



Strål
säkerhets
myndigheten

Swedish Radiation Safety Authority

Författare:

Per-Anders Oskarsson
Björn JE Johansson
Natalia Gonzalez

Forskning

2010:02

Metodutveckling för integrerad validering

Titel: Metodutveckling för integrerad validering
Rapportnummer: 2010:02
Författare: Per-Anders Oskarsson, Björn JE Johansson och Natalia Gonzalez
FOI, Totalförsvarets Forskningsinstitut, Avdelningen för Informationssystem,
Enheten för Systemvärdering, Linköping
Datum: Februari 2010

Denna rapport har tagits fram på uppdrag av Strålsäkerhetsmyndigheten, SSM. De slutsatser och synpunkter som presenteras i rapporten är författarens/författarnas och överensstämmer inte nödvändigtvis med SSM:s.

SSM perspektiv

SSM granskar tillståndshavarnas hantering av anläggningsändringar och deras hantering av tester. Myndigheten lägger stor vikt vid att dessa tester genomförs stringent och korrekt på alla sätt samt att de rapporteras på ett sätt som gör det möjligt att göra bedömningar. Men även om ISV genomförs optimalt, på bästa sätt, finns starka skäl att utveckla ISV. Det finns också starka skäl att utveckla metoder för vad som ska testas, metoderna vid testning, etc.

Totalförsvarets Forskningsinstitut (FOI) fick i uppdrag att studera metodutveckling för integrerad validering.

Bakgrund

Kärnkraftindustrin genomför en hel del anläggningsändringar med påverkan på kontrollrummen. Syftet med förändringarna är förbättringar av olika slag, som funktioner och arbetsförutsättningar. Det finns noggranna planer, genomtänkt design och många olika kompetenser inblandade men trots detta så framträder oväntade aspekter efter genomförandet – inte enbart vad som primärt var avsikten. Det är därför väldigt viktigt att alla konsekvenser av en ändring analyseras i förväg. Det är också i högsta grad viktigt att resultatet av anläggningsändringarna testas. SSM:s krav är att det inte får uppstå en situation att förbättrad säkerhet i ett avseende leder till försämrad säkerhet i ett annat avseende på ett sådant sätt att säkerheten som helhet försämras.

Syfte

Syftet med studien var att bidra till fördjupad kunskap både för SSM:s tillsyn och för ökad kunskap inom kärnkraftindustrin.

Resultat

FOI har med sina resultat lyft upp frågeställningar och problematiserat ytterligare aspekter.

Effekter på SSM:s tillsyn

SSM har fått goda förutsättningar att fortsätta vara pådrivande för en vidare utveckling av området.

Projektinformation

Projekthandläggare på SSM: Yvonne Liljeholm Johansson
Diarienummer för projektet är SSM 2009/2142
Projektnummer: 1118

Innehållsförteckning

Sammanfattning	4
Summary	5
1. Bakgrund	6
Validering	6
Kontrollrumsvalidering	7
Rapportering	9
2. Integrerad validering	11
Stegvisa uppgraderingar.....	11
Design.....	11
”The envisioned world problem”	11
Behovet av iterativ design.....	12
Behovet av iterativ validering.....	12
Relevanta scenarier	13
Förutsättningar för validering.....	13
Oklara effekter av validering.....	13
Försökspersoner	13
Träningseffekter	14
Förändrad arbetssituation	15
Rollfördelning mellan operatörer.....	15
Acceptanskriterier	15
Validering med baslinjemätning	16
Validering utan baslinjemätning.....	16
Kontextuell värdering av systemets användbarhet	17
Internationella Guidelines	18
NUREG	18
IFE.....	19
ISO	19
Guidelines från andra domäner	20
GARTEUR	20
VINTHEC.....	20
MANPRINT	21
The Command and Control Research Program	21
Synpunkter på guidelines	21
Mätmetoder vid validering.....	22
Explorativa metoder	22
Uppgiftsanalys.....	22
Walkthroughs	24
Heuristisk utvärdering.....	24
Kvantitativa metoder	25
Kvalitativa metoder	28
Intervju.....	28
Enkätundersökning.....	28
Observation.....	30
Tänka högt.....	30
Mätning av indirekta begrepp	30
Situationsmedvetande	31

Mental arbetsbelastning	32
Dynamiska mätmetoder	33
Psykofysiologiska mätmetoder	34
Systemloggar	36
Kvasidynamiska mått	37
Holistisk ansats	38
Statistisk analys vid integrerad validering	39
Kovariansanalys.....	39
Metodik för att hitta orsakssamband.....	40
Korrelationsanalys.....	40
Datareduktion.....	41
Kausalanalys – LISREL.....	42
3. Fokusgrupp	44
Metod	44
Genomförande	46
Analys	46
Deltagare	47
Resultat.....	47
Referensdokument och begreppsdefinitioner.....	47
Diskussion.....	49
Samarbete mellan verken.....	49
Diskussion.....	50
Samarbete med externa parter.....	50
Diskussion.....	52
SSM	52
Diskussion.....	54
Metoder för att samla in data	54
Diskussion.....	57
Praktiska problem	57
Diskussion.....	59
Baslinjemätning.....	59
Diskussion.....	62
Att tolka valideringsrapporter	63
Diskussion.....	63
Designprocessen	63
Diskussion.....	66
Sammanfattning.....	66
4. Förslag och synpunkter	68
Synpunkter på gemensamma riktlinjer för valideringsprocessen.....	68
Förslag: Handböcker.....	68
Förslag: Workshops.....	69
Synpunkter på baslinjemätning	69
Förslag: Acceptanskriterier.....	70
Synpunkter på SSM:s roll i valideringsprocessen	70
Förslag: Avstämningpunkter	70
Synpunkter på externa aktörers roll i valideringsprocessen	71
Förslag: Gemensamma möten	71
Förslag: Sammansättning av valideringsteam	71

Synpunkter på tidpunkt och konsekvenser av integrerad validering..	71
Förslag: Tidpunkt för integrerad validering.....	72
Förslag: Åtgärder efter integrerad validering.....	72
Synpunkter på designprocessen	72
Förslag: Designrekommendationer	72
Synpunkter på mät- och beräkningsmetoder	73
Förslag: Alternativa mätmetoder.....	73
Förslag: Alternativa statistiska metoder.....	73
5. Sammanfattning och slutsats	74
6. Referenser	76

Sammanfattning

Denna rapport har skrivits på uppdrag av Strålsäkerhetsmyndigheten (SSM). Syftet var att stödja Strålsäkerhetsmyndigheten avseende metodutveckling inom Human Factors området rörande integrerad validering vid kontrollrum.

Referensram för arbetet har varit befintliga rekommendationer och anvisningar från kärnkraftsområdet, erfarenheter från granskning av Turbic Validering (Oskarsson, Nählinder & Svensson, 2008), samt erfarenheter från valideringsstudier inom försvarsmakten, exempelvis avseende ledningscentraler och flygförare. Dessa verksamheter kännetecknas av komplexa system i extrema miljöer, ofta med höga risker där mänskliga felhandlingar kan få allvarliga konsekvenser.

En fokusgrupp har genomförts med ansvariga representanter för Human Factors frågor vid samtliga svenska kärnkraftverk. Frågor som diskuterades var bland annat vem/vad en integrerad validering (IV) är till för, vad som bör ingå i en IV, jämförelsen med Baslinjemätning, Designprocessen, SSM:s roll, vilka mätmetoder som ska användas vid IV och hur de påverkas av förändringar i kontrollrummet.

Rapporten tar upp olika frågor rörande valideringsprocessen. Kompletterande mätmetoder vid integrerad validering diskuteras, bland annat dynamiska, psykofysiologiska och kvalitativa metoder för att identifiera problem. Kompletterande analysmetoder för statistisk bearbetning ges. Studien pekar på ett antal brister i valideringsprocessen, bland annat behovet av gemensamma riktlinjer för validering och design, kriterier för olika typer av mätningar, förtydligande av SSM:s roll och rekommendationer för externa aktörers ansvar i valideringsförfarandet. Författarna ger 12 förslag på åtgärder för att åtgärda de identifierade problemen.

Summary

This report has been written on the commission of the Swedish Radiation Safety Authority (SSM). The purpose was to support SSM in the development of Human Factors methods concerning integrated validation of nuclear power plant (NPP) control rooms.

The frame of reference for this work was existing recommendations and instructions from the NPP area, experiences from the review of the Turbic Validation (Oskarsson, Nählinder & Svensson, 2008), and experiences from system validations performed at the Swedish Armed Forces, e.g. concerning military control rooms and fighter pilots. These enterprises are characterized by complex systems in extreme environments, often with high risks, where human error can lead to serious consequences.

A focus group has been performed with representatives responsible for Human Factors issues from all Swedish NPP:s. The questions that were discussed were, among other things, for whom an integrated validation (IV) is performed and its purpose, what should be included in an IV, the comparison with baseline measures, the design process, the role of SSM, which methods of measurement should be used, and how the methods are affected of changes in the control room.

The report brings different questions to discussion concerning the validation process. Supplementary methods of measurement for integrated validation are discussed, e.g. dynamic, psychophysiological, and qualitative methods for identification of problems. Supplementary methods for statistical analysis are presented. The study points out a number of deficiencies in the validation process, e.g. the need of common guidelines for validation and design, criteria for different types of measurements, clarification of the role of SSM, and recommendations for the responsibility of external participants in the validation process. The authors propose 12 measures for taking care of the identified problems.

1. Bakgrund

Denna rapport har skrivits på uppdrag av Strålsäkerhetsmyndigheten (SSM). Syftet var att stödja SSM avseende metodutveckling inom Human Factors området avseende integrerad validering vid kontrollrum.

Att bedömningar av kontrollrum genomförs på ett bra sätt är av stor vikt, så att väsentliga aspekter inte förbises. Om exempelvis arbetssättet, och samspelet mellan operatörerna, i kontrollrummet förändras, bör de metoder som används vid bedömning av kontrollrummet också kunna fånga upp detta. I vissa fall kan det också vara nödvändigt att komplettera eller förändra valideringsmetodiken om förändringen är mycket omfattande. Det är t.ex. möjligt att omställning från analoga till digitala operatörsgränssytor kräver nya värderingsmetoder.

