även ange vikten (1-10) för respektive angivet inspel.

Försök har även gjorts där metoden och den digitala enkäten använts för att bedöma stabsmedlemmars förmåga att handla enligt chefens intentioner.

Med det utvecklade programmet genereras befattningshavarnas resultat, i diagram- och tabellform, direkt efter att enkäten har besvarats. Resultaten kan således ge underlag vid genomgång direkt efter genomförd övning.

Mått på gemensam lägesuppfattning	
Krav	Beskrivning
När i utvecklingsprocessen	Bör kunna användas i samtliga utvecklingsstadier för att studera samstämmighet i arbetsgrupper/staber etc. Dvs. metoden bör även kunna prövas för mätning av gemensam lägesuppfattning för andra typer av arbetsgrupper än staber, exempelvis stridsvagnsbesättningar.
Resurser	Kräver tillgång till specialanpassad programvara. På FOI, Ledningssystem, Institutionen för MSI finns en fungerande demoversion, skriven i Visual Basic i Excel, för 2-7 befattningshavare.
Förmåga	Användning kräver endast grundläggande vana vid Excel. Anpassning av programmet kräver grundläggande programmeringskunskaper i Visual Basic och kalkylhantering i Excel.
Med/utan användare	Med användare (de som besvarar enkäten).
Teoretisk bakgrund	Metoden har utvecklats på FOI, Ledningssystem, Institutionen för MSI, för att studera vissa egenskaper hos befattningshavares gemensamma lägesuppfattning under stabsövningar (Modéer & Oskarsson, 2000; Modéer m.fl., 2001a, 2001b). Metoden har utvecklats med utgångspunkt från metoder för att beräkna grupperns ”överlappning” och ”kalibrering”, som också utvecklats på FOI, Ledningssystem, Institutionen för MSI (Andersson, m.fl., 2002).
Mognadsgrad	Metoden är nyutvecklad och därför oprövad.
Reliabilitet	Har ej studerats, men kan förväntas vara på motsvarande nivå som andra subjektiva metoder.
Validitet	Metoden har endast prövats vid ett begränsat antal stabsövningar. Studier av korrelation med andra metoder saknas.
Sensitivitet/Känslighet	Den subjektiva skattningen ger känslighet för subjektivt upplevda skillnader mellan olika befattningar, dock beroende på hur väl valda frågealternativen är. Metoden är dock ej väl utprovad, varför det kan vara svårt att veta hur relevant måttet är för gemensam lägesuppfattning.
Diagnosticitet	Upplägget med diskussion efter övning, med metodens resultat som underlag, ger möjlighet till bedömning av skillnader och likheter mellan olika befattningar. Beroende på att metoden ännu är oprövad kan det dock vara svårt att avgöra hur väl identifierade skillnader och likheter avspeglar verkliga förhållanden.
Typ av fel som hittas	Pekar på möjliga brister i gemensam lägesuppfattning. Att resultaten presenterats direkt efter genomförd övning har bidragit till diskussionsunderlag under efterföljande debriefing, vilket inneburit att exempelvis skiljaktigheter mellan olika typer av befattningar kunnat belysas. Metoden bör även kunna generaliseras till uppskattning av gemensam lägesuppfattning hos andra arbetsgrupper än staber.
Resultatanalys	Mått ges på hur varje stabsmedlem lyckats bedöma informationsbehovet hos övriga medlemmar i staben. Resultaten redovisas dels i ett stapeldiagram där de olika befattningshavarnas resultat jämförs med varandra, dels i ett stapeldiagram där varje befattningshavares överensstämmelse med var och en av de övriga stabsmedlemmarna visas. Dessutom har frekvens valda alternativ för olika befattningar visats. (Modéer m.fl., 2001a, 2001b). Efter att dessa referenser publicerats har metoden vidareutvecklats så att även poäng kan ges för beslätade svarsalternativ. Programmet kan analysera resultaten direkt efter att enkäterna besvaras,

Resultatanalys (forts. från föreg. sida)	vilket möjliggör anslutande resultatredovisning. Då analys görs med utvecklad programvara krävs endast små insatser. Vid manuell analys krävs dock stor arbetsinsats.
Andra metoder som ger motsvarande information	–
Kompletterande metoder	Samtal, eller debriefing, efter datainsamling är lämplig. Eftersom mått på gemensam lägesuppfattning kan ges direkt efter avslutad övning kan detta vara en bra utgångspunkt för diskussion, exempelvis avseende eventuella skiljaktigheter eller likheter mellan olika deltagare.
Tillämpbarhet/Implementeringskrav	Fungerande demoversion finns, för 2–7 användare. Dator med Windows och Excel installerat krävs. Metoden kräver förbestämda svarsalternativ, vilket begränsar tillämpbarheten vid fältstudier. I demoversionen finns dock möjlighet att enkelt lägga in svarsalternativ strax innan enkäten delas ut, vilket gör att det under pågående fältstudie går att bestämma svarsalternativ. För att resultaten skall kunna redovisas direkt efter att enkäten besvaras är det bra om datorerna är sammankopplade i ett nätverk. Nätverk är dock inget krav, svarsfilerna kan även skickas via mail, eller sparas på diskett, för att överföras till den dator där analysen genomförs. Vid behov kan motsvarande undersökning genomföras med pappersenkäter (se Modéer & Oskarsson, 2000), analys får dock i så fall göras i efterhand.
Störande/Påträngande	Vid genomförda stabsövningar har enkäten besvarats vid avbrott eller i samband med spelslut och har därför inte uppfattats som störande.
Användares acceptans	Under genomförda övningar har deltagarna varit mycket nöjda med att få resultatet presenterat direkt efter att de besvarat enkäten. Risk finns dock att stabsmedlemmar med lågt resultat upplever det besvärande att detta redovisas inför hela arbetsgruppen.
Referenser	<ul style="list-style-type: none"> • Andersson, J., Berggren, P., Castor, M., Magnusson, S., & Svensson, E. (2002). <i>Instrumentutveckling för mätning av: Gruppens prestationspotential</i> (FOI-R--0429--SE). Linköping: FOI (Totalförsvarets Forskningsinstitut). • Modéer, B., & Oskarsson, P.-A. (2000). <i>Mätning av stabers lägesuppfattning</i> (FOA-R--00-01815-505-SE). Linköping: Försvarets Forskningsanstalt. • Modéer, B., Wikberg, P., Oskarsson, P.-A., & Lundin, J. (2001a). <i>Nya metoder för datainsamling och värdering av stabsarbete (PM-ESÖ2)</i>. FOI Memo 01-1763/L). • Modéer, B., Wikberg, P., Oskarsson, P.-A., & Lundin, J. (2001b). <i>Mätning av samverkan vid experimentell simuleringsövning (PM-ESÖ3)</i>. FOI Memo 01-2716/L).

Ytterligare beskrivning av Gemensam lägesuppfattning

Formel för beräkning av respektive befattningshavares gemensamma lägesuppfattning med övriga stabsmedlemmar:

$$L(\text{inspel.bedömning}) = \frac{\sum_{j=1}^{N-1} \sum_{i=1}^n \left[p_i \left(1 - \frac{\text{abs}[qA_{i,j} - qB_{i,j}]}{10} \right) \right]}{N-1}$$

$L(\text{inspel.bedömning})$: varje befattningshavares gemensamma lägesuppfattning med avseende på hur väl de lyckats förutsäga de andra befattningarnas uppfattning av respektive inspel.

n : antal bedömda inspel.

N : antal befattningshavare i staben som besvarat enkäten .

p_i : graden av överensstämmelse mellan två jämförda inspel (samma inspel = 1; inspel av samma kategori = 0,5; olika inspel = 0).

$qA_{i,j}$: vikten på den information som befattning A tror att befattning B_j angivit för inspel i.

$qB_{i,j}$: vikten på den information som befattning B_j angivit sig ha för inspel i.

5.13 Kvantitativa prestationsmått

De huvudsakliga kvantitativa måtten på prestation är:

- *Mått på hastighet eller tid* – Exempelvis hur fort värden skrivs in i en dator, hur lång tid det tar att bedöma sin position på en elektronisk karta.
- *Mått på precision eller fel* – Exempelvis sannolikheten att korrekta värden skrivs in i en dator, hur stort felet är i positionsbestämning på en elektronisk karta.

Kvantitativa prestationsmått omfattar även psykofysiologiska mått på mental belastning (se t.ex. hjärtfrekvens och HRV).

Kvantitativa prestationsmått	
Krav	Beskrivning
När i utvecklingsprocessen	För loggning av prestation krävs fungerande prototyp, eller implementering i simulator. Därför vanligen relativt sent under utvecklingsprocessen. Explorativa undersökningar, exempelvis i simulator, kan dock göras tidigt innan prototyp finns utvecklad, exempelvis för att utforska grundläggande egenskaper för prototyp/system.
Resurser	Loggning av prestation vid användning av befintlig prototyp eller simulator kräver vanligen inte stora resurser. Om simulator eller prototyp behöver implementeras för att mätning av prestation skall kunna genomföras krävs givetvis stora resurser.
Förmåga	Kräver vanligen teknisk kunskap för implementering av loggningsfunktioner och dylikt. Enklare tids- och prestationsmått kan dock registreras manuellt.
Med/utan användare	Med användare.
Teoretisk bakgrund	Har använts länge. Har traditioner från både psykologi och teknik. Typiska kvantifierbara prestationsmått omfattar (Nielsen, 1997): <ul style="list-style-type: none"> • Tid för att slutföra en uppgift. • Antal, eller andel, uppgifter av olika typer som slutförs inom en given tidsgräns. • Proportion framgångsrika interaktioner, respektive fel. • Antal fel. • Antal kommandon eller delar av systemet som använts. • Antal kommandon eller delar av systemet som inte använts. • Frekvens användning av manualer eller hjälpfiler, och tidsåtgång för användning av hjälp. • Hur ofta användning av manualer eller hjälpfiler löste användarens problem.
Mognadsgrad	Har används länge, inom många områden.
Reliabilitet	Hög reliabilitet.
Validitet	Vid systemutvärdering finns risk att det som mäts är dåligt relaterat till de faktorer som man faktiskt är intresserad av att utvärdera (Nielsen, 1997). Komplettering med subjektiva metoder ger högre validitet och kan sätta in mätdata i sitt kontextuella sammanhang. Extern validitet är vanligen högre när loggning sker på funktionella system i naturlig miljö, eller i simulator, än då enskilda komponenter av system studeras i laboratorium.
Sensitivitet/Känslighet	Vanligtvis kan mycket små förändringar i mätdata registreras.
Diagnosticitet	Ofta låg förmåga att förklara orsaken till problem. Ger vanligen inte svar på vad som bör förändras för att eliminera problemet.
Typ av fel som hittas	Hög tidsåtgång, felfrekvens, precision etc.
Resultatanalys	Erhållna data analyseras vanligen med traditionell statistisk analys.
Andra metoder som ger motsvarande information	I princip inga andra metoder. Vissa enkelt observerbara kvantifierbara förlopp kan dock hjälpligt registreras med subjektiva metoder som observation, videoinspelning etc.
Kompletterande metoder	Intervjuer, enkäter, observation, psykofysiologiska mätningar.
Tillämpbarhet	Hög, går i allmänhet implementera så fort det finns fungerande prototyp eller system.

Störande/Påträngande	Vanligen inte varken störande eller påträngande, såvida användaren inte upplever vetskapen om registrering av prestation som besvärande.
Användares acceptans	Vanligen hög.
Referenser	<ul style="list-style-type: none">Nielsen, J. (1997). Usability testing. In G. Salvendy (Ed.), <i>Handbook of human factors and ergonomics</i> (2nd ed., pp. 1543-1568). New York: Wiley.

5.14 Modifierad Cooper-Harperskala

Metoden bygger på subjektiv skattning av arbetsbelastning. Försökspersonen skattar sin arbetsbelastning på en skala från 1 till 10, där 1 motsvarar mycket lätt och 10 motsvarar mycket svårt. Till sin hjälp har försökspersonen en trädstruktur för att differentiera skattningsvärdena. Skattningen kan göras antingen med papper och penna eller med dator, vilket som passar bäst.