Referensram för arbetet har varit befintliga rekommendationer och anvisningar från kärnkraftsområdet, erfarenheter från granskning av Turbic Validering (Oskarsson, Nählinder & Svensson, 2008), samt erfarenheter från valideringsstudier inom försvarsmakten, exempelvis av ledningscentraler och stridsflygare. Dessa verksamheter kännetecknas av komplexa system i extrema miljöer, ofta med höga risker där mänskliga felhandlingar kan få allvarliga konsekvenser. I uppdraget till FOI ingick att resonera öppet kring integrerad validering utifrån våra egna premisser. Syftet med detta var att skapa nya uppslag och tankar kring hur valideringsprocessen skulle kunna genomföras.

För att även inhämta mer kunskap om dessa frågor från tillståndshavarnas synvinkel, har en fokusgrupp genomförts. Representanter med ansvar för Human Factors frågor vid samtliga svenska kärnkraftverk deltog.

Validering

Det finns ett flertal definitioner av validering och oftast sätts termen i kontrast till verifiering. Validering innebär en bekräftelse genom undersökning och framläggande av bevis att de särskilda kraven för en specifik, avsedd användning har uppfyllts (Svensk standard SS 020106, citerat i SWEDAC, 2005). Verifiering innebär en bekräftelse genom undersökning och framläggande av bevis att specificerade krav har uppfyllts (Svensk standard SS 020106 citerat i SWEDAC, 2005).

Internationella Standardiseringsorganisationen (ISO) beskriver validering enligt följande: *“Confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled.”* (Draft International Standard ISO/DIS 8402 citerad i Green och Collier, 1999).

Internationella energikommissionen (IEC) skriver: *“Validation, which should be carried out after completing the verification, is generally defined as the test and evaluation to determine that a problem solution complies with the functional, performance and interface requirements. More specifically, it is the process of determining whether the physical and organisational design for operations is*

adequate to support effective integrated performance of the functions of the control room operation staff.” (Green & Collier, 1999).

Nuclear Regulatory Commission (NRC) skriver: *”The objective of validation is to provide evidence that the integrated system supports plants personnel in safe operation of the plant; i.e.; that the integrated design remains within acceptable performance envelopes.” (O’Hara, Stubler, Higgins & Brown, 1997).*

Validering är dock inte alltid konkret definierat och det finns verksamheter där det istället arbetas utifrån vad de anser är kontentan av validering. I Eurocontrols valideringshandbok, Validation Guideline Handbook, förklaras validering som den process *”by which the fitness-for-purpose of a new system or operational concept being developed is established” (MAEVA, 2004).*

Alla ovannämnda exemplen beskriver eller definierar validering på olika detaljdjup och på olika sätt. ISO lägger tyngdpunkten på att validering ska bekräfta att ett system eller en process ska uppfylla det syftet den är avsedd för. IEC och NRC lägger istället tyngdpunkten på att systemet/processen skall vara ett stöd till operatörerna i deras arbete.

Kontrollrumsvalidering

I följande avsnitt presenteras en översikt över valideringsprocessen rörande kontrollrum. Syftet med detta är att ha en bas för diskussion senare i rapporten. Observera att processen som beskrivs bygger på en rapport från 1998 och således inte är koncensus eller standard inom svensk kärnkraft, snarare ska den ses som ett exempel på hur en validering skulle kunna gå till (Rollenhagen, Bladh, Borg & Evénus, 1998). Om inte annat anges är det Rollenhagen m.fl. (1998) som är referens i följande avsnitt. Som framgår av fokusgruppsdiskussionen som presenteras senare i rapporten är det långt ifrån säkert att en validering faktiskt går till enligt denna beskrivning.

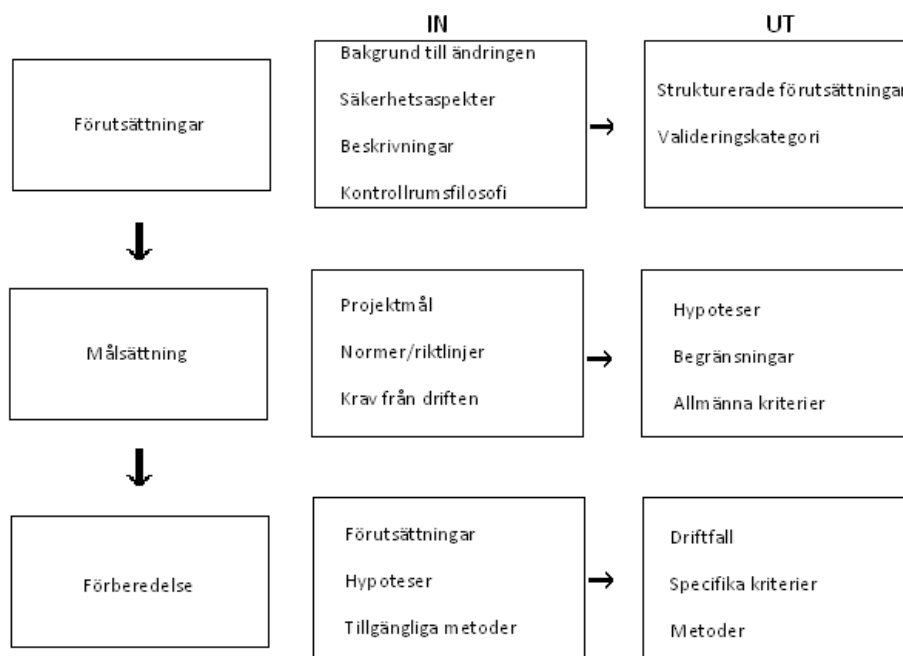
Det grundläggande syftet i validering är att granska huruvida systemet är användbart (funktionellt) för sina ändamål eller inte. Att granska detta i ett kontrollrum är långt ifrån enkelt då funktionaliteten i ett kontrollrum är beroende av samspelet hos en rad olika faktorer.

Validering av kontrollrumsändringar ska ske på ett strukturerat sätt, vilket bl.a. kräver en förbestämd plan. Valideringsplanen ska bl.a. innehålla specifikation av målsättningar, resurser, kriterier, mätinstrument och hantering av resultat. Redan i starten av ett projekt som påverkar kontrollrummet så bör valideringsaspekterna uppmärksammas. De beslut som fattas i utvecklingsfasen bör dokumenteras med avseende på vilka övervägande och/eller kompromisser som berör kontrollrumsutformningen. På så sätt underlättas valideringsarbetet genom att ett större fokus kan läggas på kritiska punkter. Valideringsprocessen bör bygga på en bred metoduppsättning, då kombinationen av exempelvis enkäter och expertbedömningar ökar möjligheten att dra valida slutsatser.

Ett kontrollrum består av flera olika specifika delfunktioner, som exempelvis människa-system gränssnitt, instruktioner, arbetssätt, personal med relevant utbildning, miljöfaktorer m.m. Valideringen av ett kontrollrum bör därför fokusera på individuella faktorer och på relationen och samfunktionen mellan dessa faktorer. För att åstadkomma detta krävs att verifieringen av enskilda komponenter når respektive krav men framförallt att samfunktionen mellan komponenterna testas i specifika situationer (driftfall). Det kan vara svårt att definiera vad som är en ”godkänd” samfunktion men genom att beakta samfunktionen i valideringen så betonas ett förhållningssätt som innebär att validering bör ha uppmärksamheten riktad mot eventuella problem i samfunktionen mellan exempelvis människa-maskin, människa-instruktion och människa-människa.

Valideringsunderlaget ska innehålla antagande om vad som förväntas uppfyllas av systemet och god praxis vid validering är att vara öppen för det oväntade samt att notera sådant som vid första anblick kan anses som ”småsaker”. Resultaten ska sedan försöka generalisera från försökssituationen till det verkliga systemet.

I *Handbok – validering av kontrollrumsändringar* föreslår Rollenhagen m.fl. (1998) en generell valideringsplan. Figur 1 nedan skildrar valideringsplanen som ett blockschema med ingångskriterier för blocken (IN) och med vilken typ av resultat som genereras (UT).



Figur 1. Valideringsplan (Rollenhagen m.fl., 1998)

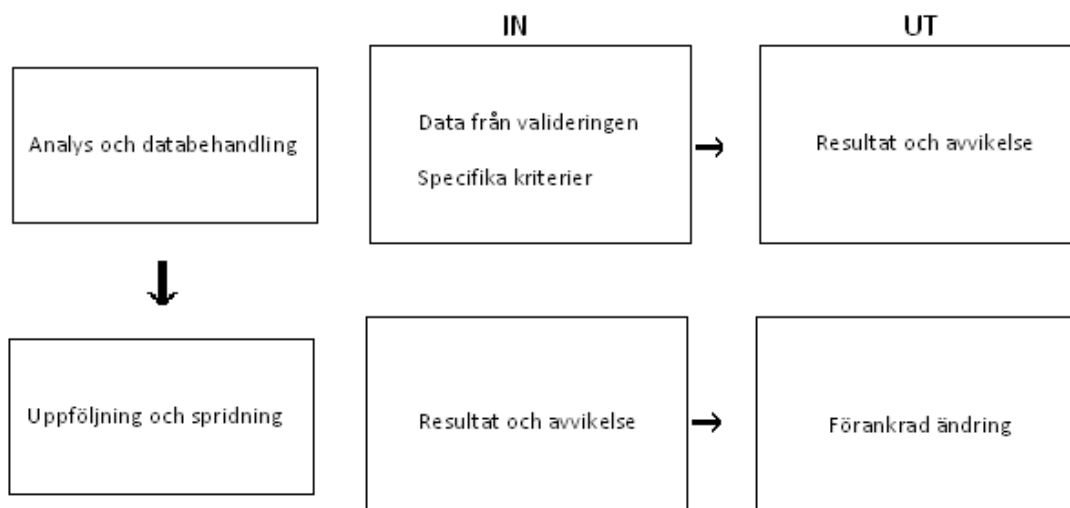
Under förutsättningar bör bl.a. en bakgrund till ändringen eller nykonstruktionen anges, likaså bör säkerhetsaspekter som reaktor- och/eller personsäkerhet klaras ut. Under målsättning anges först det övergripande målet med valideringen. Därefter ska detta brytas ner till delmål för att uppnå en mer detaljerad nivå. Hypoteserna som genereras är bl.a. till för att veta vad som ska testas i valideringen. Exempel på hypoteser kan vara: *operatörerna upplever att det nya systemet stödjer utförandet av arbetsuppgifterna*, eller: *det nya systemet fungerar tillsammans med övriga stödfunktioner*.