Modifierad Cooper-Harperskala	
Krav	Beskrivning
När i utvecklingsprocessen	När en arbetsbelastande situation kan genereras för en användare.
Resurser	Små resurser erfordras.
Förmåga	Låga krav på förkunskaper.
Med/utan användare	Med användare.
Teoretisk bakgrund	Metoden utgår ifrån att arbetsbelastning kan mätas genom subjektiv skattning. Cooper-Harperskalan började användas inom flygdomänen. Wierwille och Casali (1983) modifierade skalan till att omfatta generell mätning av mental arbetsbelastning. Ett beslutsträd används, vilket man får vandra igenom genom att svara ja eller nej på ett antal frågor, för att hamna på ett skattat värde på den mentala arbetsbelastningen.
Mognadsgrad	En mogen metod. Cooper-Harperskalan, började användas 1969 (Cooper & Harper, 1969) och den modifierade Cooper-Harperskalan 1983 (Wierwille & Casali, 1983).
Reliabilitet	Reliabiliteten är god på grund av att ett skattat värde alltid betyder samma sak (har sin motsvarighet i beslutsträdet).
Validitet	Subjektiva skattningar av arbetsbelastning anses ofta ha hög validitet, då en person som uppger sig vara belastad också ofta är det.
Sensitivitet/Känslighet	Metoden är ganska känslig för upplevda förändringar av arbetsbelastningen. Det är ofta svårt att finna en exakt tidpunkt för skiftet i arbetsbelastning, då metoden inte innebär dynamisk mätning av situationen.
Diagnosticitet	Diagnosticiteten är låg. Det är svårt att utifrån denna metod veta vad som ligger bakom skattningarna av arbetsbelastningen.
Typ av fel som hittas	Nivåer av upplevd arbetsbelastning.
Resultatanalys	Statistiska analyser.
Andra metoder som ger motsvarande information	Bedfordskalan påminner om metoden både till konstruktion och användning. NASA-TLX, är en annan metod som bygger på subjektiv skattning av arbetsbelastning.
Kompletterande metoder	Metoden kan med fördel kombineras med psykofysiologiska metoder, till exempel mätning av hjärtfrekvens och ögonrörelser, för att komplettera skattningen med dynamiska mått på arbetsbelastning. Metoden har visat sig korrelera högt med hjärtfrekvens och blinkningsfrekvens (Svensson m.fl., 1997a)
Tillämpbarhet	Metoden är lätt att använda och kan användas i många olika domäner.
Störande/Påträngande	Eftersom det går snabbt att skatta ett värde är det inte så störande. Uppgiften att fylla i skattningen kan ibland även läggas in som ett naturligt moment i uppgifterna som skall utföras.
Användares acceptans	Hög acceptans om mätningen inte sker alltför ofta, eller under kritiska skeden under försöken.

Referenser	<ul style="list-style-type: none">• Cooper, G. E. & Harper, R. P. Jr. (1969). <i>The use of Pilot Rating in the Evaluation of Aircraft Handling Qualities</i>. NASA TN D-5153.• Svensson, E., Angelborg-Thanderz, M., & van Avermaete, J. (1997a). <i>Dynamic measures of pilot mental workload, pilot performance, and situational awareness</i>. Technical Report: VINTHEC-WP3-TR01. NLR, Amsterdam.• Wierwille, W. W. & Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. In <i>Proceedings of the 27th Annual Meeting of the Human Factors Society</i> (Santa Monica, CA: Human Factors Society), pp. 129-133.
------------	--

5.15 NASA-TLX

NASA-TLX är en subjektiv metod för mätning av mental arbetsbelastning. Försökspersonen får skatta sex dimensioner av arbetsbelastning: "Mentala krav", "Fysiska krav", "Temporala krav", "Prestation", "Ansträngning" och "Frustration". Varje dimension skattas på en skala från 0–100, vilket ger ett subjektivt mått på försökspersonens totala arbetsbelastning.

NASA-TLX	
Krav	Beskrivning
När i utvecklingsprocessen	När en arbetsbelastande situation kan genereras för en användare.
Resurser	Små resurser.
Förmåga	Låga krav på förkunskaper.
Med/utan användare	Med användare.
Teoretisk bakgrund	Metoden utgår från att arbetsbelastning kan mätas genom att subjektivt skatta sex faktorer på en skala från 0 - 100: "Mentala krav", "Fysiska krav", "Temporala krav", "Prestation", "Ansträngning" och "Frustration". Försökspersonen får även vikta samtliga faktorer betydelse genom parvisa jämförelser, vilket avgör vilken vikt för den totala mentala arbetsbelastningen respektive dimension ges under resultatanalysen (Hart & Staveland, 1988).
Mognadsgrad	Metoden är mogen och har använts i flera olika domäner.
Reliabilitet	Hög. Reliabiliteten kan ibland vara lägre på grund av bristande definitioner, vilket kan leda till skiftande tolkning av skattningskoncepten.
Validitet	Alla sex dimensionerna utom prestation är starkt relaterade till varandra. Prestation kan utelämnas, med motiveringen att den avspeglar något annat än de övriga dimensionerna.
Sensitivitet/Känslighet	Metoden är känslig för upplevd förändring av arbetsbelastning. Det är dock ofta svårt att finna en exakt tidpunkt för skiftet i arbetsbelastning, då metoden inte innebär dynamisk mätning av situationen.
Diagnosticitet	Eftersom skattningen sker i flera dimensioner, är diagnosticiteten högre, än vid skattning i en dimension. Eftersom flera av de skattade dimensionerna är starkt relaterade till varandra, är diagnosticiteten ändå inte särskilt stor.
Typ av fel som hittas	Nivåer av upplevd arbetsbelastning.
Resultatanalys	Skattningen för var och en av delskalorna kan vägas ihop till ett värde på mental arbetsbelastning. Ofta får försökspersonen genom parvisa jämförelser även bedöma betydelsen av respektive dimension. Detta ger en viktning för respektive dimension som avgör dess betydelse för den totala mentala arbetsbelastningen. Genom att använda samma vikt för samtliga dimensioner vinner man dock i enkelhet och tydlighet och förlorar inte mycket i exakthet.
Andra metoder som ger motsvarande information	Andra subjektiva skattningar av arbetsbelastning, t.ex., SWAT, Modifierade Cooper-Harper och Bedford skalor ger motsvarande information.
Kompletterande metoder	Metoden kan med fördel kompletteras med psykofysiologiska mått på arbetsbelastning för att få en bild av arbetsbelastningsdynamiken. Intervjuer kan komplettera.
Tillämpbarhet	Metoden är tillämpbar i många olika domäner och är lätt att använda.
Störande/Påträngande	Eftersom skattningen sker i flera dimensioner, tar den lite tid. Den kan därför upplevas som störande om den återkommer ofta under försöken. Om den används sparsamt under försöken, eller om den används efter genomgången försök, upplevs den inte som störande.
Användares acceptans	Ganska hög användaracceptans, om metoden inte används för tätt under pågående försök. Man bör även förklara vad varje dimension skall skattas stå för.

Referenser	Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), <i>Human mental workload</i> (pp. 139-183). Amsterdam: Elsevier Science Publishers Company.
------------	--

5.16 Observation med video

Observation med video innebär att en eller flera videokameror filmar användarna när de arbetar med ett system.

Samtidigt kan även data samlas in genom att observatörer iakttar och för anteckningar, eller genom loggning av användarnas interaktioner med systemet. Om dessa metoder kombineras kan de ge en mycket god bild av användarens aktiviteter. För att synkronisera metoderna krävs dock ganska mycket förberedelser. Det är också synnerligen tidskrävande att analysera de stora datamängder som genereras (Preece m.fl., 1994).

Observation med video	
Krav	Beskrivning
När i utvecklingsprocessen	Vid uppgiftsanalys, uppföljningsstudier (Nielsen, 1993). Vid studier av användares interaktioner av system, prototyper etc.
Resurser	Begränsade, om enstaka standardmodell av "hemvideokamera" kan användas.
Förmåga	Observationsvana är en fördel. Eftersom videon oftast inte fångar all relevant interaktion kan egna iakttagelser och noteringar vara av stort värde.
Med/utan användare	Med användare.
Teoretisk bakgrund	Videoinspelning används huvudsakligen när vissa typer av interaktion behöver detaljstuderas (Nielsen, 1993).
Mognadsgrad	Har använts relativt mycket.
Reliabilitet	Hög reliabilitet för själva inspelningen som registrerar det som faktiskt händer. En risk kan vara att en kamerabild vanligen endast registrerar en liten del av användarnas interaktioner. Det är därför inte säkert att det som syns på filmen är det mest väsentliga. Jämfört med vanlig observation kan reliabiliteten vid analys ytterligare ökas genom att mer än en expert ges möjlighet att granska den videoinspelade observationen (Hennessy, 1990).
Validitet	God ekologisk validitet, visar användares verkliga uppgifter. (Nielsen, 1993). Att bli filmad kan uppfattas som störande (Hawthorne effekten) och därmed påverka användarnas beteende. Risken kan dock minskas om videokameran installeras flera dagar före filmning (Preece m.fl., 1994), så att användarna hinner vänja sig.
Sensitivitet/Känslighet	Känslig för kvalitativa skillnader mellan hur olika personer använder ett system. Mindre känsligt för trender, förändringar i tiden.
Diagnosticitet	God förmåga att upptäcka oförutsedda användarproblem som forskaren inte kunnat förutse, och därför inte skulle designat för i en traditionell laboratoriestudie.
Typ av fel som hittas	Användarproblem i verkliga arbetssituationer. Kan ge uppslag avseende funktioner och egenskaper hos systemet (Nielsen, 1993).
Resultatanalys	Det är tidskrävande att studera videoinspelningar. Vid detaljerad analys av videoinspelningar nämns ofta kvoten 5:1 – en timmas videoband kan ta fem timmar, eller t.o.m. en hel dag att analysera (Preece. m.fl., 1994). Inspelningar bör därför helst vara korta. Tid kan även sparas om man vid inspelning noterar när något intressant händer så det vid analys går att snabbspola mellan relevanta delar av inspelningen (Faulkner, 2000). Mindre formella videoinspelningar går dock fortare att analysera, och kan t.ex. vara till stor nytta för att visa tvivlande designers eller konstruktörer vilka problem som användaren faktiskt har. Analys kan även göras i efterhand tillsammans med användare (se <i>Retrospektiv testning</i>). För detaljerad analys av videoinspelningar finns två huvudsakliga analysmetoder (Preece m.fl., 1994): 1) <i>Uppgiftsbaserad analys</i> som syftar till att avgöra hur användaren tacklar uppgifterna; var huvudsakliga problem finns och vad som kan åtgärdas. 2) <i>Prestationsbaserad analys</i> som syftar till att utvinna tydligt definie-

Resultatanalys (forts. från föreg. sida)	<p>rade prestationsmått från insamlade data, av vilka de vanligaste är:</p> <ul style="list-style-type: none"> • Frekvens korrekt utförd uppgift • Uppgiftens tidpunkt • Använda kommandon • Felfrekvens • Tidsåtgång för olika kognitiva aktiviteter, som paus under och mellan kommandon
Andra metoder som ger motsvarande information	I vissa fall, observation utan video.
Kompletterande metoder	<p>Vid samtidig observation och loggning av interaktionen kan en mycket bra bild av användarnas aktiviteter erhållas.</p> <p>Intervjuer för klarläggande, exempelvis av handlingar som observatören inte förstått.</p> <p>Intervjuer eller enkäter för att samla in data om användarnas subjektiva värderingar.</p>
Tillämpbarhet	<p>Videoinspelningar är särskilt fördelaktiga i följande situationer (Hackos & Redish, 1998):</p> <ul style="list-style-type: none"> • När observatörer inte kan vara närvarande i användarens arbetsmiljö (t.ex. kabin i stridsflygplan, i sådan miljö kan det dock vara svårt att få tillåtelse att montera kamera). • När det är svårt att observera användarnas handlingar. • När komplexa arbetsuppgifter utförs. Vid analys kan videobandet spelas upp bild för bild, eller visas i slow motion. • Vid retrospektiv testning, där inspelningen i efterhand går igenom tillsammans med användaren (se beskrivning av <i>Retro-spektiv testning</i>).
Störande/Påträngande	<p>Att bli filmad och iakttagen under arbete kan upplevas som störande, vilket också kan påverka prestationen både negativt och positivt (Hawthorne effekten). Denna risk kan dock rimligen minimeras om videokameran installeras flera dagar före filmning (Preece m.fl., 1994), så att användarna hinner vänja sig.</p>
Användares acceptans	Att bli filmad kan upplevas störande.
Referenser	<ul style="list-style-type: none"> • Faulkner, X. (2000). <i>Usability engineering</i>. Basingstoke: Macmillan Press Ltd. • Hackos, J. T., & Redish, J. C. (1998). <i>User and task analysis for interface design</i>. New York: Wiley. • Hennessy, R. T. (1990). Practical human performance testing and evaluation. In H. R. Booher (Ed.), <i>MANPRINT an approach to systems integration</i> (pp. 433-470). New York: Van Nostrand Reinhold. • Nielsen, J. (1993). <i>Usability engineering</i>. Boston: Academic Press, Inc. • Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). <i>Human-Computer Interaction</i>. Wokingham: Addison-Wesley.

5.17 SAGAT

SAGAT (Situation Awareness Global Assessment Technique) är en subjektiv metod för bedömning av situationsmedvetande som är utvecklad för att användas vid simulering. Vid valda tillfällen i simuleringen fryser man simuleringen och ställer slumpvis utvalda frågor till försökspersonen. Metoden är ursprungligen utvecklad för militär flygsimulering. Frågorna handlar om saker som konstruktören av frågorna anser att försökspersonen bör kunna redogöra för, för att anses ha god situationsmedvetenhet. Det kan handla om att svara på frågor om flygplan i omgivningen, eller om situationer som kan uppkomma genom flygplanens agerande. Genom att jämföra personens svar med ett upprättat facit, kan en bedömning göras av personens situationsmedvetenhet.

SAGAT	
Krav	Beskrivning
När i utvecklingsprocessen	Metoden är utvecklad för att användas vid simulering.
Resurser	Ej så stora. Att implementera metoden i en simulering kan dock ta lång tid första gången.
Förmåga	Domänkunskap och uppgiftsanalys behövs för att själv konstruera frågor. För att använda färdiga frågor behövs inga tidigare erfarenheter, dock behövs viss insikt i problematiken för att inse att de färdiga frågorna fångar relevanta aspekter för situationsmedvetande i försökssituationen.
Med/utan användare	Med användare.
Teoretisk bakgrund	Metoden baseras på en teori om situationsmedvetenhet (Endsley, 1988, 1990, 1995), som beaktar situationsmedvetenhet bestående av tre nivåer: 1) uppfattande av relevanta element, 2) förståelse av vad elementen betyder, och 3) förutsägelse av framtida skeenden.
Mognadsgrad	Metoden är inte gammal. Den har främst använts inom flygsimuleringsområdet, där den är utvecklad.
Reliabilitet	Höga krav ställs på de slumpvis ställda frågorna för att de vid upprepade tillfällen skall ge en rättvisande bild av personens situationsmedvetenhet. Om frågorna är alltför lika, finns också risk för inlärningseffekter, till exempel, att personen efter ett tag börjar bemöda sig att försöka uppmärksamma saker som sannolikt uppkommer i kommande frågor.
Validitet	God, förutsatt att frågorna är relevanta för situationen.
Sensitivitet/Känslighet	Metoden har inte hög känslighet, för det är ofta svårt att avgöra värdet av enskilda frågor. Det är först då flera frågor används som slumpens inverkan blir mindre.
Diagnosticitet	Metoden har låg diagnosticitet, då de bakomliggande orsakerna till varför man svarar fel på frågorna inte avslöjas. Metoden har dock viss diagnosticitet, då den separerar brister i situationsmedvetenhet i tre nivåer utifrån bakomliggande teori (Endsley, 1988).
Typ av fel som hittas	Brister i uppmärksamhet, brister i tolkningen av information och brister i att förutse konsekvenser av den aktuella situationen, kan detekteras.
Resultatanalys	Ett SAGAT-värde räknas fram som mått på personens situationsmedvetenhet, utifrån svaren på frågorna.
Andra metoder som ger motsvarande information	Flera varianter av konceptet att frysa simuleringen för att ställa kontrollfrågor har utvecklats.
Kompletterande metoder	Metoden skiljer sig från metoder som använder sig av personens svar som subjektiva skattningar av situationsmedvetenhet, till exempel, SART. Metoderna SAGAT och SART har även jämförts i flera studier, där de inte har visat sig korrelera. Metoden anses som objektiv, då den jämför personens svar med objektiva förhållanden. Metoden kan även kompletteras med enkäter, intervjuer och psyko-fysiologiska mätningar.
Tillämpbarhet	En simuleringsmiljö som kan skapa en trovärdig situation krävs.
Störande/Påträngande	Att frysa en pågående simulering kan upplevas som störande. Metodens förespråkare anser dock inte att detta påverkar den situationsmedvetenhet som mäts.

Användares acceptans	Metoden har hög acceptans bland användare. Dock kan frysningen av simuleringen, av vissa användare, uppfattas som störande.
Referenser	<ul style="list-style-type: none">• Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT). <i>In Proceedings of the IEEE National Aerospace and Electronics Conference</i>, (pp. 789-795). Dayton, OH.• Endsley, M. R. (1990). Predictive Utility of an Objective Measure of Situation Awareness. <i>In Proceedings of the Human Factors Society - 34th Annual Meeting</i>, (pp. 41-45). Santa Monica, CA: Human Factors Society.• Endsley, M. R. (1995). Measurement of Situation Awareness in Dynamic Systems. <i>Human Factors</i>, 37(1), pp 65-84.

5.18 SART och CC-SART

Situational Awareness Rating Technique (SART) (Taylor, 1990) och den senare utvecklade CC-SART (Taylor, 1995a, 1995b) som också beaktar kognitiv kompatibilitet (Cognitive Compatibility) är metoder för subjektiv skattning av situationsmedvetenhet. Försökspersonen får först genomgå ett slags kalibreringsförfarande, för att sedan skatta sin situationsmedvetenhet utifrån olika dimensioner. Utifrån skattningarna kan man antingen räkna ut ett uppskattat värde på situationsmedvetenheten, eller välja att analysera de bakomliggande orsakerna, genom att beakta varje dimension, eller enskild skattning.

SART och CC-SART	
Krav	Beskrivning
När i utvecklingsprocessen	När en prototyp finns eller något annat som kan simulera en tillräckligt realistisk situation, för att situationsmedvetenheten som mäts är relevant.
Resurser	Små materiella resurser. Att öva sig själv och försökspersonerna för att kalibrera dem, kan ta ganska lång tid.
Förmåga	Ringa eller inga förkunskaper behövs. Metoden bör dock prövas i försök eller genom annan träning innan den används skarpt, om den inte har använts tidigare.
Med/utan användare	Med användare.
Teoretisk bakgrund	Genom att be personen uppskatta, till exempel, krav och förutsättningar för uppmärksamhet och hur väl man förstår situationen, kan ett värde på situationsmedvetenheten tas fram. Skattningarna sker i olika dimensioner. Det finns en 3-dimensionell variant av SART och en 10-dimensionell variant.
Mognadsgrad	Metoden har inte använts så länge.
Reliabilitet	Kalibreringen (övningen) av personerna som skall skatta värdena ökar reliabiliteten.
Validitet	Validiteten anses som hög av metodens konstruktörer. Detta är dock svårt att visa.
Sensitivitet/Känslighet	Det är ofta svårt att finna en exakt tidpunkt för skiftet i situationsmedvetenhet, då metoden inte innebär dynamisk mätning av situationen.
Diagnosticitet	Metoden har hög diagnosticitet, genom att de olika dimensionerna som används reflekterar olika aspekter av försökssituationen.
Typ av fel som hittas	Metoden kan hjälpa till att hitta fel som orsakar försämrad situationsmedvetenhet.
Resultatanalys	Dimensionerna kan analyseras var för sig, eller sättas ihop till ett värde på situationsmedvetenheten.
Andra metoder som ger motsvarande information	Andra metoder för subjektiv skattning av situationsmedvetenhet kan ge liknande information. Den höga diagnosticiteten är kanske metodens största fördel, jämfört med andra metoder för subjektiv skattning av situationsmedvetenhet.
Kompletterande metoder	Metoden kan kompletteras med enkäter och intervjuer för att ytterligare höja diagnosticiteten.
Tillämpbarhet	Metoden kan tillämpas i många domäner.
Störande/Påträngande	Metoden är inte störande eller påträngande, om skattningen inte görs vid ett krävande tillfälle.
Användares acceptans	Metoden kan upplevas som något omständlig och mer krävande av personen som gör skattningen, än andra metoder för subjektiv skattning av situationsmedvetande.

Referenser	<ul style="list-style-type: none">• Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a toll for aircrew systems design. In <i>Situational awareness in aerospace operations</i> (AGARD-CP-478; pp. 3/1-3/17). Neuilly-Sur-Seine, France: NATO-Advisory Group for Aerospace Research and Development.• Taylor, R. M. (1995a). Experimental measures: Performance-based self ratings of situational awareness. In <i>International Conference on Experimental Analysis and Measurement of Situation Awareness</i>, Daytona Beach, 1-3 November, 1995.• Taylor, R. M. (1995b). <i>CC-SART: The Development of an experimental measure of cognitive compatibility in systems design</i>, Report to TTCP UTP-7 Human Factors in Aircraft Environments, Annual Meeting, DCIEM, Toronto, Canada, 12-16 June, 1995.
------------	--

5.19 Strukturerad intervju

Vid strukturerade intervjuer ställs förbestämda frågor på ett förbestämt sätt. Strukturerade intervjuer påminner därför om *enkäter*, med skillnaden att det är en människa som ställer frågorna. Jämfört med *enkäter* har strukturerade intervjuer därför bland annat fördelen att användaren ges möjlighet att uttrycka sina åsikter och förklara sina svar (Salvendy & Carayon, 1997).

Strukturerade intervjuer används i huvudsak då forskaren vet vad som är centralt att få reda på av användaren.

Strukturerad intervju	
Krav	Beskrivning
När i utvecklingsprocessen	Intervjuer kan användas under samtliga designstadier (Popovic, 1999).
Resurser	Dyrt, jämfört med enkäter; betydligt större tidsåtgång. Bortsett från eventuell användning av bandspelare e.d. för ljudupptagning krävs ingen teknisk apparatur.
Förmåga	För att bli skicklig på att genomföra intervjuer krävs kunskap och övning. Strukturerade intervjuer är dock den enklaste intervjumetoden.
Med/utan användare	Med användare.
Teoretisk bakgrund	Inom socialvetenskap har ett flertal olika intervjuformer utvecklats, som kan användas vid MSI-utvärderingar, bl.a. strukturerad intervju (Faulkner, 2000). Strukturerade intervjuer används huvudsakligen när forskaren vet vad som är centralt att få reda på av användaren/försökspersonen.
Mognadsgrad	Har använts länge
Reliabilitet	Beroende på intervjuarens förmåga. Intervjuarens närvaro kan påverka användaren, så att den som intervjuas kan ge de svar som han/hon tror är de "rätta" (Popovic, 1999).
Validitet	Medelgod.
Sensitivitet/Känslighet	Känslig för användares subjektiva uppfattningar. Vissa följdfrågor kan ställas, ger dock ej lika god bild av en frågeställning som ostrukturerad intervju.
Diagnosticitet	God förmåga att diagnosticera oförutsedda problem. Dock bättre förmåga hos ostrukturerad intervju.
Typ av fel som hittas	Beroende på användningsområde.
Resultatanalys	Frågornas fasta struktur medger statistisk behandling av svaren. Kompletterande frågor, där användaren uttrycker sina åsikter och förklarar sina svar, kräver dock ofta kvalitativ analys.
Andra metoder som ger motsvarande information	Enkäter, möjligheten att insamla information genom kompletterande frågor, som inte planerats i förväg, är dock reducerad för enkäter. Ostrukturerade intervjuer, dock sämre möjlighet att samla in data som kan genomgå statistisk analys, och kräver större kompetens av den som genomför intervjun.
Kompletterande metoder	Beroende på när i utvecklingsprocessen strukturerade intervjuer används.
Tillämpbarhet	Kan användas i de flesta situationer. Används huvudsakligen då forskaren vet vad som är centralt att få reda på av användaren/försökspersonen.
Störande/Påträngande	Genomförs vanligen vid naturliga avbrott i arbetsuppgifter, eller efter arbetets slutförande, och uppfattas därför vanligen inte som speciellt störande. Vissa människor kan dock uppfatta alltför utforskande frågor som påträngande.
Användares acceptans	Vanligen hög, förutsatt att intervjuaren uppträder smidigt och professionellt.

Referenser	<ul style="list-style-type: none">• Faulkner, X. (2000). <i>Usability engineering</i>. Basingstoke: Macmillan Press Ltd.• Popovic, V. (1999). Product evaluation methods and their importance in designing artifacts. In W. S. Green & P. W. Jordan (Eds.), <i>Human Factors in product design: Current practice and future trends</i> (pp. 26-35). New York: Taylor & Frances.• Salvendy, G., & Carayon, P. (1997). Data collection and evaluation of outcome measures. In G. Salvendy (Ed.), <i>Handbook of human factors and ergonomics</i> (2nd ed., pp. 1451-1470). New York: Wiley.
------------	---

5.20 *Tänka högt*

Tänka högt innebär att användarna tänker högt medan de utför sina arbetsuppgifter, vilket medför att de under varje sekvens under arbetets gång talar om vad de gör, varför de gör det och vad de tänker på när de gör det (Hackos & Redish, 1998).

Under tänka högt ger användare också ofta olika kommentarer om olika detaljer hos gränssnittet/systemet. Nielsen (1993) menar att det är en fördel att informella åsikter om irritationsmoment, som ofta inte kan uppfångas med andra metoder, kan insamlas. Nielsen varnar dock för att genomföra ändringar av ett gränssnitt/system på grundval av enstaka användares kommentarer, eftersom olika användare ofta har mycket avvikande åsikter.