Begränsningarna som ställs upp är till för att fånga in valideringens omfattning och/eller begränsning. I detta ingår även att kategorisera de olika valideringarna för att olika fokus och vikt ska kunna läggas på valideringens olika delar. Under allmänna kriterier ska övergripande kriterier beskrivas. Flera av dessa bör vara generella för flera valideringar, exempelvis *god kontrollbarhet*. Majoriteten av de övergripande kriterierna ska baseras på kvalitativa bedömningar och dessa ska i sin tur vara grundade på en sammanvägning av strukturerade intervjuer, observationer samt expertbedömningar.

I förberedelserna ingår det att bl.a. ta fram en valideringsgrupp som speglar kompetenser från olika områden, exempelvis systemkunskap och MTO-kunskap. Det ska även anges vilka driftfall som valts att fungera som valideringssituationer, samt motiveringen till dessa.

Rapportering

Valideringsplanen ska göras innan insamling av data och efter genomförandet av valideringsplanen produceras valideringsrapporten. Valideringsrapporten behandlar analys, databearbetning, resultat och uppföljning (se Figur 2 nedan).



Figur 2. Valideringsrapporten (Rollenhagen m.fl., 1998)

En valideringsrapport ska kunna läsas fristående. Därför krävs att den inleds med en kort beskrivning av de scenarier som har genomförts. Resultatet av

valideringen ska först presenteras genom att ge den sammanvägda bedömningen av hela systemet/ändringen. En gränsdragning mellan godkänt, varning och icke godkänt kan exempelvis göras. Om varning eller icke godkänt anges ska även en förklaring till "betyget" finnas med. Resultaten ska även kopplas till de hypoteser som ställs upp i valideringsplanen.

Det är viktigt att rådata som samlas in under valideringen inte går förlorad efter sammanställningen, utan att samtliga rådata arkiveras för att vara spårbara i senare skeden av anläggningens livscykel.

I resultatdelen ska avvikelser som uppmärksammas under validering anges. Det ska även dokumenteras hur dessa avvikelser hanterats och vad som anses vara orsaken till att de uppstått från första början. Det ska dock stå klart att valideringsteamet inte i första hand har i uppgift att hitta orsaker och åtgärder till upptäckta avvikelser, såvida de inte är uppenbara. Huvuduppgiften är att finna avvikelser och rapportera dessa.

2. Integrerad validering

Övergripande behandlar detta avsnitt olika mätmetoder, statistiska metoder och riktlinjer för integrerad validering. Med integrerad validering avses, enligt Bakgrund, validering av hela det sammansatta kontrollrummets funktionalitet med samtliga operatörer ur ett Human Factors perspektiv. Vanligtvis används en baslinjemätning, där kriteriet för den integrerade valideringen är att interaktionen med det nya systemet ska vara fungera minst lika bra som med det gamla systemet där baslinjemätningen genomfördes. Riktlinjer för integrerad validering ges av NUREG - NUREG 0711 (O'Hara, Higgins, Perensky, Lewis & Bongarra, 2004); NUREG/CR 6393 (O'Hara m.fl., 1997).

Att genomföra en integrerad validering är en mycket omfattande och komplex process, vilket givetvis medför ett stort antal problem. Detta avsnitt syftar till att lyfta fram och belysa denna typ av problem, samt i viss utsträckning ge förslag på åtgärder. Samtliga förslag är inte helt realistiska, utan ska betraktas ur perspektivet att vi vill "kasta upp bollen" för att problemet ska komma till diskussion. Att vi pekar på vissa problem utan att ge konkreta förslag på åtgärder kan möjligen betraktas som en negativ kritik av valideringsprocessen. En förhoppning är dock vi genom att tydliggöra problem kan bidra till en process som i förlängningen kan generera lösningar. Vi tycker också att det är viktigt att befintliga problem belyses, för att säkerställa att resultaten vid valideringsprocessen ska tolkas med hänsyn till de problem och osäkerhetsfaktorer som faktiskt finns.

Stegvisa uppgraderingar

Ett problem för valideringsprocessen är att en uppgradering av ett kontrollrum kan ske i flera faser över tiden. Under detta tidsförlopp kan även förändringar i personal och arbetsrutiner förekomma (Norros, Savioja & Salo, 2009). Under sådana förhållanden är det svårt att genomföra en traditionell integrerad validering som jämför det nya och gamla kontrollrummet. Detta måste inte betyda att arbetssättet är sämre än att genomföra en enskild större förändring (se nedan, Avsnitt Design).

Design

"The envisioned world problem"

Ett fundamentalt problem med design är att designen baseras på studier av hur något är, men att designen i sig sedan förändrar förutsättningarna. Vid varje förändring uppstår alltså förändringar som är svåra att förutse, något som i litteraturen benämns "The envisioned world problem" (Hollnagel & Woods, 2005). En övergripande förändring av ett kontrollrum kan t.ex. leda till att en arbetsuppgift försvinner helt, eller att en annan person än tidigare får ansvar för den. Större flexibilitet i gränssnitt, t.ex. vid digitalisering kan också leda till att personalen själv omorganiserar sin arbetsfördelning, vilket gör att planerade test i valideringen omkullkastas. Relevansen i ett mått kan alltså påverkas av den

förändring den är ämnad att mäta. I komplexa system kan detta vara mycket svårt att förutse.

Behovet av iterativ design

Iterativ design är egentligen den enda lösningen på ”the envisioned world problem”. I en iterativ designprocess är planen redan från början att arbeta med versioner av det tänkta systemet till dess att en tillfredsställande funktionalitet är uppnådd. Samtidigt fungerar den iterativa designen åt två håll. Designerns bild av processen som han/hon designar för påverkas av hur väl den senaste versionen av systemet fungerar och bilden av vad som är målet förändras vid varje iteration. Skillnaden mellan det tänkta och det faktiska minskar alltså gradvis vid varje iteration av designen, till dess att en acceptabel nivå uppnås. Den tänkta världen närmar sig den reella världen, varför ”the envisioned world problem” till viss del kan hanteras. Iterativ design av system är därför att föredra när det är möjligt. Många gånger är det dock svårt att arbeta iterativt pga. kostnads- och tidsaspekter, speciellt i verksamheter som har behov av att fungera kontinuerligt.

En mycket viktig komponent i allt designarbete är att ta hänsyn till användarnas åsikter. De tänkta användarna bör därför medverka vid varje iteration och det är viktigt att fånga upp deras synpunkter på systemets förändringar. Samtidigt bör man vara uppmärksam på att användare kan vara konservativa och vilja behålla gamla lösningar, samt att de inte alltid har tillräckliga kunskaper om möjligheterna med nya tekniska lösningar för att fullt ut förstå potentiella fördelar.

Behovet av iterativ validering

Av samma anledning som att design bör följa ett iterativt förlopp så bör validering inte heller ske vid endast vid ett tillfälle. Istället bör validering ses som en naturlig del av designprocessen – vid varje iteration bör en validering ske som bekräftar vilka av designmålen som uppnåtts och i så fall hur väl.

Att endast genomföra iterativa valideringar av delsystem under utvecklingsprocessen räcker dock inte. När delsystem sammanfogas till ett enhetligt system kan man inte räkna med att det enhetliga systemets funktionalitet kommer att vara lika med summan av delsystemens funktionalitet. Även om delsystemen fungerar bra var för sig, kan mycket väl negativ interaktion mellan delsystemen förekomma. Om olika systemutvecklingsgrupper har designat respektive delsystem, kan divergerande designlösningar göra att det behövs olika strategier för att interagera med respektive delsystem. En integrerad validering av det enhetliga systemet är därför nödvändigt för att säkerställa att delsystemen fungerar väl tillsammans.

Iterativ validering under utvecklingsarbetet kan dock bidra till att minimera antalet problem med det enhetliga systemet. Exempelvis kan samma kriterier användas för validering av respektive delsystem, vilket bland annat kan förhindra divergerande designlösningar som gör att olika strategier behöver användas vid interaktion med respektive delsystem.

Relevanta scenarier

Att skapa scenarier som är relevanta för både baslinjemätning med gammalt system och för moderniserat system är av stor vikt vid integrerad validering

Att skapa relevanta scenarier vid en baslinjemätning kan vara problematiskt. När baslinjemätning genomförs kan det ibland vara svårt att förutse exakt vilka uppdateringar som kommer att genomföras. Ett problem vid Turbic-valideringen vid O2 (Braarud & Strand, 2007a, 2007b) var exempelvis att endast turbinoperatörens kontrollstation hade uppdaterats, medan scenarierna vid baslinjemätningen reflekterade arbetet i hela kontrollrummet. Därför användes endast 3 scenarier från baslinjemätningen, medan tre nya scenarier som reflekterade turbinoperatörens arbete skapades. För de nyskapade scenarierna fanns det därmed inte någon baslinjemätning att jämföra med.

Ett mer generellt problem med scenarier för baslinjemätning uppenbarar sig i ljuset av "the envisioned world problem" som presenterats ovan. Scenarier som ska gå att använda för framtida designlösningar måste vara av generisk karaktär för att inte bli irrelevanta om en stor förändring av kontrollrummet görs. Det är därför positivt om scenarier omfattar grundläggande funktioner och processer som måste styras oavsett gränssnittets utformning. Någon garanti för att detta är möjligt finns dock inte eftersom t.ex. framtida avancerad automation kan göra att människor inte längre utför vissa sysslor, eller att reaktorer byggs om på ett sådant sätt att kontrolluppgiften förändras fundamentalt. Enligt synpunkter vid fokusgruppen är de scenarier som rekommenderas i NUREG/CR-6393 främst utformade för testning vid nybyggnation.

Förutsättningar för validering

Oklara effekter av validering

När en integrerad validering genomförs kan det vara mycket kort tid kvar tills den planerade implementeringen av det nya kontrollrummet ska genomföras. Vi upplever det som oklart vilka riktlinjer som finns för hur stora brister som behöver framkomma vid en integrerad validering för att implementeringen ska påverkas. Exempelvis vad som krävs för att direkt stoppa implementeringsarbetet för att ge tid för korrigerande av identifierade brister.

Även om inte direkta felaktigheter identifierats som måste korrigeras, kan intervjuer med operatörerna, och deras svar på öppna frågor, ge värdefulla synpunkter på förbättringar. Vid ett allt för kort tidsintervall mellan validering och implementering förloras, eller begränsas, möjligheterna till att implementera sådana förslag.

Försökspersoner

Det kan vara svårt att ha exakts samma försökspersoner vid baslinjemätning och validering av moderniserat system. Om det är något år mellan mätningarna så har en del operatörer nyanstälts, en del har slutat och en del har befordrats.