Tänka högt	
Krav	Beskrivning
När i utvecklingsprocessen	<p>Kan användas före design, t.ex. på äldre eller liknande system, som stöd för uppgiftsanalys.</p> <p>Kan användas under tidig utvärdering av pappersprototyper, senare för utvärdering av implementerade prototyper, och för utvärdering av färdiga system.</p> <p>Störst nytta antas dock vara för utvärdering av implementerade prototyper, som stöd för omdesign.</p>
Resurser	<p>Låg tidsåtgång. Dock förutsatt att analysen begränsas till identifikation av användarens missuppfattningar om gränssnittet etc.</p> <p>Är en billig testmetod (Nielsen, 1993).</p>
Förmåga	<p>Behövs inte speciellt stor förmåga, men resultatet blir bättre, ju större vana forskaren som leder testet har.</p> <p>Kräver domänkunskap.</p>
Med/utan användare	<p>Det är användarna som tänker högt, dvs. utvärderingen görs med användare. Nielsen (1993) rekommenderar 3-5 användare.</p>
Teoretisk bakgrund	<p>Syftet är att komma åt användarnas resonemang, intuition och mentala modeller, men också anledningen till varje steg som de tar i uppgiften, och de beslut de tar när de utför uppgiften. Enligt Nielsen (1993) innebär detta att konkret information om vad som användarna missuppfattar hos ett gränssnitt kan utvinnas, vilket kan ge stöd för omdesign.</p> <p>Att tänka högt innebär att användarna tvingas utföra två saker samtidigt, både själva arbetsuppgiften och att tänka högt under arbetet. Forskning inom kognitiv psykologi har dock visat att människor är dåliga på att bibehålla delad uppmärksamhet under längre tid än några minuter (Preece m.fl., 1994), vilket gör att användarens dialog ofta avtar över tiden. Preece m.fl. menar därför att det är nödvändigt att hitta på olika sätt att stötta användarna om de skall tänka högt. Det är t.ex. vanligt att den som utför studien måste driva på användaren genom att ställa frågor som, "Vad tänker du på nu?" och "Vad tror du att det här meddelandet betyder?" (Nielsen, 1993). Om användaren frågar om han/hon kan utföra ett visst kommando, menar Nielsen (1993) att den som utför experimentet inte skall svara på frågan, utan istället svara med en motfråga av typen "Vad tror du händer om du utför kommandot?". Vidare påpekar Nielsen (1993) att eftersom den som utför en studie skall följa den allmänna principen att inte ingripa i användarens interaktion med systemet, skall frågor som "Vad tror du meddelandet längst ned på skärmen betyder?" inte ställas, såvida användaren inte redan lagt märke till meddelandet.</p> <p>En av fördelarna med att användarna tänker högt under tiden de arbetar är att det efter att en uppgift har utförts kan vara svårt att komma ihåg varför man gjort något, eller vad man tänkte på när man gjorde det (Hackos & Redish, 1998). Dessutom kan efterkonstruktioner och bortförklaringar undvikas, som annars kan förekomma vid samtal med användare efter att en studie avslutats (Nielsen, 1993). Att prata med användarna under tiden som de utför en uppgift är dessutom ofta en framkomlig väg för att komma åt deras kognitiva aktiviteter, och därmed hur de faktiskt utfört uppgiften (Nielsen, 1993).</p>

Mognadsgrad	Användes ursprungligen som forskningsmetod inom psykologi, men har fått ökad användning för praktisk utvärdering av gränssnitt för människa-dator-interaktion (Nielsen, 1993).
Reliabilitet	Olika användare kan tänka på olika sätt.
Validitet	<p>Tänka högt förutsätter förmåga att språkligt uttrycka perceptuella och kognitiva aktiviteter, vilket kan begränsas av stor vana med system där användarens handlingarna blivit automatiserade (Kanis, 1999)</p> <p>Tänka högt, innebär en extra belastning på användaren som får utföra två uppgifter samtidigt, arbetsuppgiften och att tänka högt, och är därför inte en naturlig arbetssituation. Användarnas prestation kan påverkas på två sätt (Nielsen, 1993):</p> <ul style="list-style-type: none"> • Behovet att verbalisera kan göra att arbetet går långsammare. • Verbalisering av tankar kan ge ökad koncentration på arbetsuppgiften, därmed färre fel och effektivare arbete. <p>Metoden är därför inte lämpad för användning när:</p> <ul style="list-style-type: none"> • Samtidig insamling av prestationsmått sker. • Uppgiften kräver hög koncentration. • Arbetet sker under tidspress. • Användaren samtidigt kommunicerar verbalt med andra aktörer.
Sensitivitet/Känslighet	<p>Är känslig för när i tiden en användare börjar tänka annorlunda, dock inte alltid känslig för orsaken till det. Detta kan dock ofta klargöras genom att observatören ställer frågor.</p> <p>Är känslig för att genom användarnas kommentarer identifiera irritationsmoment hos gränssnittet, vilket vanligen inte framkommer vid kvantitativa användartest (Nielsen, 1993).</p> <p>Experter med stor vana vid systemet kan ha svårt att verbalisera sina handlingar (Nielsen, 1993).</p>
Diagnosticitet	God förmåga att identifiera fel som beror på att användaren missförstått gränssnittets funktioner.
Typ av fel som hittas	<p>Användares missuppfattningar om hur gränssnittet fungerar (Nielsen, 1993).</p> <p>När användarna tänker högt ger de ofta kommentarer om systemet/gränssnittet. Dvs. informella åsikter om irritationsmoment som ofta inte kan uppfångas med andra metoder.</p>
Resultatanalys	<p>En stor mängd kvalitativ data kan insamlas från ett relativt litet antal användare (Nielsen, 1993).</p> <p>Ger kvalitativ data om hur användare missuppfattar eventuellt missförstår olika delar av gränssnittet, hur de tolkar olika delar av gränssnittet etc. (Nielsen, 1993).</p>
Andra metoder som ger motsvarande information	Till viss del retrospektiv testning, som kan genomföras i vissa situationer där "Tänka högt" är mindre lämplig (se validitet).
Kompletterande metoder	Intervjuer, för att identifiera orsaker till varför användaren tänkt på ett visst sätt, eller orsaker till varför användaren börjat tänka annorlunda.
Tillämpbarhet	Kan användas i de flesta fall, för att studera medvetna kognitiva processer och problemlösning.

Störande/Påträngande	<p>En onaturlig arbetsituation för användare. Användare med mycket stor vana (experter) kan ha svårt att verbalisera sina aktiviteter (Nielsen, 1993)</p> <p>Att tänka högt kan kännas mindre konstigt för användare om försöksledaren före försöket demonstrerar metoden genom att själv tänka högt för någon enkel arbetsuppgift, eller visar en filmsekvens med en annan användare som tänker högt (Nielsen, 1993).</p>
Användares acceptans	Medelgod
Referenser	<ul style="list-style-type: none"> • Hackos, J. T., & Redish, J. C. (1998). <i>User and task analysis for interface design</i>. New York: Wiley. • Kanis, H. (1999). Design centered research into user activities. In W. S. Green & P. W. Jordan (Eds.), <i>Human Factors in product design: Current practice and future trends</i> (pp. 36-46). New York: Taylor & Frances. • Nielsen, J. (1993). <i>Usability engineering</i>. Boston: Academic Press, Inc. • Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). <i>Human-Computer Interaction</i>. Wokingham: Addison-Wesley.

5.21 Användaranalys

Vissa system används huvudsakligen av särskilda kategorier av användare. För att kunna anpassa systemet för den tänkta användarkategorin bör användaranalys genomföras.

De vanligaste metoderna för att analysera användare är (Faulkner, 2000): Informella och formella diskussioner, observationsstudier, att ta med användare i designteamet, enkäter och intervjuer.

Användaranalys	
Krav	Beskrivning
När i utvecklingsprocessen	Tidigt, innan designprocessen påbörjas.
Resurser	Små materiella resurser, kan dock gå åt en hel del arbetstid för besök av tilltänkta användare.
Förmåga	Lämplig kompetens för att genomföra analysen.
Med/utan användare	Vanligen med användare. I vissa fall, kan analytisk klassificering utan användare genomföras.
Teoretisk bakgrund	<p>För kategorisering av användare finns två huvudsakliga dimensioner: kompetensnivå, och sätt att interagera med systemet.</p> <p>Kompetensnivå kan ges följande indelning (Hackos & Redish, 1998):</p> <ul style="list-style-type: none"> • <i>Nybörjare</i> – är rädda att göra fel, saknar praktisk erfarenhet. • <i>Avancerade nybörjare</i> – gör ej uppgiften regelbundet, börjar få en mental modell, fokuserar på konkreta arbetsuppgifter • <i>Kompetenta</i> – har en pålitlig mental modell av ämnesrådet och systemet, fokuserar på att utföra komplexa uppgifter • <i>Experter</i> – är intresserade av att lära sig begrepp och teorier bortom systemets design, vill ha mer funktioner <p>Sätt att interagera med systemet kan ges följande indelning (Faulkner, 2000):</p> <ul style="list-style-type: none"> • <i>Direkta användare</i> – Användare som personligen använder systemet. (t.ex. en kontorist som själv gör analyser med ett datorprogram). • <i>Indirekta användare</i> – Användare som ber andra människor att använda systemet åt dem (t.ex. en resebyrå som ber resebyrå boka flygbiljetten). • <i>Fjärranvändare</i> – Användare som använder systemet indirekt (t.ex. bankkunder som är beroende av bankens informations-system avseende sina konton). • <i>Stöd-användare</i> – Användare som tillhör ett företags tekniska stöd eller datorstödsavdelning. <p>Nielsen (1993) gör även följande indelning:</p> <ul style="list-style-type: none"> • <i>Systemvana</i> – Vana vid systemet, eller vid motsvarande typer av system (Nybörjare – Experter). • <i>Datorvana</i> – Allmänna datorvana (Minimal datorvana – Mycket stor datorvana). • <i>Domänkunskap</i> – Kunskap om det verksamhetsområde som systemet arbetar med (Okunniga – Kunniga). <p>Indelning av användare i kategorier bör dock ske med försiktighet. När man klassificerar användare så måste man göra det med avseende på kontexten, och vara tydlig om kategoriernas innebörd. Kategorierna skall inte vara statiska, utan formbara verktyg som kan anpassas efter omständigheterna där de används. Kategorierna måste också vara meningsfulla för designteamet. Det kan också vara svårt att kategorisera människor, eftersom människor inte alltid överensstämmer med uppställda regler (Faulkner, 2000).</p>
Mognadsgrad	Har använts länge.
Reliabilitet	Svag, dock mycket beroende kompetens hos den som utför analysen.
Validitet	Beroende på hur representativa analyserade användare är.
Sensitivitet/Känslighet	Kriteriet är ej helt entydigt pga. att metoden används så tidigt att man ofta inte har en klar bild av systemet.

Diagnosticitet	Beroende på hur lätt avsedda kategorier av användare kan särskiljas.
Typ av fel som hittas	Används inte för att hitta fel, utan för att undvika framtida problem som kan uppstå om inte systemet anpassas efter den kategori av människor som förväntas använda det.
Resultatanalys	Framst kvalitativ analys för kategorisering av användare och dess konsekvenser för systemets utformning.
Andra metoder som ger motsvarande information	I vissa fall uppgiftsanalys.
Kompletterande metoder	Uppgiftsanalys.
Tillämpbarhet/Implementeringskrav	Kan användas för utvärdering av användare till alla typer av system och arbetsuppgifter.
Störande/Påträngande	Ej speciellt störande.
Användares acceptans	Stor acceptans. Om analysen görs på användare som faktiskt skall använda systemet kan de känna sig mer delaktiga i utvecklingsprocessen.
Referenser	<ul style="list-style-type: none"> • Faulkner, X. (2000). <i>Usability engineering</i>. Basingstoke: Macmillan Press Ltd. • Hackos, J. T., & Redish, J. C. (1998). <i>User and task analysis for interface design</i>. New York: Wiley. • Nielsen, J. (1993). <i>Usability engineering</i>. Boston: Academic Press, Inc.

5.22 GOMS

GOMS (Goals Operators Methods and Selection rules) är en välkänd typ av uppgiftsanalys för att modellera procedurell kunskap (Preece m.fl., 2002). Metoden används för att beskriva rutinarbete med datorbaserade system, vilket baseras på att tid för prestation förutsägs (McLeod & Sherwood-Jones, 1992) GOMS-modellen förutsätter att den tid som åtgår för att utföra en aktivitet kan uppdelas i tre komponenter: En perceptuell processkomponent (T_p), en kognitiv processkomponent (T_c), och en motorkomponent (T_m). Total tidsåtgång beräknas sedan enligt följande: Total tidsåtgång = $T_p + T_c + T_m$ (McLeod & Sherwood-Jones, 1992).

GOMS	
Krav	Beskrivning
När i utvecklingsprocessen	Kan utföras både tidigt innan prototyp finns, och på implementerat system.
Resurser	Inga speciella materiella resurser krävs. Att skapa GOMS-modeller kan dock vara tidsödande (Seamster, Redding & Kaemf, 1997).
Förmåga	GOMS modeller kan bli synnerligen komplexa. Det krävs också stor kunskap för att framgångsrikt tillämpa GOMS (Seamster m.fl., 1997).
Med/utan användare	Tidiga utvärderingar kan göras analytiskt utan användare. Då prototyp finns kan GOMS-modellen även testas på användare.
Teoretisk bakgrund	<p>GOMS utvecklades i början av åttiotalet av Card, Moran och Newel (1983). GOMS är förkortning för analysens grundläggande komponenter (Goals, Operators, Methods, och Selection rules):</p> <ul style="list-style-type: none"> • <i>Goals (Mål)</i> – det tillstånd som användaren vill uppnå. • <i>Operators (Operatorer)</i> – de kognitiva processer och fysiska handlingar som måste utföras för att mål skall uppnås. Skillnaden mellan ett mål och en operator är att ett mål uppnås, medan en operator utförs. • <i>Methods (Metoder)</i> – inlärd procedureller för att uppnå mål. • <i>Selection rules (Urvalsregler)</i> – för att bestämma vilken metod som ska väljas när det finns mer än en tillgänglig metod för att utföra en given del av en uppgift (Preece, Rogers & Sharp, 2002). <p>Det som GOMS modellerar är procedurella aspekter av användbarhet, vilket avser mängd, samstämmighet och effektivitet hos procedurer som användarna måste följa (Kieras, 1997).</p> <p>För att kunna genomföra en GOMS-analys är det nödvändigt att identifiera de mål som användarna vill uppnå. Dvs. GOMS ersätter inte den mest kritiska processen vid design av användbara system, att förstå användarnas situation, arbetsmiljö och mål (Kieras, 1997).</p>
Mognadsgrad	GOMS har genomgått betydligt mer empirisk testning än andra analytiska modeller för människa-dator interaktion (Luczak, 1997).
Reliabilitet	GOMS är synnerligen användbar då sekvenser av arbetsmoment är kända. Metoden är inte så tillförlitlig när arbetsmomenten är okända, eller då man arbetar med oerfarna användare eller nybörjare (Seamster m.fl., 1997).
Validitet	<p>Risk finns att komponenter i analysen, t.ex. Goals, inte alltid stämmer med verkligheten.</p> <p>Metoden är mindre tillförlitlig när arbetsmomenten är okända, eller då man arbetar med oerfarna användare eller nybörjare (Seamster m.fl., 1997).</p>
Sensitivitet/Känslighet	<p>Analysen kan göras med olika detaljeringsgrad.</p> <p>Stor känslighet för vilken tid det tar att utföra både delar av och hela arbetsmoment.</p>
Diagnosticitet	<p>GOMS kan användas för att förutsäga användares prestation, validera empiriska forskningsresultat, undersöka teoretiska förklaringar till observerad prestation och för att undersöka effekten av olika designalternativ (Seamster m.fl., 1997).</p> <p>GOMS kan ge mycket detaljerad information avseende underliggande kognitiva företeelser. En nackdel är dess bristande flexibilitet (Seamster m.fl., 1997).</p>

Typ av fel som hittas	Resultat från GOMS medger relativt enkel jämförelse mellan olika gränssnitt eller datorsystem (Preece m.fl., 2002).
Resultatanalys	Metoden ger kvantitativa mätdata som kan analyseras med statistiska analyser.
Andra metoder som ger motsvarande information	”Key-stroke level model” är nära besläktad med GOMS, även den utvecklade av Card, Moran och Newell (1980). Modellen beskriver tidsåtgång för en erfaren användare att utföra en uppgift med en given metod på ett interaktivt datorsystem. Tidsåtgången modelleras som en funktion av fyra fysiska motoroperatorer K, key press [tangentryckning]; P, pointing [peka]; H, homing [dra tillbaka handen]; D, drawing a line [rita en linje]; en mental operator M och en operator för systemrespons R: Tidsåtgång = $T_k + T_p + T_h + T_d + T_m + T_r$ (Luczak, 1997).
Kompletterande metoder	Uppgiftsanalys, t.ex. kognitiv uppgiftsanalys och ACTA.
Tillämpbarhet	<p>Det enklaste fallet är att konstruera en GOMS-modell för ett redan existerande system, beroende på att en stor del nödvändig information då kan erhållas från själva systemet (Kieras, 1997).</p> <p>Det är inte nödvändigt att ett system redan är implementerat eller i bruk för att en GOMS-analys skall kunna utföras. Det krävs endast att forskaren kan specificera beståndsdelarna i GOMS-modellen. (Kieras, 1997).</p> <p>Om utvärderingen görs tidigt har forskaren inte några användare att tala med. I så fall går det att samtala med konstruktören/designern för att avgöra avsikter och antaganden om användarnas mål och metoder, och sedan konstruera motsvarande GOMS-modell för att tydliggöra dessa antaganden och för att undersöka deras konsekvenser (Kieras, 1997).</p> <p>GOMS brukar inte användas för utvärdering. Metodens tillämpning begränsas av att den endast tillfredsställande modellerar datorbaserade arbetsuppgifter som består av kortare sekvenser av strikt rutinmässiga uppgifter av inmatning av data (t.ex. tangentryckningar). Metoden syftar även huvudsakligen till att förutsäga experters prestation, och medger inte modellering av felhandlingar. Detta gör det mycket svårt att förutsäga hur en medelanvändare kommer att utföra sina uppgifter (Preece m.fl., 2002).</p> <p>GOMS är synnerligen användbar då sekvenser av arbetsmoment är kända. Metoden är dock inte så tillförlitlig när arbetsmomenten är okända, eller då man arbetar med oerfarna användare eller nybörjare (Seamster m.fl., 1997).</p> <p>Metoden kan även tillämpas på kognitiva arbetsuppgifter, t.ex. beslutsfattande, och då visa tidsåtgång för olika delprocesser.</p>
Störande/Påträngande	Upplevs vanligen inte som störande.
Användares acceptans	Vanligen stor acceptans.

Referenser	<ul style="list-style-type: none"> • S. K., Moran, T. P., & Newel, A. (1980). The keystroke level model for user performance time with interactive systems. <i>Communications of the ACM</i>(23), 396-410. • Card, S. K., Moran, T. P., & Newel, A. (1983). <i>The psychology of human-computer interaction</i>. Hillsdale, NJ: Lawrence Erlbaum Associates. • Kieras, D. (1997). A guide to GOMS Model usability evaluation using NGOMSL. In M. Helander & T. K. Landauer & P. Prabhu (Eds.), <i>Handbook of human-computer interaction</i> (Second, completely revised edition ed., pp. 733-766). Amsterdam: Elsevier Science B.V. • Luczak, H. (1997). Task analysis. In G. Salvendy (Ed.), <i>Handbook of human factors and ergonomics</i> (2nd ed., pp. 340-416). New York: Wiley. • McLeod, R. W., & Sherwood-Jones, B. M. (1992). Simulation to predict operator workload in a command system. In B. Kirwan & L. K. Ainsworth (Eds.), <i>A guide to task analysis</i> (pp. 301-310). London: Taylor & Frances. • Preece, J., Rogers, Y., & Sharp, H. (2002). <i>Interaction design: Beyond human-computer interaction</i>. New York: Wiley. • Seamster, T. L., Redding, R. E., & Kaempf, G. L. (1997). <i>Applied cognitive task analysis in aviation</i>. Aldershot: Ashgate.
------------	--

5.23 Axiomatisk design

Axiomatisk design är namnet på ett försök att skapa: a) ett teoretiskt ramverk för design, b) ett stöd för att göra designbeslut, c) en top-down styrd designprocess uttryckt med matematisk formalism.

Enligt axiomatisk design kan designprocessen beskrivas med hjälp av krav och specifikationer av förhållandet mellan relevanta faktorer/variabler i ”fyra världar”: Kundens behov (Customer needs), Funktionella krav (Functional requirements), Designparametrar, Processvariabler.

Att arbeta med axiomatisk design är ett teoretiskt arbete med att ta fram en kravspecifikation och sedan se hur olika krav och begränsningar i de ”4:a världarna” kan syntetiseras till det bästa designbesluten.

Vid exempelvis design av en ny display skulle kundens behov kunna vara en lättanvänd display med tydlig symbolik. De funktionella kraven skulle bl.a. kunna vara att symbolerna syns (inte klottrig display) och att symbolernas innebörd är tydlig. Designparametrar representerar olika sätt att lösa detta t.ex. olika färgskalor, declutterfunktioner och 2D- eller 3D-symboler. Processvariabler skulle kunna vara att 3D-displayer kanske är bättre, men svårare att konstruera, men att det finns en expertgrupp tillgänglig.

Axiomatisk design	
Krav	Beskrivning
När i utvecklingsprocessen	Axiomatisk design används huvudsakligen under inledande faser av konceptualisering och kravspecifikation av ett nytt system
Resurser	Relativt liten arbetsinsats, inga särskilda materiella resurser krävs.
Förmåga	Det teoretiska ramverket kan tyckas vara onödigt tillkrånglat formulerat.
Med/utan användare	Inga slutanvändare behövs under analysen. Deras input dock viktig för att kunna definiera Kundens behov. Bra resultat/kravspecifikationer förutsätter dock att någon expert finns tillgänglig.
Teoretisk bakgrund	<p>Axiomatisk design introducerades kring 1970 och finns beskriven i ett antal publikationer (t.ex. Suh, 1998)</p> <p>Axiomatisk design är ett försök att skapa: a) ett teoretiskt ramverk för design, b) ett stöd för att göra designbeslut, och c) en top-down styrd designprocess uttryckt med matematisk formalism.</p> <p>En designprocess kan ses som skapandet av syntetiserade lösningar som tillfredsställer de upplevda behoven genom att koppla varje enskilt funktionellt krav till en specificerad designparameter. Axiomatisk design erbjuder då stöd när man skall välja mellan olika designlösningar.</p> <p>Hela processen grundar sig på två axiom som alltid är giltiga. Axiom 1 (Independence Axiom). Detta axiom innebär att de funktionella kraven man sätter upp skall vara fristående från varandra. Detta oberoende är viktigt för att man skall få full frihet att hitta de bästa designparametrarna utan att en ändring i en suboptimerar valet av en annan.</p> <p>Axiom 2 (Information Axiom). Detta axiom används för att välja mellan de designlösningar som passerat Axiom 1. Axiom 2 innebär att man skall välja den designlösning där det behövs minst mängd ytterligare information för att veta om de funktionella krav man satt upp går att genomföra (dvs. om man har en uppsättning funktionella krav där man vet att alla går att uppfylla så är det bättre än att man har en uppsättning krav där man är osäker på om/hur vissa skall uppfyllas).</p> <p>Till dessa två axiom kommer 27 teorem och 8 så kallade naturliga slutsatser (corollaries) som ger ytterligare hjälp.</p>
Mognadsgrad	Mogen. Har dock ännu inte fått full spridning
Reliabilitet	Man gör knappast samma analysarbete igen för samma system så begreppet reliabilitet är inte så givande. Om man däremot gör en ny analys av nytt system så ger teorin och tillvägagångssättet viss stadga och därigenom ett resultat som liknar det tidigare.
Validitet	Upp till analytikernas domämförståelse
Sensitivitet/Känslighet	Upp till analytikerna att välja hur mycket man bryter ned det.
Diagnosticitet	Strukturen i metoden stödjer en hög diagnosticitet.
Typ av fel som hittas	En designlösning där de funktionella kraven är beroende av varandra leder ofta till att slutprodukten blir dålig och svårändrad eftersom förändringar på en designparameter slår igenom och allt måste ändras.
Resultatanalys	Beroende på applikationen kan det vara olika svårt att avgöra om designen är optimal.
Andra metoder som ger motsvarande information	-
Kompletterande metoder	-
Tillämpbarhet/Implementeringskrav	Axiomatisk design erbjuder ett teoretisk ramverk för utvecklandet av en kravspecifikation.

Störande/Påträngande	–
Användares acceptans	–
Referenser	<ul style="list-style-type: none">• Suh, N. P. (1998). <i>Axiomatic Design Theory for Systems. Research in Engineering Design</i>, 10:189-209. London: Springer-Verlag.

5.24 Walkthroughs

Walkthroughs är en samling strukturerade metoder där användbarhetsexperter och/eller designers genom att simulera användarens arbete går igenom ett system. Syftet är att hitta problem tidigt i utvecklingscykeln. För att en walkthrough skall kunna genomföras bör de uppgifter som blivande användare skall utföra omsorgsfullt konstrueras med utgångspunkt från tillgängliga systemspecifikationer eller skisser ("mock-ups") av skärmbilder. En genomgång av de aktiviteter (kognitiva och operationella) som krävs för att gå från en skärmbild till en annan är, enligt Preece m.fl. (1994), ett typiskt exempel på en walkthrough.

Kognitiv walkthrough fokuserar på användarens kognitiva processer. Arbets sättet är nära besläktat med kognitiv uppgiftsanalys (Preece m.fl., 1994). Det kan beskrivas som att experten försöker imitera den tänkta användarens kognitiva arbete med systemet, t.ex. beslutsfattande (Preece m.fl., 2002). Syftet är att försöka upptäcka de problem som användaren skulle råka ut för vid respektive arbetsmoment (Faulkner, 2000).

Pluralistisk walkthrough innebär att systemet går igenom av personer med olika typer av kompetens, som produktutvecklare/designers, användbarhetsexperter och användare. Pluralistisk walkthrough är betydligt mindre formell än kognitiv walkthrough, och kan exempelvis användas för att kontrollera gränssnittets överensstämmelse med vedertagen standard för gränssnittutformning (Preece m.fl., 1994). Att både användare och designers samtidigt deltar i utvärderingen är dessutom fördelaktigt, eftersom det möjliggör en preliminär insamling av användarsynpunkter redan under ett tidigt designstadium (Nielsen, 1993).

Innan experter genomför en walkthrough bör de fastställa exakt vilken uppgift som skall utföras, kontexten i vilken den skall utföras, och vilka antaganden de har om den tilltänkta användarpopulationen (Preece m.fl., 1994). Därefter går de igenom uppgiften, granskar de handlingar som är nödvändiga för att utföra uppgiften, och försöker förutsäga hur användarna sannolikt kommer att bete sig och vilka problem de kommer att möta (Preece m.fl., 1994). Avsikten med walkthroughs vid systemutvärdering är att upptäcka problem mycket tidigt för att därmed också kunna avlägsna dem tidigt.