Dessutom har de under den tid som gått mellan mättillfällena fått mer erfarenhet av sina arbetsuppgifter. Detta medför att det kan vara svårt att jämföra mätningar över tid – skillnader i mätning kan bero på att operatörerna skiljer sig snarare än att den nya designen är bättre eller sämre. Det enda sättet att hantera detta ur en experimentell synvinkel vore att genomföra simulatorprov med både den gamla och den nya designen för samma grupp, vid samma tidpunkt. Detta är dock troligen omöjligt att genomföra i praktiken.

Viktigt är att de försökspersoner som används vid valideringarna är representativa för de operatörer som normalt arbetar i kontrollrummet. Vilka som i praktiken deltar i en validering varierar troligen då kontrollrumsoperatörer trots allt är en relativt liten grupp. Arbetsbelastning, sjukdom och andra förhinder kan även påverka vilka som faktiskt kan delta.

Träningseffekter

Operatörerna har vanligtvis mycket stor vana vid det ”gamla” systemet som används för baslinjemätning, men har mindre vana av det ”nya” systemet vid validering av moderniserat system. Detta är problematiskt eftersom en lägre prestation på det nya systemet kan vara beroende av mindre vana vid detta. NUREG/CR-9393 rekommenderar därför att operatörerna får träning på det nya systemet innan valideringen genomförs, så att de uppvisar en nästan asymptotisk prestation, med vilket de menar en stabil prestation som inte varierar mellan olika försökstillfällen. Om kärnkraftverket inte har ett stort ”överskott” på operatörer finns dock rimligtvis ekonomiska och personella begränsningar för hur mycket träning som kan genomföras.

Att operatörerna i kontrollrummet har mycket stor träning, och är att betrakta som experter, medför också problem vid jämförande valideringsstudier. Det vill säga, operatörerna är så skickliga att de klarar av de flesta situationer i kontrollrummet, och anpassar sig därmed vanligen mycket bra till det nya systemet. Detta medför att uppmätta skillnader i prestation mellan kontrollrummen ofta blir mycket små. De kvantitativa mått som huvudsakligen används vid integrerad validering visar därför ofta inte några statistiskt säkerställda skillnader mellan kontrollrummen. Detta pekar på behovet av att i större utsträckning inkludera kvalitativa mätningar vid integrerad validering (se t.ex. Norros m.fl., 2009).

Att Operatörerna har mycket stor vana medför också att det är svårt för en extern observatör att bedöma om det är lätt att använda ett system. Att operatörerna till synes arbetar smidigt och friktionsfritt med systemet behöver inte innebära att det är lätt att använda, utan kan mycket väl bero på att de på grund av stor skicklighet kompenserar för brister hos systemet.

Att det går viss tid mellan baslinjemätning och validering innebär också att operatörerna som deltar har ackumulerat avsevärd arbetserfarenhet när valideringen genomförts.

Förändrad arbetssituation

Uppgradering av kontrollrum innebär ofta att äldre analog teknik byts ut mot datoriserade lösningar. Den analoga tekniken bestod av väggpaneler och strömbrytare och kontrollampor som samtliga operatörer kan se. I ett datoriserat kontrollrum är detta ofta utbytt ut mot individuella datorskärmar. Detta innebär att hela arbetssituationen förändras. Savioja, Norros och Salo (2008) ger exempel på några av de mest typiska konsekvenserna:

- Samarbetet mellan operatörerna förändras. I ett kontrollrum med väggpaneler är hela teamet medvetet om vad en enskild operatör gör.
- Uppgiftsfördelningen blir flexibel. I ett datoriserat kontrollrum kan varje operatör ta upp den skärmbild som han/hon önskar. Detta lyfter fram frågan om hur arbetsuppgifterna ska fördelas i olika situationer.
- Tidigare fanns instruktionerna på papper. Nu kan de läggas in digitalt. Det blir därför enklare att använda instruktioner samtidigt som en uppgift utförs.
- Operatörernas processbild (eng. process picture) förändras. Detta är sannolikt den viktigaste förändringen, och innebär att operatörerna får en förändrad mental modell av processen, strukturen hos dess delar (eng. component structure), ömsesidiga beroenden mellan delar, funktionaliteter etc.

Dessa förändringar innebär, enligt Savioja m.fl. (2008), att det behövs nya teoretiska koncept för att analysera operatörernas arbete och de verktyg som de använder under arbetet. De har därför skapat en utvärderingsmetod som de kallar för Contextual Assessment of Systems Usability (CASU) (se nedan, Avsnitt Kontextuell värdering av systemets användbarhet).

Rollfördelning mellan operatörer

När ett kontrollrum digitaliseras kan rollfördelningen mellan operatörerna påverkas på ett flertal sätt. Om kontrollrummet blir helt digitalt skulle operatörerna i princip kunna byta roller med varandra. Hierarkin mellan operatörerna kan delvis förändras när turbinoperatören, som vid OKG, som vanligtvis har kortast arbetstid som operatör får den mest moderna utrustningen. Turbinoperatören fick därmed initialt bättre kunskap om hur turbinen kontrolleras än reaktoroperatören, som i regel har avancerat från att vara turbinoperatör till att bli reaktoroperatör.

Acceptanskriterier

NUREG rekommenderar en stor mängd olika typer av mätmetoder. Detta är bra för att få en helhetsbild av systemet. Det kan dock medföra problem avseende tolkning av data. Mätmetoderna visar ofta endast små skillnader mellan baslinjemätning och integrerad validering, vilket gör att slumpvisa variationer får relativt stort utslag. Om olika mätmetoder visar små skillnader för olika scenarier och operatörer, kan det därför vara svårt att tolka resultaten.

Vid jämförelse med baslinjemätning är kravet att resultaten i det nya kontrollrummet ska vara lika bra eller bättre än för det gamla systemet. Detta är problematiskt eftersom varje kärnkraftverk då använder sitt eget gamla kontrollrum som baslinje. Det vill säga om det finns skillnader i hur bra de

gamla kontrollrummen är på de olika kärnkraftverken, så kan dessa skillnader fortleva i nya generationer av kontrollrum.

Acceptanskriterier då baslinjemätning saknas är också problematiskt. Vår tolkning är att respektive valideringsteam skapar acceptanskriterier med avseende på de scenarier som ingår i valideringen. Detta kan medföra skillnader mellan vad som anses acceptabelt mellan valideringar av olika kontrollrum.

Validering med baslinjemätning

Baslinjemätning brukar göras vid det gamla kontrollrummet för att användas som referensmått vid integrerad validering av det moderniserade kontrollrummet. Det finns dock flera problem med att jämföra med baslinjemätningen vid en integrerad validering

Uppdatering av kontrollrummet kan ske i flera steg, exempelvis kan först en uppdatering genomföras på turbinsidan och något år senare på reaktorsidan. Att baslinjemätning från det "gamla" kontrollrummet i så fall ska användas som referensmätning vid valideringen av uppdateringen av den första uppgraderingen vid turbinsidan är självklart. Men vid den följande uppdateringen av reaktorsidan är det inte självklart om det är det "gamla" kontrollrummet, eller kontrollrummet med uppdaterad turbinsida som bör användas som baslinjemätning. Det vill säga om uppgraderingar genomförs i etapper så är det inte uppenbart om etapp två av uppgraderingen ska jämföras med det ursprungliga systemet, eller med systemet från uppgraderingen vid etapp ett.

Ibland har en viss tid gått sedan baslinjemätningen genomfördes, vilket kan innebära att planer har ändrats, varför scenarierna vid baslinjemätningen inte är helt relevanta för den integrerade valideringen vid den genomförda uppgraderingen. Så var delvis fallet vid den integrerade valideringen av O2, där baslinjemätningen hade genomförts med syftet att utgöra baslinjemätning för hela kontrollrummet. Eftersom uppgraderingen endast avsåg turbinoperatörens del av kontrollrummet var det tre scenarier vid baslinjemätningen som inte var helt relevanta. Dessa togs då bort och tre nya skapades vid valideringen, för vilka jämförelse inte kunde göras med baslinjemätningen (se Braarud m.fl., 2008a).

Observera att vårt syfte med att påtala problem med jämförelsen med baslinjemätning inte är att avråda från att baslinjemätning genomförs och används vid integrerad validering. Tvärtom, det finns flera fördelar med dessa jämförelser. Vi vill dock tydliggöra nackdelarna, vilket förhoppningsvis kan leda till en större användning av alternativa metoder för att komplettera den integrerade valideringen med andra metoder och acceptanskriterier.

Validering utan baslinjemätning

När en integrerad validering genomförs kan det vara så att data från baslinjemätning saknas. Fallet kan också vara att baslinjemätning finns, men att jämförelsen inte är relevant. Baslinjemätningen kan exempelvis ha gjorts med syfte på en annan uppdatering, varför de scenarier som används inte är

anpassade för den implementerade uppdateringen som ska valideras (se exempel ovan). Fallet kan också vara att förändringen av kontrollrummet är så stor att förändringarna i operatörernas interaktion med systemet och med varandra blir diametralt annorlunda. I så fall kan det vara tveksamt om det är relevant att jämföra med det gamla systemet. Detta diskuteras också i fokusgruppen (se nedan Avsnitt 3 Fokusgrupp).

Inom det militära området uppstod exempelvis detta ”problem” när 39-systemet (JAS) skulle valideras. Skillnaden mot 37 systemet (Viggen) var så diametral att det snarast var att betrakta om ett paradigmskifte, med ett fullständigt nytt system som kunde lösa helt nya uppgifter på ett helt nytt sätt.

När det gäller förändring av kontrollrum i kärnkraftverk kanske förändringen mellan ett analogt och digitalt kontrollrum inte är lika dramatisk eftersom det i princip fortfarande är samma arbetsuppgifter. Det skulle dock gå att tänka sig förändringar som gör att antalet operatörer förändras. Jämförelsen med 37 vs 39 problematiken kan dock bli aktuell vid eventuell nybyggnation av kärnkraftverk. Nya tekniska lösningar skulle kunna medföra så stora skillnader gentemot nuvarande reaktorer och turbiner att operatörernas arbetsuppgifter förändras på ett avgörande sätt.