Termen kognitiv jogthrough används ibland för en snabbare och mindre formell genomgång av systemet än den som görs vid vanlig walkthrough

Walkthroughs	
Krav	Beskrivning
När i utvecklingsprocessen	Används tidigt i utvecklingsprocessen.
Resurser	I den senaste versionen av kognitiv walkthrough är den tidigare relativt långa utvärderingstiden nedkortad (Lewis & Warton, 1997).
Förmåga	Den senaste versionen av kognitiv walkthrough ska kunna användas av alla oavsett formell träning (Lewis & Warton, 1997).
Med/utan användare	De flesta typer av walkthroughs är analytiska utvärderingsmetoder och genomförs utan användare. Pluralistisk walkthrough inkluderar dock användare.
Teoretisk bakgrund	Kognitiv walkthrough skapades i början av 1990-talet med utgångspunkt från teorier om lärande genom utforskning. Sedan dess har metoden modifierats ett antal gånger (Lewis & Warton, 1997).
Mognadsgrad	Metoden har använts sedan början på 1990-talet.
Reliabilitet	Medelhög, dock beroende på utvärderarens kompetens.
Validitet	När en kognitiv walkthrough genomförs av en MSI-expert, finns risk att expertens försök avviker från vad en riktig användare skulle ha gjort. Anledningen är att användare i regel har mindre kunskap om gränssnittets funktionalitet och sammansättning, jämfört med experten. Fallet kan dock även vara det omvända, att vissa uppgifter kan vara svårare för experten, pga. användarnas större domänkunskap (Lewis & Warton, 1997). Vid pluralistisk walkthrough deltar dock även användare i utvärderingen.
Sensitivitet/Känslighet	Låg, då det kan vara svårt att förutse detaljers inverkan. Risk finns att experten som genomför utvärderingen inte fokuserar på samma saker som riktiga användare i naturlig miljö.
Diagnosticitet	Låg, svårt att kartlägga orsakssamband vid brist på detaljerad kunskap om produkt och kontext.
Typ av fel som hittas	I huvudsak övergripande problem.
Resultatanalys	Företrädesvis kvalitativ analys. Enskilda utvärderare kan analyseras separat eller sammanställas i en övergripande analys.
Andra metoder som ger motsvarande information	Till viss del heuristisk utvärdering.
Kompletterande metoder	Heuristisk utvärdering som fokuserar på själva gränssnittet, medan kognitiv walkthrough fokuserar på användarens uppgifter. Metoderna bör därför ses som komplement till varandra, och inte som likvärdiga alternativ (Lewis & Warton, 1997). Användartest för att identifiera allvarliga och återkommande problem (Jeffries m.fl., 1991). Användartestning kan inte genomföras innan en funktionell prototyp finns implementerad. Det är dock viktigt att användartestning genomförs för att identifiera vissa allvarliga fel. Dvs. kognitiv walkthrough ersätter inte riktiga användartest. Uppgiftsanalys, t.ex. kognitiv uppgiftsanalys, hierarkisk uppgiftsanalys, ACTA och användaranalys. Dessa metoder bör användas före en walkthrough och ligga till grund för denna.
Tillämpbarhet	Kan nästan alltid tillämpas. Kräver ingen fungerande prototyp.
Störande/Påträngande	Användare inkluderas endast i pluralistisk walkthrough. Metoden uppfattas då i vanligen inte som speciellt störande.

Användares acceptans	Vid pluralistisk walkthrough där användare inkluderas uppfattar de det i regel positivt att tidigt under utvecklingsstadiet kunna ge synpunkter på ett system som de i framtiden kommer att arbeta med.
Referenser	<ul style="list-style-type: none"> • Faulkner, X. (2000). <i>Usability engineering</i>. Basingstoke: Macmillan Press Ltd. • Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. M. (1991). User interface evaluation in the real world: A comparison of four techniques. In S. P. Roberson & G. M. Olson & J. S. Olson (Eds.), <i>Human Factors in computing systems CHI'91 Conference Proceedings</i> (pp. 119-124). New York: ACM Press. • Lewis, C., & Wharton, C. (1997). Cognitive walkthroughs. In M. Helander & T. K. Landauer & P. Prabhu (Eds.), <i>Handbook of human-computer interaction</i> (Second, completely revised edition ed., pp. 717-732). Amsterdam: Elsevier Science B.V. • Nielsen, J. (1993). <i>Usability engineering</i>. Boston: Academic Press, Inc. • Preece, J., Rogers, Y., Sharp, H., Benyon, D., Simon, H., & Tom, C. (1994). <i>Human-Computer Interaction</i>. Wokingham: Addison-Wessley. • Preece, J., Rogers, Y., & Sharp, H. (2002). <i>Interaction design: Beyond human-computer interaction</i>. New York: Wiley.

5.25 ACTA

ACTA (Applied Cognitive Task Analysis), (tillämpad kognitiv uppgiftsanalys), går ut på att få information om en uppgift genom att intervjua experter inom området. Man har stöd i metoden av texter och tabeller som man rekommenderas använda. (Hutton & Militello, 1997; Klein & Militello, 1997; Militello & Hutton, 1998). Man börjar med att göra ett *uppgiftsdiagram*, där man beskriver uppgiften i mellan tre till sex steg. Exempel på frågor man kan ställa är: "Tänk på hur du gör när du utför (aktuell uppgift)". "Kan du bryta ner uppgiften i tre till sex steg?" och "Vilka av de här stegen som du just identifierade är kognitivt krävande?". Därefter väljer man ut en eller flera uppgifter som intressanta för vidare analys, till exempel, för att de är kognitivt krävande. Man gör då en *kunskapsgranskning* av de uppgifterna. Nu upprättar man en tabell över intressanta frågor. Man har stöd av rekommenderade "primära frågor" och "tilläggsfrågor". I tabellen fyller man sedan i resultatet från en intervju med experten, under rubrikerna "Ledtrådar" och "Varför svårt". Nästa steg är att göra en *simuleringsintervju*. Simuleringsintervjun sätter expertens problemlösning i ett sammanhang och ger detaljerad information om expertens kognitiva processer. Även här har man stöd av metoden, genom en rekommenderad tabell, med rubrikerna "Händelser", "Handlingar", "Situationsbedömning", "Avgörande ledtrådar" och "Möjliga misstag". Till sist kan man sammanställa data från flera intervjuer i en *tabell över kognitiva krav*. Då får man också data organiserade så att det kan vara lättare att se mönster i datamängden.

ACTA	
Krav	Beskrivning
När i utvecklingsprocessen	ACTA kan med fördel användas tidigt i utvecklingsprocessen.
Resurser	Eftersom metoden begränsar sig till samtal och intervjuer med experter är metoden ganska billig.
Förmåga	ACTA är ganska lätt att använda. Man får stöd i metodiken av färdiga förslag till frågor, vilket underlättar. Man behöver inte vara mycket insatt i problemområdet för att använda metoden, även om resultatet naturligtvis blir beroende av det (Lützhöft, 1999; Susi, 1999).
Med/utan användare	ACTA utförs ihop med experter inom det studerade området.
Teoretisk bakgrund	Uppgiftsanalys har tillämpats länge. Användandet av kognitiv uppgiftsanalys har ökat i takt med att många arbetsuppgifter har kognitiva krav som en viktig del av kritiska skeden i arbetet. Teorier om naturalistiskt beslutsfattande utgör en viktig bakgrund för metoden (t.ex., Klein m.fl., 1993).
Mognadsgrad	Metoden har inte funnits så länge, dock har flera olika domäner provats.
Reliabilitet	Resultatet beror till en del på vilka man väljer ut som experter, hur många och hur man väger ihop deras synpunkter.
Validitet	En nackdel är att det är svårt att verifiera resultatet från analysen. En annan nackdel är att man bara pratar med experter. Det kan ju faktiskt vara så att även nybörjare har något att bidra med. Till exempel, delar av utbildningen kanske inte var densamma för experterna som för nybörjarna.
Sensitivitet/Känslighet	Metoden har inte hög känslighet. Ofta får man dock reda på generella mönster.
Diagnosticitet	Ganska bra. Ibland kan det vara svårt att analysera de bakomliggande orsakerna.
Typ av fel som hittas	En fördel är att man inte i förväg måste veta vilka kritiska situationer man vill studera, utan det "växer fram" efterhand.
Resultatanalys	För metoden typiska tabeller används att få en tydlig struktur i data-mängden.
Andra metoder som ger motsvarande information	Andra kognitiva uppgiftsanalysmetoder kan ge motsvarande information. Även uppgiftsanalysmetoder, som ej primärt fokuserar på kognitiva aspekter, kan ge liknande information. Kännetecknande för ACTA är att endast experter används.
Kompletterande metoder	Metoden bör kompletteras med intryck från icke-expert, eller från andra typer av experter än användare.
Tillämpbarhet	Tillgång på experter (användare) är ett krav. Hög tillämpbarhet över domän och tid.
Störande/Påträngande	Ej särskilt störande/påträngande.
Användares acceptans	Hög.

Referenser	<ul style="list-style-type: none"> • Hutton, R. J. B., & Militello, L. G. (1997). Applied cognitive task analysis (ACTA): A practitioner's window into skilled decision making. In: D. Harris (Ed.), <i>Engineering Psychology and Cognitive Ergonomics (Vol 2): Job Design and Product Design</i>. Aldershot, UK: Ashgate. • Klein, G., Orasanu, J., Calderwood, R., & Zsombok, C. E. (1993). <i>Decision Making in Action: Models and Methods</i>. Norwood, NJ: Ablex • Klein, G. & Militello, L. G. (1997). <i>Cognitive Task Analysis for Display Design</i>. Workshop 15-16 September 1997, Swedish Centre for Human Factors in Aviation. Division of Industrial Ergonomics, Linköping Institute of Technology, Linköping. • Militello, L. G. & Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. <i>Ergonomics</i>, (Vol 41, No. 11, pp.1618-1641). • Lützhöft, M. (1999). <i>ACTA - Tillämpad kognitiv uppgiftsanalys. II: Utvärdering av tabell över kognitiva krav (CDT)</i>. HFA Degree Project 1999-02 • Susi, T. (1999). <i>ACTA - Tillämpad kognitiv uppgiftsanalys. I: Utvärdering av dess tillämpning i praktiken</i>. HFA Degree Project 1999-03
------------	--

5.26 Heuristisk utvärdering

Heuristisk utvärdering är en analytisk expertmetod, som genomförs av användbarhetsexperter, designers, eller motsvarande, för att systematisk granska/inspektera ett gränssnitts design med avseende på dess användbarhet (Nielsen, 1993). Målsättningen är att tidigt hitta användbarhetsproblem hos ett gränssnitt. Vid heuristisk utvärdering undersöker utvärderarna gränssnittet och bedömer dess överensstämmelse med kända användbarhetsprinciper (Nielsen, 1993). Vanligen används Nielsens principer (se ytterligare beskrivning, sid. 100).

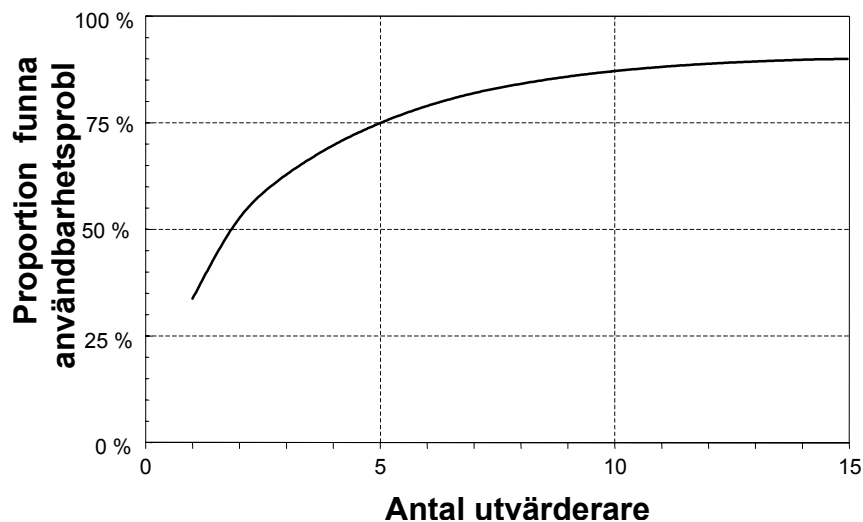
Heuristisk utvärdering	
Krav	Beskrivning
När i utvecklingsprocessen	Vanligen relativt tidigt. Enligt Faulkner (2000), synnerligen lämpligt under den tidiga utvärderingsprocessen för att allvarigare fel skall kunna elimineras innan riktiga användarförsök genomförs. Kan även användas parallellt med användartestning, eftersom metoderna vanligen inte hittar samma typ av fel.
Resurser	Billig metod. I synnerhet jämfört med användarstudier som vanligen är mycket kostsamma. Kräver flera utvärderare. Mest kostnadseffektivt är 5 utvärderare (Nielsen, 1993).
Förmåga	Utvärderarna bör helst vara användbarhetsexperter eller motsvarande. Utnyttjandet av en lista med principer för design, underlättar dock genomförandet, varför även andra typer av experter, exempelvis gränssnittsdesigners och konstruktörer vid behov kan delta i utvärderingen. Kräver kunskap om, eller vana vid, systemet som skall utvärderas. Vanligen krävs även domänkunskap.
Med/utan användare	Utan användare.
Teoretisk bakgrund	Vid Heuristisk utvärdering testar MSI-experterna systemets överensstämmelse med uppställda användbarhetsprinciper. Ofta används Nielsens principer (se ytterligare beskrivning, sid. 100). Detta görs vanligen tidigt, för att eliminera uppenbara fel, innan prototypen genomgått testning med användare/försökspersoner. I princip kan heuristisk utvärdering utföras av en enda utvärderare. Enskilda utvärderare tenderar dock att missa flertalet användbarhetsproblem, varför bäst resultat fås genom summering av resultatet från flera utvärderare (Nielsen, 1993). Ju fler utvärderare som används, desto fler fel hittas. Fem utvärderare brukar dock uppskattningsvis hitta 75 % av användbarhetsproblemen, och anses därför vanligen som det mest kostnadseffektiva antalet (Preece m.fl., 1994). Mindre än tre utvärderare bör dock inte användas (Nielsen, 1993). Med fler än fem utvärderare avtar nyttan av varje ytterligare utvärderare (se figur 5, sid. 100). Antal utvärderare är dock även beroende på förhållandet mellan kostnad och nytta, varför fler utvärderare kan användas om användbarhet är mycket kritiskt (Nielsen, 1993).
Mognadsgrad	Relativt ny metod.
Reliabilitet	Relativt god, men beroende på utvärderarens kompetens.
Validitet	Både testsituationen och MSI-experten som utvärderare skiljer sig från systemets naturliga arbetsmiljö. Låg ekologisk validitet.
Sensitivitet/Känslighet	God känslighet för grova uppenbara fel. Kan vara svårt att upptäcka sekvenser av fel och vilka konsekvenser ett fel kan få.
Diagnosticitet	Huvudsakligen kvalitativ analys. Statistiska analyser är vanligen deskriptiva.
Typ av fel som hittas	Identifierar många problem, med fem utvärderare hittas vanligen 75 % av användbarhetsproblemen (Nielsen, 1993).
Resultatanalys	Huvudsakligen kvalitativ. Kvantifiering kan göras av frekvens av olika typer av problem.
Andra metoder som ger motsvarande information	Andra analytiska expertmetoder, exempelvis kognitiv walkthrough och pluralistisk walkthrough.
Kompletterande metoder	Användartestning, pluralistisk walkthrough.
Tillämpbarhet	Kan tillämpas på de flesta system/prototyper.
Störande/Påträngande	–

Användares acceptans	–
Referenser	<ul style="list-style-type: none">• Faulkner, X. (2000). <i>Usability engineering</i>. Basingstoke: Macmillan Press Ltd.• Nielsen, J. (1993). <i>Usability engineering</i>. Boston: Academic Press, Inc.• Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). <i>Human-Computer Interaction</i>. Wokingham: Addison-Wesley.

Ytterligare beskrivning av Heuristisk utvärdering

Nielsens principer (Nielsen, 1993):

- *Enkel och naturlig dialog* – All överflödiga information konkurrerar med relevant information. All information skall ges i naturlig och logisk ordning.
- *Tala användarens språk* – Dialogen skall vara tydligt formulerad i ord och uttryck som är bekanta för användaren, och inte i systemorienterade termer.
- *Minimera användarens minnesbelastning* – Användaren skall inte behöva komma ihåg information mellan olika delar av dialogen. Hjälper för användning av systemet skall vara synlig, eller lättåtkomlig när den behövs.
- *Överensstämmelse* – Användaren skall inte behöva vara osäker på om olika ord, situationer eller handlingar innebär samma sak.
- *Återkoppling* – Systemet skall hela tiden hålla användaren informerad om vad som händer, genom tillräcklig återkoppling inom rimlig tid.
- *Tydligt markerade utgångar* – Användare väljer ofta funktioner av misstag och behöver därför tydligt markerade ”nödutgångar” för att enkelt avsluta funktioner.
- *Genvägar* – Ökar hastigheten för expertanvändare. Gör att systemet passar för både nybörjare och experter.
- *Bra felmeddelanden* – De skall uttryckas med lättbegripligt språk som noggrant anger felet och konstruktivt föreslår en lösning.
- *Förebygga fel* – Ännu bättre än bra felmeddelanden är en omsorgsfull design som förhindrar att problem uppstår.
- *Hjälper och dokumentation* – Bäst är om systemet kan användas utan dokumentation. Om dokumentation ändå behövs skall information vara lätt att hitta, vara fokuserad på användarens uppgift, lista konkreta steg, och inte vara för omfattande.



Figur 5. Proportion användbarhetsproblem som hittas med heuristisk utvärdering som funktion av antal utvärderare (anpassad efter Nielsen, 1993).

6 Referenser

- American Association of Psychology (1953). *Ethical standards of psychologists*. Washington DC: Author.
- American Association of Psychology. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist*, 57(12), 1060-1073.
- American Association of Psychology (2002). *Ethical principles of psychologists and code of conduct*. URL: <http://www.apa.org/ethics/code2002.html> [2003, 02-21].
- Andersson, J., Berggren, P., Castor, M., Magnusson, S., & Svensson, E. (2002). *Instrumentutveckling för mätning av: Gruppens prestationspotential* (FOI-R--0429--SE). Linköping: FOI (Totalförsvarets Forskningsinstitut).
- Angelborg-Thanderz, M. (1990). *Prisvärd militär flygning med rimliga risker* (FOA rapport C 500083 - 5.1). Stockholm.
- Berggren, P. (2000). *Situational awareness, mental workload, and pilot performance - relationships and conceptual aspects* (FOA-R--00-01438-706--SE). Linköping.
- Brookings, J. B. & Wilson, G. F. (1994). Physiological and workload changes during a simulated air traffic control task. In *Proceedings of the Human Factors and Ergonomics Society, 38th Annual Meeting*.
- Bullinger, H. J., Bauer, W., & Braun, M. (1997). Virtual environments. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1725-1759). New York: Wiley.
- Card, S. K., Moran, T. P., & Newel, A. (1980). The keystroke level model for user performance time with interactive systems. *Communications of the ACM*(23), 396-410.
- Card, S. K., Moran, T. P., & Newel, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Casali, J. & Wierwille, W. (1983). A comparison of rating scale, secondary task, physiological, and primary task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors*, 25, 623-642.
- Castor, M., Hanson, E., Svensson, E., Nählinder, S., LeBlaye, P., MacLeod, I., Wright, N., Alfredson, J., Ågren, L., Berggren, P., Juppet, V., Hilburn, B., & Ohlsson, K. (2003). *GARTEUR Handbook of mental workload measurement* (GARTEUR FM AG13 Final Report - GARTEUR TP 145).
- Charlton, S. G. (1996). Questionnaire techniques for test and evaluation. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of human factors testing and evaluation* (pp. 81-99). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Charlton, S. G., & O'Brien, T. G. (1996). The role of Human Factors testing and evaluation in systems development. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of human factors testing and evaluation* (pp. 13-26). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Cooper, G. E. & Harper, R. P. Jr. (1969). *The use of Pilot Rating in the Evaluation of Aircraft Handling Qualities*. NASA TN D-5153.

Derefeldt, G., Skinnars, Ö., Alfredson, J., Eriksson, L., Andersson, P., Westlund, J., Berggrund, U., Holmberg, J., & Santesson, R. (1999). Improvement of tactical situation awareness with colour-coded horizontal-situation displays in combat aircraft. *Displays*(20), 171-184.

Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT). In *Proceedings of the IEEE National Aerospace and Electronics Conference*, (pp. 789-795). Dayton, OH.

Endsley, M. R. (1990). Predictive Utility of an Objective Measure of Situation Awareness. In *Proceedings of the Human Factors Society - 34th Annual Meeting*, (pp. 41-45). Santa Monica, CA: Human Factors Society.

Endsley, M. R. (1995). Measurement of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1), pp 65-84.

Faulkner, X. (2000). *Usability engineering*. Basingstoke: Macmillan Press Ltd.

Geddie, J. C., Boer, L. C., Edwards, R. J., Enderwick, T. P., Graff, N., Pfendler, C., Ruisseau, J.-Y., & Van Loon, P. A. (2001). *NATO Guidelines on human engineering testing and evaluation* (RTO-TR-021). Neuilly-Sur-Seine Cedex, France: Nort Atlantic Treaty Organization.

Goodwin, C. J. (1995). *Research in psychology: Methods and design*. New York: Wiley.

Hackos, J. T., & Redish, J. C. (1998). *User and task analysis for interface design*. New York: Wiley.

Hankins, T. C. and Wilson, G. F. (1998), A Comparison of Heart Rate, Eye Activity, EEG and Subjective Measures of Pilot Mental Workload During Flight. *Aviation Space and Environmental Medicine*, vol. 69, 360-367.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: Elsevier Science Publishers Company.

Hennessy, R. T. (1990). Practical human performance testing and evaluation. In H. R. Booyer (Ed.), *MANPRINT an approach to systems integration* (pp. 433-470). New York: Van Nostrand Reinhold.

ISO 13407 (1999). *Human-centred design processes for interactive systems*. Geneva: ISO.

Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. M. (1991). User interface evaluation in the real world: A comparison of four techniques. In S. P. Roberson & G. M. Olson & J. S. Olson (Eds.), *Human Factors in computing systems CHI'91 Conference Proceedings* (pp. 119-124). New York: ACM Press.

- Hutton, R. J. B., & Militello, L. G. (1997). Applied cognitive task analysis (ACTA): A practitioner's window into skilled decision making. In: D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics (Vol 2): Job Design and Product Design*. Aldershot, UK: Ashgate.
- Jorna, P. G. A. M. (1993), Heart rate and workload variations in actual and simulated flight, *Ergonomics*, vol. 36, 1043-1054.
- Kanis, H. (1999). Design centered research into user activities. In W. S. Green & P. W. Jordan (Eds.), *Human Factors in product design: Current practice and future trends* (pp. 36-46). New York: Taylor & Frances.
- Kieras, D. (1997). A guide to GOMS Model usability evaluation using NGOMSL. In M. Helander & T. K. Landauer & P. Prabhu (Eds.), *Handbook of human-computer interaction* (Second, completely revised edition ed., pp. 733-766). Amsterdam: Elsevier Science B.V.
- Kirwan, B. & Ainsworth, L., K. (1992). *A guide to task analysis*, (pp.95-104). London. Taylor & Francis.
- Klein, G. & Militello, L. G. (1997). *Cognitive Task Analysis for Display Design*. Workshop 15-16 September 1997, Swedish Centre for Human Factors in Aviation. Division of Industrial Ergonomics, Linköping Institute of Technology, Linköping.
- Klein, G., Orasanu, J., Calderwood, R., & Zsombok, C. E. (1993). *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex.
- Kramer, A. F. (1991). Physiological metrics of mental workload: a review of recent progress. I D. L. Damos [Ed.], *Multiple Task Performance*. London: Taylor & Francis.
- Krebs, M. J., Wingert, J. W., & Cunningham, T. (1977). *Exploration of an Oculometer-Based Model of Pilot Workload*. NASA technical report CR-145153. Minneapolis, Minnesota: Honeywell Systems & Research Center.
- Lewis, C., & Wharton, C. (1997). Cognitive walkthroughs. In M. Helander & T. K. Landauer & P. Prabhu (Eds.), *Handbook of human-computer interaction* (Second, completely revised edition ed., pp. 717-732). Amsterdam: Elsevier Science B.V.
- Liu, Y. (1997). Software - user interface design. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1689-1724). New York: Wiley.
- Linell, P. (1994) *Transkription av tal och samtal: Teori och praktik*. Linköping: Institutionen för TEMA Kommunikation, Linköpings Universitet.
- Luczak, H. (1997). Task analysis. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 340-416). New York: Wiley.
- Lützhöft, M. (1999). *ACTA - Tillämpad kognitiv uppgiftsanalys. II: Utvärdering av tabell över kognitiva krav (CDT)*. HFA Degree Project 1999-02
- Magnusson, S. (2002). Similarities and Differences in Psychophysiological Reactions Between Simulated and Real Air-to-Ground Missions. *The International Journal of Aviation Psychology*. Vol. 12(1), 49-61.

MANPRINT in acquisition: A handbook. (2000). Prepared by office of the deputy chief of staff for personnel. Personnel Technologies Directorate 300 Army Pentagon, Washington DC. [URL: http://www.manprint.army.mil/manprint/guides/handbookacquis/handbookacq.htm](http://www.manprint.army.mil/manprint/guides/handbookacquis/handbookacq.htm) (2003-02-04).

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., & Brannan, J. R., (1990). Eye Movements indices of Mental Workload, *Acta Psychologica*, 75, 75-89.

McLeod, R. W., & Sherwood-Jones, B. M. (1992). Simulation to predict operator workload in a command system. In B. Kirwan & L. K. Ainsworth (Eds.), *A guide to task analysis* (pp. 301-310). London: Taylor & Frances.

Meister, D. (1996). Human Factors test and evaluation in the twenty-first century. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of human factors testing and evaluation* (pp. 313-322). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Meister, D., & O'Brien, T. G. (1996). The history of Human Factors testing and evaluation. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of human factors testing and evaluation* (pp. 3-11). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Militello, L. G. & Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, (Vol 41, No 11, pp. 1618-1641).

Miller, R. B. (1953) *A Method for Man-machine Task Analysis*. Wright-Patterson Air Force Base, Ohio, USA: WADC Tech Rept. No 53, 137.

Modéer, B., & Oskarsson, P.-A. (2000). *Mätning av stabers lägesuppfattning* (FOA-R--00-01815-505-SE). Linköping: Försvarets Forskningsanstalt.

Modéer, B., Söderberg, H., & Elmquist, K. (2000). *Datainsamling med elektroniska enkäter* (FOA-R--00-01614-505-SE). Linköping: Försvarets Forskningsanstalt.