En möjlighet är att t.ex. använda kriterier från säkerhetsföreskrifter i kombination med förväntade prestationsmått. I en sådan situation skulle någon typ av gemensamma acceptanskriterier vara till hjälp. En mycket viktig aspekt vid en sådan värdering är att också ta hänsyn till användaracceptans. Operatörens åsikter om systemet bör fångas upp så tidigt som möjligt.

Kontextuell värdering av systemets användbarhet

Savioja m.fl. (2008) menar att ett system måste ha bra informationsinnehåll och informationspresentation för att fungera. Informationsinnehållet måste vara kontextuellt meningsfullt och presenteras så att operatörerna kan tolka och använda informationen på rätt sätt. Savioja m.fl. har därför utvecklat en valideringsmetod som de kallar för *Kontextuell värdering av systemets användbarhet (CASU, Contextual Assessment of Systems Usability)* (se t.ex. Savioja & Norros, 2008; Savioja m.fl., 2008; Norros & Savioja, 2006). Metoden bygger på teori och tillämpning från core-task analys och består av fyra åtskilda faser: modellering, datainsamling, dataanalys och värdering av systemets användbarhet.

Under modelleringsfasen ska de kriterier formuleras som ska användas för att värdera systemets användbarhet. Målsättningen är att definiera vilka scenarier (process situations) som ska användas och vilka kriterier som mätningarna ska jämföras emot.

Datainsamlingen syftar till att samla in information om hur systemet används. Datainsamling görs både kvantitativt och kvalitativt. Efter datainsamling i simulerat kontrollrum, genomförs även intervjuer med operatörerna.

Analysfasen syftar till att skapa en bild av hur väl systemet uppfyller kraven på användbarhet. De viktiga resultaten från analysfasen är analys av prestation avseende processkontroll, analys av arbetssättet och analys av systemets förväntade lämplighet. I värderingsfasen jämförs resultaten med de mått som valdes i modelleringsfasen. Syftet är att kunna dra slutsatser om hur väl det nya systemet uppfyller kraven på användbarhet. För analys av systemets instrumentella funktionalitet är prestationsdata det viktigaste måttet vid processkontroll. För analys av psykologisk funktionalitet studeras operatörernas sätt att arbeta. Exempelvis hur operatörerna agerar, hur de söker information, hur de utnyttjar olika operationella möjligheter och hur de använder instruktioner. Den här informationen ger tillsammans med de avslutande intervjuerna en värdering av hur väl systemet passar för operatörernas krav och mentala modeller av processen.

För utförligare beskrivning av CASU, se t.ex. Savioja m.fl. (2008) samt Norros och Savioja (2006). Savioja m.fl. (2008) beskriver även en fallstudie där CASU använts vid validering av kontrollrummet i ett kärnkraftsverk. I fallstudien användes följande mått för att värdera hur framgångsrik processkontrollen var i olika scenarier:

- Tid för att identifiera en grundläggande funktion som var hotad.
- Tid för att korrekt diagnostisera felet
- Antal korrekta och adekvata beslut och operationer som utförts.
- Nivå på vilken verket stabiliserats.

Denna metod liknar i stort den handledning för validering som Rollenhagen m.fl. (1998) tagit fram (presenterad ovan i Avsnitt 1 Bakgrund). Den har också likheter med den syn på validering som tillämpas av IFE. Skillnaden ligger till stor del i kontexten, dvs. att Savioja m.fl. gör en poäng av att utgå från en modell av systemet som ska förändras, vilken bygger på analys av existerande system och arbetsprocesser. Detta är inte på samma sätt uttalat i de andra metoderna, där man istället pratar om förutsättningar och hypoteser. Hur till vida metoden som föreslås av Savioja m.fl. har några fördelar jämfört med de som används i Sverige idag är oklart.

Internationella Guidelines

Detta avsnitt presenterar kort några av de mest kända riktlinjerna som vi stött på under arbetet med rapporten.

NUREG

NUREG (US Nuclear Regulatory Commission Regulation) har skapat de mest etablerade riktlinjerna för design och validering av kontrollrum vid kärnkraftverk.

NUREG/CR-6393 (O'Hara m.fl., 1997) ger ett utförligt ramverk för hur en integrerad validering ska genomföras. Den innehåller ett stort antal kriterier och mätmetoder för integrerad validering. Flertalet rapporter som vi har läst använder rekommendationer från NUREG/CR6393.

Att rekommendationerna är så omfattande innebär att en integrerad validering som följer dem kan bli mycket tids- och kostnadskrävande. Det skulle möjligen ha varit en fördel om det funnits olika nivåer med rekommendationer, beroende på hur stor uppgradering av kontrollrum som gjorts. Braarud och Skraaning (2006) kommenterar att NUREG/CR-6396 inte ger mycket vägledning för hur man ska bedöma vad som är en ”betydelsefull” (noteworthy) skillnad mellan baslinjemätning och moderniserat system. NUREG ger inte heller rekommendationer avseende hur operatörernas förändrade arbetsituation vid uppdatering från analogt till digitalt kontrollrum ska hanteras.

IFE

IFE som ledande forskningsinstitut inom EU inom kärnkraftsområdet kanske borde få i uppdrag att utveckla riktlinjer för validering av kontrollrum. Detta skulle kunna säkerställa att samma krav ställs vid samtliga anläggningar inom EU. Vår uppfattning är att IFE arbetar utifrån NUREG, men även har ett antal interna dokument som används vid integrerad validering. Dessa är dock inte publika, varför vi avstår från att referera dem. IFE förtjänar trots det att nämnas då de är tongivande för många av de tankegångar som finns rörande integrerad validering och därtill ofta är inblandade i eller utförare av sådana valideringar.

ISO

Internationella Standardiserings Organisationen (ISO, International Organization for Standardization) och IEC (International Electrotechnical Commission) har utarbetat ett stort antal riktlinjer för gränssnittsdesign. Dokumenten kan inte direkt användas till en integrerad validering. De kan dock vara intressanta vid preliminära utvärderingar med explorativa metoder, som t.ex. walkthroughs och heuristisk utvärdering, exempelvis för att bedöma om gränssnittets design lever upp till ställda krav. Här ges några exempel på designriktade guidelines från ISO:

- ISO/TR 18529:2000 – Vägledning för ergonomi och Human Factors centrerade processer, med inriktning mot dem som genomför design och utvärdering.
- ISO 13407:1999 - Översikt av ergonomi Human Factors centrerade designaktiviteter, avsedda för projektledare och chefer över designprocessen.
- ISO:11064-4:2004 – Rekommendationer för ergonomi vid arbetsstationer i kontrollrum.
- *ISO 9241-12:1998* – Rekommendationer för presentation och representation av information på visuella displayer, t.ex. alfanumeriska och grafiska/symboliska koder, skärmlayout, och design).

- *ISO 9241-13:1998* – Rekommendationer för design och utvärdering av användargränssnitt, omfattande markörer, återkoppling, status, on-line hjälp, och felhantering.
- *ISO 9241-14: 1997* – Rekommendationer för ergonomisk design av menyer vid datorbaserade dialoger.
- *ISO/IEC 11581-1:2000; ISO/IEC 11581-2: 2000; ISO/IEC 11581-3: 2000; ISO/IEC 11581-6:1999* – Ett allmänt ramverk för utveckling och design av ikoner
- *IEC 80416-3:2002; ISO 7000:2004* – Principer för grafiska symboler vid systemdesign.

ISO/TR 16982:2002 ger information om Human Factors centrerade metoder för design och validering. Information ges om för- och nackdelar hos olika metoder. Dokumentet vänder sig i första hand till projektledare för övergripande förståelse för metoderna. Konkreta beskrivningar om hur metoderna används ges därför inte

Guidelines från andra domäner

Vi tar här upp några andra guidelines, främst från det militära området. Deras rekommendationer är inte direkt applicerbara på kontrollrumsdesign och validering av kontrollrum i kärnkraftverk. Flera rekommendationer och resultat i dessa guidelines är dock generella och kan därför bidra med värdefull information. De metoder som exempelvis GARTEUR (Castor m.fl., 2003) beskriver för mätning av mental arbetsbelastning är även intressanta för motsvarande mätningar inom kärnkraftsområdet.

GARTEUR

GARTEUR (Group for Aeronautical Research and Technology in Europe är en ramverksorganisation för europeisk flygforskning). En arbetsgrupp (GARTEUR FM AG13 Action Group on Mental Workload Measurement) har utarbetat vägledande principer för val av mätmetoder vid bedömning av mental arbetsbelastning. Under 1999-2003 utfördes detta arbete av nio forskningsinstitut och flygindustrier. Ansvar för ordförandeskapet togs av FOI. En slutrapport har publicerats av Castor m.fl. (2003).

VINTECH

VINTECH (Visual Interaction and Human Effectiveness in the Cockpit) är ett EU-finansierat civilt flygforskningsprojekt med syftet att utveckla metodik för värdering av enskilt och gemensamt situationsmedvetande (team-SA eller gemensam lägesuppfattning) hos flygbesättningar. Ett flertal flygforskningsorganisationer, bland andra FOI och flygindustrin (Saab) har deltagit i projektet. VINTECH I fokuserade huvudsakligen på situationsmedvetande hos den enskilde piloten. VINTECH II har vidareutvecklat metodiken från VINTECH I och därutöver utforskat värdering av team-aspekter för flygbesättningar, med inriktning mot gemensamt situationsmedvetande, t.ex. samordning inom besättningen och interaktion med flygledning, autopilot och

andra automatiserade system. Metodiken fokuserade på att studera ögonrörelser hos piloterna (Zon m.fl., 2004).

MANPRINT

MANPRINT är ett program som utvecklats inom den amerikanska armén och för dess organisation. Beteckningen MANPRINT står för MANpower and PeRsonnel INTeGration. MANPRINT kan i vissa avseenden betraktas som en militär version av civil Människa-System-Interaktion (MSI) och Human Factors forskning (se t.ex. Booher, 1990; MANPRINT Handbook, 2005).

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. Detta ska, enligt MANPRINT Handbook (2005), uppnås genom fortlöpande integration av sju humanrelaterade områden, så kallade MANPRINT-områden: manpower, personalens kapacitet, träning, Human Factors engineering, systemsäkerhet, hälsorisker och soldatens överlevnadsförmåga.

The Command and Control Research Program

The Command and Control Research Program är ett forskningsprogram inom det amerikanska försvarsdepartementet (DoD). Inom forskningsprogrammet har ett flertal sammanställningar gjorts av tillämpad försöksmetodik för det militära området. Nedan presenteras ett par av dessa publikationer som även kan vara intressanta för kärnkraftsområdet.

- *Nato Code of best practice for C2 Assessment* (2002) behandlar experiment och försöksmetodik vid utvärdering av ledning och ledningssystem.
- *Code of best practice for experimentation* (Alberts & Hayes, 2002) behandlar planering, design och genomförande av experiment, samt användning av resultat från experiment. Syftet är att ge ökad medvetenhet och förståelse för olika typer av experimentella aktiviteter. De diskuterar fördelarna med experimentkampanjer, i stället för individuella experiment, för bredare fokus och säkrare resultat genom upprepade resultat under olika förhållanden.

Dessa publikationer, och ett flertal ytterligare, kan laddas ner från:
http://www.dodccrp.org/html4/books_downloads.html

Synpunkter på guidelines

De guidelines som NUREG har skapat är utförliga och täcker till mycket stor del behovet för integrerad validering. Det verkar dock finnas vissa saker som saknas.

Det skulle behövas tydligare riktlinjer för hur man ska bedöma skillnaden mellan moderniserat system och baslinjemätning. Det står heller inte mycket om

hur man ska hantera skillnaderna i arbetssituation för operatörerna när ett gammalt analogt kontrollrum byts ut mot ett digitaliserat.

Dynamiska mätningar tas upp av NUREG, men borde lyftas fram tydligare.

När det gäller baslinjemätning skulle det vara praktiskt med någon typ av referensmått. Integrerad validering utgår från baslinjemätning vid det gamla kontrollrummet i det kärnkraftsverk som moderniserats. Detta är problematiskt på så sätt att det kan skilja sig åt vid olika kärnkraftverk med avseende på hur bra det gamla kontrollrummet är ur ett Human Factors perspektiv. Vilket i sin tur innebär att kraven vid validering av ett moderniserat kontrollrum kan skilja sig åt vid olika kärnkraftverk.

När det gäller kraven på design för kontrollrum kanske det borde finnas någon typ av referensanläggning för att man vid utformning av ett kontrollrum ska veta vad som är bra design och vad som redan har visat sig vara dåliga lösningar. Vi inser dock att detta kan vara problematiskt då det är flera olika konkurrerande aktörer som utvecklar kontrollrum.

Mätmetoder vid validering

Att fånga upp vad som faktiskt händer i kontrollrummet kan vara problematiskt vid integrerad validering. Händelseförlopp kan utspelas mycket snabbt i ett kontrollrum, där vältränade operatörer vid en incident mycket väl kan ha fått en situation under kontroll innan en extern observatör ens börjat förstå vad som händer (Savioja & Norros, 2008). Detta problem undanröjs till stor del genom specialtränade observatörer med stor domänkunskap, som dessutom har kännedom om sceneriet under valideringen, och därmed i förväg känner till vad som kommer att hända.

Detta stycke tar upp en relativt stor mängd av mätmetoder. Stycket ska dock inte ses som någon typ av guidelines för integrerad validering. Syftet är snarast att belysa för- och nackdelar med ett antal av de metoder som används, samt att ge förslag på kompletterande metoder.

Explorativa metoder

Explorativa metoder används för att kravställa system och upptäcka problem innan systemet är färdigutvecklat. Uppgifts- och användaranalys används företrädesvis innan systemdesign påbörjas. Walkthroughs och heuristiska utvärderingar kan användas i tidiga faser under systemutveckling för att identifiera problem innan systemet är färdigutvecklat.

Uppgiftsanalys

Uppgiftsanalys är en generell term för en relativt mångfacetterad samling metoder som bland annat omfattar framtagning av beskrivningar av vilka uppgifter människor utför, för att representera dessa beskrivningar, förutsäga problem, utvärdera system med avseende på användbarhet eller funktionella krav, förutsäga prestation, mäta systemets komplexitet och lärbarhet (Hollnagel,

1998). Uppgiftsanalys beskrivs av Kirwan och Ainsworth (1992) som processen att kritiskt granska uppgiftens beståndsdelar – operatörens resurser, begränsningar och vad han/hon föredrar i syfte att fastställa hur detta påverkar människans operationer för att uppnå systemets mål.

Anledningen till att uppgiftsanalys utförs är att de kan ge en tydlig bild av vilka uppgifter ett system ska utföra (Faulkner, 2000). Uppgiftsanalys används inte bara till beskrivningar av de handlingar och uppgifter som behöver utföras för att lösa en arbetsuppgift, utan syftar även till att identifiera prestationsproblem och felkällor och föreslå lösningar för dessa (Annett, 2003). Uppgiftsanalys kan antingen fokusera på fysiska handlingar eller kognitiva operationer (kognitiv uppgiftsanalys) som måste utföras för att målet ska uppnås (Wickens & Hollands, 2000).

Vid hierarkisk uppgiftsanalys används en generell struktur för att dela upp uppgifter i underuppgifter, tills man kommer till en nivå där uppgifterna är så enkla att vidare uppdelning inte längre behövs. Eftersom strukturen är generell kan den även användas vid de flesta typer av uppgiftsanalys.

Det finns ett stort antal av olika metoder för uppgiftsanalys. Här ges exempel på några olika typer:

- *ACTA (Applied Cognitive Task Analysis)* Metod för kognitiv uppgiftsanalys. Används för allmän kognitiv analys av operatörers kognitiva processer. Består av uppgiftsdiagram, kunskapsgranskning och simuleringsintervjuer.
- *Concept mapping and expert design storyboarding* En analytiker genomför ostrukturerade intervjuer med en expert, vilket ligger till grund för begreppskartor.
- *Critical Incidents Technique* Utgångspunkt är att experter som utför normala, välbekanta uppgifter inte ger mycket insikt. I stället fokuseras på situationer då de har varit nära att begå misstag, eller på svåra och ovanliga situationer.
- *GOMS (Goals Operators Methods and Selection rules)* Ett antal besläktade metoder för uppgiftsanalys. Används för modellering av procedurell kunskap.
- *Hierarkisk uppgiftsanalys* Uppgifter delas upp i deluppgifter, vilka i sin tur genomgår en egen uppgiftsanalys, osv. tills tillräcklig detaljeringsnivå uppnåtts.
- *Kognitiv uppgiftsanalys* Fokuserar på användarnas kognitiva processer. Bygger på att människa system interaktion vanligen består av komplexa kognitiva processer, och därför inte kan studeras med traditionella beteendeinriktade uppgiftsanalys.
- *Konceptuell grafanalys* Konceptuella grafer, bestående av semantiska nätverk skapas. Metoder som används är bl.a. intervjuer och videoinspelningar, av experter som arbetar, vilka i efterhand analyseras av analytiker och experter tillsammans.

- *PARI* Metod för kognitiv uppgiftsanalys. Experter genererar först själva problem som de är välbekanta med. Därefter paras varje expert ihop med en annan expert, med vilken de fortsätter att diagnostisera problemet.
- *CREAM* Cognitive Reliability and Error Analysis Method. Egentligen andra generationens HRA-metod. Används för att utvärdera säkerhet i ett system. Bygger på uppgiftsanalys och analys av arbetssituationen.

Vid integrerad validering används uppgiftsanalys främst vid scenarioutveckling för att undersöka vilka arbetsuppgifter som bör studeras (se NUREG; Eriksson m.fl., 1998)

Walkthroughs

Walkthroughs är en typ av strukturerade metoder där användbarhetsexperter, i detta fall processexperter och erfarna operatörer tillsammans med Human Factors experter, går igenom systemet/prototypen med syftet att hitta användbarhetsproblem tidigt i utvecklingscykeln. Vid genomgången följer experten/operatören ett antal noggrant definierade arbetsuppgifter (se t.ex. Alfredson, Oskarsson, Castor & Svensson, 2004b; Lewis & Warton, 1997; Preece m.fl., 1994; Preece, Rogers & Sharp, 2002).

Vid kognitiv walkthrough fokuseras på kognitiva processer, med målsättningen att identifiera potentiella användarproblem. Arbetssättet är, enligt Preece m.fl. (1994), närbesläktat med kognitiv uppgiftsanalys. Vid kognitiv walkthrough kombineras följaktligen den walkthrough som görs av systemet, gränssnittet eller programvaran med en walkthrough med avseende på användarnas kognitiva (eller mentala) modeller. Enligt Faulkner (2000) kan det beskrivas som att experten försöker imitera den tänkta användarens arbete med systemet och därmed också försöka upptäcka de problem som användaren skulle råka ut för vid respektive stadium.

Pluralistisk walkthrough innebär att systemet gås igenom av personer med olika typer av kompetens, som produktutvecklare/designers, användbarhetsspecialister och användare. Detta kan vara fördelaktigt eftersom Human Factors experters genomgångar kan avvika från riktiga användares, beroende på att användare i regel har mindre kunskap om gränssnittets funktionalitet och sammansättning, jämfört med experten.

Termen kognitiv jogthrough förekommer även, och betecknar en snabbare och mindre formell genomgång av systemet än den som görs vid vanlig walkthrough

Heuristisk utvärdering

Heuristisk utvärdering är en metod för systematisk granskning/inspektion av ett gränssnitts design med avseende på användbarhet. Vid heuristisk utvärdering går experter igenom systemet. Ofta används Nielsens principer (Nielsen, 1993, 1994, 2004), vilket är en lista med tio principer, eller tumregler för hur ett system bör vara utformat. Fem utvärderare brukar uppskattningsvis hitta 75 procent av problemen med systemets användbarhet, och anses därför vanligen som det mest kostnadseffektiva antalet (Preece m.fl., 1994).