Modéer, B., Wikberg, P., Oskarsson, P.-A., & Lundin, J. (2001a). *Nya metoder för datainsamling och värdering av stabsarbete (PM-ESÖ2)*. FOI Memo 01-1763/L).

Modéer, B., Wikberg, P., Oskarsson, P.-A., & Lundin, J. (2001b). *Mätning av samverkan vid experimentell simuleringsövning (PM-ESÖ3)*. FOI Memo 01-2716/L).

Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press, Inc.

Nielsen, J. (1997). Usability testing. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1543-1568). New York: Wiley.

Nilsson, L., & Nåbo, A. (1996). *Evaluation of application 3: Intelligent cruise control simulator experiment*. Linköping: Statens Väg- och transportforskningsinstitut. VTI särtryck 266.

O'Brien, T. G. (1996). Test safety. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of human factors testing and evaluation* (pp. 323-331). Mahwah, New Jersey: Lawrence Erlbaum Associates.

- Oskarsson, P.-A. (2002). *Översikt över metodik för MSI-utvärdering vid systemutveckling* (Metodrapport FOI-R--0583--SE). Linköping: Totalförsvarets Forskningsinstitut.
- Popovic, V. (1999). Product evaluation methods and their importance in designing artifacts. In W. S. Green & P. W. Jordan (Eds.), *Human Factors in product design: Current practice and future trends* (pp. 26-35). New York: Taylor & Frances.
- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-Computer Interaction*. Wokingham: Addison-Wesley.
- Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design: Beyond human-computer interaction*. New York: Wiley.
- Price, H. E. (1990). Conceptual system design and the human role. In H. R. Booher (Ed.), *MANPRINT an approach to systems integration* (pp. 161-203). New York: Van Nostrand Reinhold.
- Roscoe, A. H. (1993). Heart rate as a psychophysiological measure for in-flight workload assessment. *Ergonomics*, vol. 36, 1055-1062.
- Salvendy, G., & Carayon, P. (1997). Data collection and evaluation of outcome measures. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1451-1470). New York: Wiley.
- Sanders, M. S., & McCormick, E. J. (1993). *Human factors in engineering and design* (7th ed.). New York: McGraw-Hill.
- Seamster, T. L., Redding, R. E, & Kaempf, G. L. (1997). *Applied cognitive task analysis in aviation*. Aldershot: Ashgate.
- Stein, E. S. (1992). *Air Traffic Control Visual Scanning*. FAA Technical Report DOT/FAA/CT-TN92/16. US Department of Transportation.
- Stern, J. A. (1993). The Eyes: Reflector of attentional processes (synopsis by J. J. Kelly). In *CSERIAC Gateway*, 4 (4), 7-12.
- Suh, N. P. (1998). *Axiomatic Design Theory for Systems*. *Research in Engineering Design*, 10:189-209. London: Springer-Verlag.
- Susi, T. (1999). *ACTA - Tillämpad kognitiv uppgiftsanalys. I: Utvärdering av dess tillämpning i praktiken*. HFA Degree Project 1999-03
- Svensson, E. (2003). Statistical techniques for data reduction and modelling. In M. Castor & E. Hanson & E. Svensson & S. Nählinder & P. LeBlaye & I. MacLeod & N. Wright & J. Alfredson & L. Ågren & P. Berggren & V. Juppet & B. Hilburn & K. Ohlsson (Eds.), *GARTEUR Handbook of mental workload measurement* (pp. 122-132): GARTEUR FM AG13 Final Report - GARTEUR TP 145.
- Svensson, E., Angelborg-Thanderz, M., & van Avermaete, J. (1997a). *Dynamic measures of pilot mental workload, pilot performance, and situational awareness*. Technical Report: VINTHEC-WP3-TR01. NLR, Amsterdam.

Svensson, E., Angelborg-Thanderz, M., Sjöberg, L., & Olsson, S. (1997b). Information complexity - Mental workload and performance in combat aircraft. *Ergonomics*, 40(3), 362-380.

Svensson, E., Angelborg-Thanderz, M., & Wilson, G. F., (1999). *Models of pilot performance for systems and mission evaluation, psychological and psychophysical aspects*. AFRL-HE-WP-Tr-1999-0215.

Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a toll for aircrew systems design. In *Situational awareness in aerospace operations* (AGARD-CP-478; pp. 3/1-3/17). Neuilly-Sur-Seine, France: NATO-Advisory Group for Aerospace Research and Development.

Taylor, R. M. (1995a). Experimental measures: Performance-based self ratings of situational awareness. In *International Conference on Experimental Analysis and Measurement of Situation Awareness*, Daytona Beach, 1-3 November, 1995.

Taylor, R. M. (1995b). *CC-SART: The Development of an experimental measure of cognitive compatibility in systems design*, Report to TTCP UTP-7 Human Factors in Aircraft Environments, Annual Meeting, DCIEM, Toronto, Canada, 12-16 June, 1995.

Tsang, P., & Wilson, G. F. (1997). Mental workload. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 417-449). New York: Wiley.

Wierwille, W. W. & Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. In *Proceedings of the 27th Annual Meeting of the Human Factors Society* (Santa Monica, CA: Human Factors Society), pp. 129-133.

Wilson, G., Balkin, T., Beaumont, M., Burov, A., Edgar, G., John, Fraser, W., Gilliland, K., Grandt, M., Gundel, A., Hockey, B., Schlegel, R. E., Stone, H., Svensson, E., Van Orden, K., Varoneckas, G., & Veltman, H. (2003). *Operator functional state assessment* (NATO RTO HFM-056/TG-008).

Appendix. Ordlista och Begreppsförklaringar

ACTA (Applied Cognitive Task Analysis) Metod för kognitiv uppgiftsanalys. Består av uppgiftsdiagram, kunskapsgranskning och simuleringsintervjuer.

Beroendevariabel Den eller de variabler som man mäter vid ett test. Exempelvis tid för att utföra en viss uppgift.

BFRS (BedFord Rating Scale) Subjektiv metod för att mäta mental arbetsbelastning. Har utvecklats från Cooper-Harper skalor.

CC-SART (Cognitive Compatibility-SART [Situation Awareness Rating Technique]) Subjektiv metod för att mäta situationsmedvetande. Är en vidareutveckling av SART där försökspersonerna även får skatta kognitiv kompatibilitet.

Cooper-Harper skalor Subjektiv metod för att mäta mental arbetsbelastning hos piloter.

Cooper-Harper skalor specialanpassade för situationsmedvetande Subjektiv metod för att mäta situationsmedvetande. Metoden har utvecklats av FOI och består av en modifierad version av Cooper-Harper skalor.

Delad uppmärksamhet Förmågan att samtidigt uppmärksamma två eller flera informationskällor.

Ekologisk validitet Huruvida själva genomförandet av utvärderingen (mätmetoden etc.) påverkar användarnas/försökspersonernas beteende och därmed resultatet för utvärderingen.

EOG (Elektro-okulografi) Mätning av ögonrörelser via elektroder placerade runt ögonen. EOG-måttet förändras med de visuella kraven för en uppgift.

EPOG (Eye-Point-of-Gaze) EPOG innebär blickpunktsregistrering. Mätning kan göras av vart blicken riktas, t.ex. på vilken display en operatör tittar vid en viss tidpunkt.

Extern validitet Att resultaten från en undersökning kan generaliseras till andra liknande situationer.

Fixering Låsning av ögonens riktning till en detalj i omgivningen. Under fixering kan informationsinhämtning ske.

Gränssnitt eller Användargränssnitt. Inom MSI gränssnittet mellan systemet och användaren: den del av systemet som användaren arbetar med, dvs. skärmbilder, mätare, knappar, reglage etc.

Hawthorne effekten Att observatörens närvaro, eller studiens genomförande, medför att beteendet hos de människor som observeras omedvetet avviker från deras normala beteende.

Hierarkisk uppgiftsanalys Uppgiftsanalys där uppgifter delas upp i deluppgifter, vilka i sin tur genomgår en egen uppgiftsanalys, osv. tills tillräcklig detaljeringsnivå uppnåtts.

Heuristisk utvärdering Expertmetod där systematisk genomgång av ett system med avseende på användbarhetsproblem genomförs av användbarhetsexperter.

Intern validitet Att en mätmetod mäter vad den avser att mäta.

Iterativ design En cyklisk designprocess där exempelvis en prototyp designas, utvärderas, designas om, utvärderas igen, osv. tills uppställda krav uppnås.

Kognition Inom psykologi, samlingsbegrepp för: varseblivning, minne, begreppsbildning, resonerande, problemlösning och uppmärksamhet

Kognitiv uppgiftsanalys Uppgiftsanalyser som fokuserar på användarnas kognitiva processer. Bygger på att människa system interaktion vanligen består av komplexa kognitiva processer, och därför inte kan studeras med traditionella beteendeorienterade uppgiftsanalyser.

Kognitiv walkthrough Walkthrough som fokuserar på kognitiva processer.

Kvalitativa metoder Utvärderingsmetoder där forskaren själv befinner sig i den miljö som analyseras. Försöker att uppfånga både handlingar och innebörden av handlingarna och därmed ge en helhetsbeskrivning. Ger ofta data som kan kategoriseras, men inte översättas till numeriska mått.

Kvantitativa metoder Utvärderingsmetoder som ger data bestående av numeriska värden.

Likertskala Skattningsskala där varje alternativ är ett påstående.

Lingvistisk Språkvetenskaplig, som avser språket.

Mentala modeller Förmågan att organisera minnen som behövs för att förstå specifika situationer. Används för att organisera erfarenheter av sig själv, andra människor, miljön och av de tekniska system man interagerar med.

Mentala processer Varseblivande, tänkande, värderingar, emotioner etc.

Modifierade Cooper-Harper skalor Subjektiv metod för att mäta mental arbetsbelastning. Vidareutveckling av Cooper-Harper skalor som ej begränsas till piloter, utan mäter även generell mental arbetsbelastning.

NASA-TLX (Task Load Index) Subjektiv metod för att mäta mental arbetsbelastning. Bygger på deltagarnas skattning av mentala krav, fysiska krav, tidspress, prestation, ansträngning och frustration.

Nielsens principer Tio principer för användbarhet, används ofta vid heuristisk utvärdering.

Oberoende variabel Den variabel som man vill få veta något om när ett försök genomförs. Exempelvis typ av system, hur olika versioner av design påverkar prestationen.

Pappersprototyper Prototyper skapade med vanligt ritpapper, används vanligen för att modellera olika skärmbilder.

Psykofysiologiska metoder Metoder där mätningar av fysiologiska reaktioner i kroppen görs. Exempelvis: hjärtfrekvens, variation i hjärtfrekvens, pupillstorlek, ögonrörelser och ögonblinkningar. När resultaten analyseras görs antagandet att uppmätta fysiologiska reaktioner beror på psykologiska processer, t.ex. mental arbetsbelastning eller brist på situationsmedvetande.

Reliabilitet Huruvida en mätmetod är stabil. Att den ger samma resultat vid upprepade mättillfällen.

Retrospektiva metoder Utvärderingsmetoder där en användarstudie spelas in på video, och inspelningen efter genomförd utvärdering går igenom av försöksledaren och användaren tillsammans.

Sackad Förflyttning av blicken från en fixering till nästa.

SAGAT (Situation Awareness Global Assessment Technique) Subjektiv metod för att mäta situationsmedvetande, där försökspersoner får besvara frågor när en simulering fryses.

SART (Situation Awareness Rating Technique) Subjektiv metod för att mäta situationsmedvetande, där försökspersoner bl.a. skattar hur mycket uppmärksamhet de har och behöver att ha.

Situationsmedvetande Förmågan att uppfatta vad som händer i omgivningen, förstå vad som händer, och kunna förutsäga vad som kommer att hända.

Skattningsskala Metod för att besvara enkätfråga, består av flera numrerade alternativ, där ett lägre värde exempelvis innebär en mer positiv bedömning, eller tvärt om.

Störvariabel Okontrollerad yttre variabel som kan påverka resultatet vid ett försök. Exempelvis störande ljud som kan påverka prestationen vid ett försök, eller olika förutsättningar för grupper av användare som jämförs.

SWAT (Subjective Workload Assessment Technique) Subjektiv metod för att mäta mental arbetsbelastning. Bygger på varje deltagares skattade nivå av temporal belastning, mental ansträngning och stress.

Tänka högt Utvärderingsmetod där användarna under arbetet genom att "tänka högt" talar om vad de gör, varför de gör det, och vad de tänker på när de gör det.

Uppgiftsanalys Generell term för analyser som görs av de uppgifter som användare utför med ett system. Syftet är att förutsäga problem, utvärdera system med avseende på användbarhet eller funktionella krav, förutsäga prestation, mäta systems komplexitet och lärlarhet.

Validitet Att en utvärderingsmetod mäter vad det som den har designats att mäta.

Walkthroughs Strukturerade metoder där användbarhetsexperter själva går igenom system/prototyper med syftet att hitta användbarhetsproblem tidigt i utvecklingscykeln.

Wizard of Oz Simuleringsmetod där användaren arbetar med vad han/hon tror är ett datoriserat system. I själva verket är det en människa som från en annan plats/terminal simulerar systemet, eller delar av det, genom att mata in relevanta svar till användaren.