Nielsens principer bör kallas för "heuristiker" eftersom de snarare ska ses som tumregler än specifika riktlinjer för användbarhet (Nielsen, 1993). Principerna kan även vara ett bra stöd för den som designar och konstruerar ett gränssnitt. Efter att de ursprungliga principerna publicerats (Nielsen & Molich, 1990), har Nielsen med hjälp av faktoranalys av 249 användbarhetsproblem förfinat heuristiken, vilket resulterat i en reviderad uppsättning riktlinjer (Nielsen, 1994, 2004) vilka presenteras här:

- *Synlig systemstatus* – Systemet ska informera användarna om vad som händer genom feedback inom rimlig tid.
- *Överensstämmelse mellan systemet och den verkliga världen* – Systemet ska tala användarens språk, följa vedertagna regler och presentera information i naturlig och logisk ordning.
- *Användarkontroll och frihet* – Användare aktiverar ofta funktioner av misstag, och behöver då tydligt markerade "nödutgångar". Dvs. ge stöd för möjlighet att ångra.
- *Konsistens och standard* – Användare ska inte behöva vara osäkra på om olika ord, situationer eller handlingar innebär samma sak eller inte. Dvs. följ systemets vedertagna terminologi.
- *Förhindra fel* – Bättre än bra felmeddelanden är omsorgsfull design som förebygger att problem uppstår.
- *Igenkänning istället för minne* – Gör objekt, handlingar etc. synliga. Användaren ska inte behöva minnas information mellan olika delar av interaktionen med systemet. Användarinstruktioner ska vara synliga och lättillgängliga när de behövs.
- *Flexibilitet och effektiv användning* – Genvägar, osynliga för nybörjare, kan påskynda experters interaktion med systemet. Möjliggör för användarna att själva anpassa frekvent använda funktioner.
- *Eстетisk och minimalistisk design* – Interaktionsdialogen ska inte innehålla information som är irrelevant eller sällan behövs. Varje extra informationsenhet konkurrerar med relevant information och minskar dialogens relativa synlighet.
- *Stöd användarnas igenkänning, diagnos och återhämtning från fel* – Felmeddelanden ska uttryckas med vanligt språk (inga koder), exakt ange problemet, och konstruktivt föreslå en lösning.

Hjälp och dokumentation – Information ska vara lätt att söka, vara fokuserad på användarens uppgift, räkna upp konkreta åtgärder och inte vara för omfattande.

Kvantitativa metoder

Kvantitativa metoder avser mätmetoder som samlar in kvantitativa mätdata. Typiska mått är reaktionstid, utförandetid och felfrekvens. Kvantitativa data kan samlas in på ett flertal olika sätt. Exempelvis automatiskt genom loggning eller psykofysiologisk mätning, men även genom subjektiva egenskattningar och observation, som exempelvis kvantifieras på en skattningsskala.

Vid studier av enskilda isolerade fenomen i kontrollerade experimentella laboriestudier, där potentiella störvariabler kan elimineras, konstanthållas

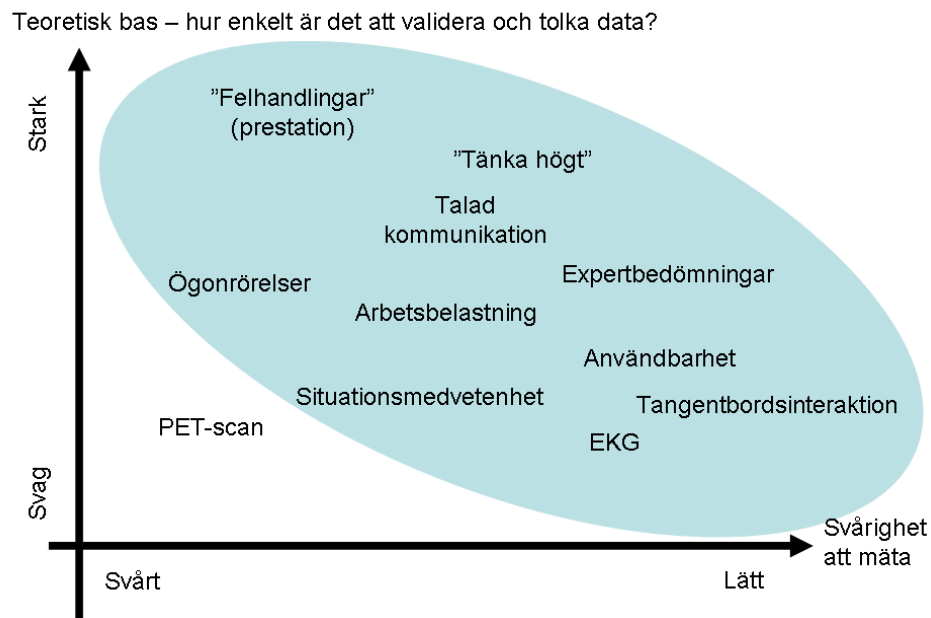
eller balanseras ut, är kvantitativa mätningar vedertagen metodik. Inom exempelvis experimentell perceptions- och kognitionspsykologisk forskning uppfylls vanligen dessa kriterier.

Vid utvärderingar och valideringar avseende användbarhet vid systemutveckling finns vanligen inte samma möjlighet till kontroll. Logiskt sett kan det tyckas att traditionella vetenskapliga laboriestudier alltid borde vara tillämpbara om det bara fanns mer tid, mer data samlades in och kontrollen över testvillkoren vore större. Hennessy (1990) anser exempelvis att detta inte är praktiskt möjligt eftersom realism och kontroll knappast kan bibehållas samtidigt. Hennessy menar också att ett problem med kvantitativa mätningar är att data ofta kommer från enkla händelser på låg nivå, exempelvis en knapptryckning, medan knapptryckningen kan föregås av en lång och sammansatt process av perception och beslutsfattande.

Man bör också vara medveten om att kvantitativa mätmetoder inte alltid är helt objektiva. Om man som i Hennessys (1990) exempel mäter knapptryckningar är det vanligen inte antal knapptryckningar i sig som är det intressanta, utan knapptryckningarnas innebörd för användarens interaktion med systemet. Även om sambandet kan byggas på välgrundad teori, finns ofta ett subjektivt inslag i forskarens tolkning av detta samband.

Utvärderingar avseende användbarhet och valideringsstudier är ofta explorativa. Det vill säga det är inte alltid det finns någon tydlig teori om vilka problem som kan finnas hos systemet. Kontrollerade laboriestudier behöver dock vanligen designas för att en viss fråga ska kunna studeras. Vid utvärdering avseende användbarhetsfrågor är detta problematiskt eftersom en designad studie bygger på forskarens teori om vilka användbarhetsproblem som kan finnas hos ett system, vilket innebär att oförutsedda användbarhetsproblem som studien inte designats för att undersöka ofta inte upptäcks (Nielsen, 1993).

Kvantitativa mått måste vara möjliga, reliabla och meningsfulla. Detta gäller självfallet även kvalitativa data. Mätningar kan göras på en mängd olika sätt, t.ex. instrumenterat genom EKG, loggning i olika operatörsgränssnitt/handlingar, genom observationsprotokoll, subjektiva bedömningar etc. I enlighet med diskussionen ovan kan mått vara mer eller mindre väl kopplade till teori – ett mått kan vara enkelt att samla in, men ha lite värde och vice versa kan måttet vara svårt att samla in, men ha stort värde om det är möjligt. I de flesta studier är måtten som används kompromisser – dvs. de är de mått som är rimligast att samla in jämfört med ansträngningen att göra det. Hollnagel och Woods (2005) beskriver förhållandet mellan några typiska ”Mått” och deras teoretiska förankring respektive svårighet att samla in dem (se Figur 3).



Figur 3. Förhållandet mellan några typiska "Mått" och deras teoretiska förankring, respektive svårighet att samla in dem (Anpassat från Hollnagel & Woods, 2005).

I likhet med Hennessy (1990) påpekar Hollnagel & Woods (2005) att t.ex. tangentbordsinteraktion kan vara en mycket reliabel datakälla, men trots det sakna förklaringsvärde. Den underliggande modell som används vid valideringen är därför av största vikt, något som talar för Savioja och Norros (2008) kontextuella ansats (se ovan, Avsnitt Kontextuell värdering av systemets användbarhet) vilken utgår från en modell av den anläggning som ska förändras. Detta eliminerar dock inte "the envisioned world problem" (se ovan).

I linje med detta är det inte trivialt att i ett kontrollrum där flera operatörer interagerar bedöma vad som är en bra prestation. Att bedöma hur fort en uppgift A utförs kan ge en skev bild ifall operatörerna har prioriterat uppgift B. Det kan mycket väl vara så att uppgift A måste lösas, men att det inte spelar någon roll om det tar lite tid, varför andra uppgifter kan prioriteras.

Det går visserligen att skapa scenarier, där endast en uppgift ska lösas samtidigt, eller där flera lika prioriterade uppgifter ska lösas samtidigt. Detta kan dock leda till konstlade scenarier med lägre validitet hos den integrerade valideringens resultat. Det kan även vara svårt att tolka innebörden av prestationsmått, exempelvis vad det egentligen betyder för systemets användbarhet om en operatör har tryckt x antal gånger på en knapp. Sätt att mäta och samla in data diskuteras i fokusgruppen.

Dessa invändningar innebär dock inte att kvantitativa prestationsmått bör undvikas, utan att de bör kompletteras med andra mätmetoder. För att få ett bra mått på prestation är därför kvalitativa bedömningar av experter som observerar, samt operatörernas egenskattningar av prestationen av stor vikt.

Kvalitativa metoder

Med kvalitativa metoder avses vanligen mätmetoder som ger data som kan kategoriseras, men inte översättas till numeriska mått. Kvalitativa metoder kan exempelvis vara intervjuer, observation och att forskaren själv befinner sig i den miljö som analyseras, med syftet att försöka att uppfånga både användarnas handlingar och innebörden av deras handlingar och därmed kunna ge en helhetsbeskrivning.

Fördelar med kvalitativa metoder är bland annat att de kan ge information som är svårt att komma åt med kvantitativa metoder. Exempelvis om problem och fördelar hos system, vilket kan ge konkret information om vilka förändringar som bör genomföras. Det är också lättare att upptäcka oförutsedda användbarhetsproblem med kvalitativa metoder. Kvalitativa metoder kräver heller inte samma typ av försökskontroll som kvantitativa metoder.

Det finns dock även nackdelar med kvalitativa metoder. Insamling av observationsdata är exempelvis mycket krävande, vilket kan påverka observatörens tillförlitlighet (Charlton & O'Brien, 1996). Även om observatörer och utvärderare som använder subjektiva metoder vanligen är mycket vältränade för detta, finns risk för personligt godtycke. Ett annat problem är att personer som genomför utvärderingar med observationsmetoder behöver ha kunskap om både tekniken och om de människor som ska använda systemet. Det kan också krävas att de är ämnesområdesexperter inom systemets arbetsområde för att förstå dess funktionalitet. Dessutom bör de vara Human Factors specialister för att ha tillräcklig kunskap om hur utvärderingar ska genomföras. Tillgången till kompetenta observatörer är därmed vanligen en bristvara.

Intervju

Strukturerade intervjuer med operatörer efter ett genomfört scenario kan ge mer konkret information om både för- och nackdelar med systemet. Här skulle resultat från exempelvis systemloggar, psykofysiologisk data och observation kunna användas. Exempelvis skulle kritiska episoder kunna spelas upp för operatören under intervjun. Exempelvis episoder där systemloggar visar avvikelser, där psykofysiologisk data indikerar förändring hos operatörens mentala status, eller där observatören identifierat felaktigt, eller avvikande, beteende.

Enkätundersökning

Enkätfrågor kan formuleras på olika sätt. Vanligtvis används slutna frågor med i förväg definierade alternativ. Vid den enklaste formen finns endast svarsalternativen: ja, nej och vet ej. Fler alternativ ger dock bättre underlag för statistisk analys och tolkning. Vid skattningsskalor med flera alternativ innebär ett lägre värde exempelvis en mer negativ bedömning, eller tvärtom.

Skattningsskalor där varje alternativ är ett påstående kallas för Likertskalor. Oavsett vilken typ av skala som används så brukar 1-7 eller 1-5 alternativ användas, vilket ger ett neutralt mittvärde. Ibland används skalor utan mittvärde, t.ex. 1-6. Nackdelen med att utelämna mittkategorin är dock att vissa svar då kan avspegla artificiella uppfattningar om något som de som svarar inte har någon riktig uppfattning om (se t.ex. Faulkner, 2000).

Ibland används även rangordning, där den som svarar får rangordna olika svarsalternativ. Nackdelen är att analys endast ger ordningen mellan alternativen, men inte storleken på skillnaden det är mellan dem. Om i stället skattningsskalor från 1-7 används för varje alternativ ger analysen medelvärden som visar både rangordning och storleksskillnad, och dessutom ett spridningsmått.

Öppna frågor kan ge mycket värdefull kvalitativ information som kan förtydliga och förklara negativa och positiva skattningar. Dessutom kan öppna frågor bidra till att problem och fördelar identifieras som man inte ens tänkt på att fråga om. Svar från öppna frågor kan dock vara svårtolkade, och kräva mycket bearbetning.

Med avseende på i vilken ordning frågor ska komma kan följande vara tänkvärt. Vanligen bör generella frågor komma före specifika, och frågor om frekvent förekommande händelser bör komma före frågor om sällsynta händelser (Charlton, 1996b).

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, som all annan metodik för användbarhetsstudier, därför testas genom pilotstudier innan de används vid utvärderingar (Nielsen, 1993).

För att utveckla och validera ett enkätinstrument så kan faktoranalytisk ansats på insamlade svar vara ett lämpligt tillvägagångssätt. Detta kan verifiera kategoriindelning och reducera antalet frågor. Verifiering av kategoriindelning är synnerligen viktigt om ett antal frågor som antas höra samman används för att skapa ett övergripande medelvärde för en kategori. Att reducera antalet frågor är viktigt eftersom om ett frågeformulär är för långt så finns risk att den som svarar inte lägger ner tillräckligt med tid på frågorna, vilket negativt påverkar svarens kvalitet. Vid undersökningar i kontrollrum vid kärnkraftverk är dock antalet deltagare så litet att faktoranalytisk ansats ej är tillämpligt. Metaanalys av en större mängd enkäter från ett flertal valideringar borde dock kunna genomföras för utveckling och verifiering av enkätinstrument.

Enkäter kan exempelvis användas för subjektiva självskattningar av indirekta begrepp (se nedan, Avsnitt Mätning av indirekta begrepp). Denna typ av självskattningar kan i allmänhet ge pålitliga mått av både situationsmedvetande och mental arbetsbelastning (Angelborg-Thanderz, 1990; Charlton, 1996a;

Svensson, m.fl., 1997a, 1997b; Svensson, Angelborg-Thanderz & Wilson, 1999; Svensson, & Wilson, 2002; Alfredson, 2001; Berggren, 2000).

Vid integrerad validering används enkäter i stor utsträckning för att samla in data. Både deltagande operatörer och observatörer brukar besvara enkäter. Exempel på enkät utvecklad för att besvara efter genomförda driftsfall vid kontrollrum i kärnkraftverk finns i Eriksson m.fl. (1998, Bilaga M).

Observation

Vid observation ligger fokus vanligen på vad som händer. Att däremot under en observation tränga djupare in i varför något händer är betydligt svårare, beroende på att anledning och innebörd till en individs beteende vanligtvis inte är direkt observerbart (Savioja & Norros, 2008). Vid arbete med datoriserade system där operatören interagerar med skärmar och tangentbord vid sin arbetsplats blir detta dessutom ännu svårare att komma åt. Det vill säga i takt med att systemen digitaliseras kommer operatörernas handlingar i allt större utsträckning att bestå av komplexa kognitiva processer, snarare än enkla yttre observerbara beteenden.

Vid observation av traditionella kontrollrum där operatören interagerade med reglage och mätare på väggen var det enklare observera operatörens yttre beteende. Problemet att tolka anledning och innebörd till beteende fanns dock även då.

Tänka högt

Tänka högt är en metod 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. Detta kan vara ett sätt att komma åt användarnas kognitiva processer under arbetet (se t.ex. Ericsson & Simon, 1984; Alfredson m.fl., 2004b; Hackos & Redish, 1998; Nielsen, 1993).

Mätning av indirekta begrepp

Inom beteendevetenskaplig forskning är det vanligt att man inte direkt kan mäta de begrepp man vill komma åt för att kunna säga något om människans interaktion med ett system. Eftersom det är svårt att fastställa vad prestation faktiskt innebär, vid styrning av komplexa processer, tvingas man ofta att mäta sådant som antas påverka eller vara en förutsättning för prestation. Sådana mått kallas indirekta mått och är typiskt mått som situationsförståelse eller mental arbetsbelastning. Hit hör också begrepp som intelligens, mentala modeller, gemensam lägesuppfattning, tillit, rädsla, oro osv. Denna typ av mått kan mycket väl användas, men om endast indirekta mått används är det mycket svårt att avgöra om en validering är meningsfull eller ej.

Två av de vanligaste indirekta begreppen i detta sammanhang är situationsmedvetande och mental arbetsbelastning.

Situationsmedvetande

Situationsmedvetande (SA) har sitt ursprung i det amerikanska flygvapnet, där en arbetsgrupp 1991 formulerade följande definition av situationsmedvetande: ”pilotens ständiga perception av sig själv och flygplanet i relation till den dynamiska omgivningen bestående av flygning, hot, och uppdrag, och förmåga att förutsäga och sedan utföra uppgifter baserade på dessa perceptioner” (Waag & Bell, 1997). Enligt Endsley (1996a) har situationsmedvetande senare även utvecklats till ett betydelsefullt begrepp inom andra områden där människor hanterar komplexa dynamiska system. Endsley (1997) beskriver situationsmedvetande som en inre föreställning av den aktuella situationen och en drivkraft i beslutsprocessen. Vidare uppdelar Endsley (1997) situationsmedvetande i följande tre nivåer:

- *Nivå 1* – Att uppfatta kritiska faktorer i omgivningen.
- *Nivå 2* – Att förstå vad dessa faktorer innebär, i synnerhet när de samordnas med varandra, i förhållande till den egna målsättningen.
- *Nivå 3* – Att förstå vad som kommer att hända med systemet i den närmaste framtiden.

Vid interaktion med automatiserade system kan operatörens roll reduceras till systemövervakning. När människors arbetsuppgifter reduceras till systemövervakning får de vanligen sämre situationsmedvetande och agerar långsammare när de väl behöver ingripa (Endsley, 1996a). Detta beror, enligt Endsley bland annat på att operatören vid ingripandet måste bestämma systemets tillstånd för att förstå vad som händer. Ett minimikrav på ett automatiserat system är därför att det ger operatören information om vilket tillstånd det befinner sig i, och att övervakade parametrar alltid presenteras tydligt och lättolkat (Endsley, 1996a).

Exempel på mätmetoder för värdering av situationsmedvetande:

- *SAGAT (Situation Awareness Global Assessment Technique)* (Endsley, 1988) mäter situationsmedvetande genom att försökspersoner, när en simulering fryses, får besvara slumpmässigt utvalda frågor om simuleringens innehåll. Nackdelar är att frysningarna kan ha en störande effekt på arbetet och därmed påverka försökspersonens SA. Dessutom kan frågorna styra försökspersonens SA genom att han/hon i fortsättningen koncentrerar sig på det som efterfrågades (Castor m.fl., 2003).
- *SACRI (Situation Awareness Control Room Inventory)* (Hogg, Follesø, Volden, Torralba, 1995) är en anpassning av SAGAT för användning vid mätning av SA i kontrollrum vid kärnkraftverk. Dvs. när simuleringen fryses får operatörerna besvara frågor om status på olika systemparametrar.
- *Cooper Harper skala specialanpassad för situationsmedvetande* är en FOI (Totalförsvarets Forskningsinstitut) modifierad version av Cooper Harper skalan för att mäta situationsmedvetande. Skalan har utarbetats för att användas i det europeiska projektet VINTHEC (Berggren, 2000).
- *SART (Situation Awareness Rating Technique)* (Taylor, 1990) mäter situationsmedvetande genom att försökspersoner får skatta krav på

uppmärksamhet, tillgång till uppmärksamhet och förståelse, vilket antas vara tre elementära dimensioner av situationsmedvetande.

- *CC-SART (Cognitive Compatibility-SART)* (Taylor, 1995) är en vidareutveckling av SART som även tar hänsyn till kognitiv kompatibilitet (överensstämmelse).
- *Självskattad SA* innebär att försökspersoner helt enkelt skattar hur bra SA de har på en skala 1 – 7. Måttet har god korrelation med övriga metoder. Ett problem är dock att försökspersoner som inte har förstått situationen felaktigt kan tro att de har bra SA (Berggren, 2005).

Det finns vissa problem med mätningar av situationsmedvetande. SACRI och liknande metoder som SAGAT innebär att simuleringen fryses och deltagarna får svara på frågor om sin uppfattning av olika parametrar i simuleringen. Dessa avbrott kan ha en störande effekt på operatörernas arbete. Dessutom kan själva avbrottet påverka försökspersonens SA, och därmed resultatet. Förutom detta kan frågorna som ställs påverka vad operatören fokuserar på under resten av försöket. Om exempelvis en fråga ställs om en viss parameter, kan det leda till att operatören under resten av försöket lägger omotiverat mycket av sina mentala resurser (och SA) på att hålla reda på utvecklingen av just den parametern, vilket kan leda till försämrad prestation och SA avseende viktigare parametrar. Detta kan i sin tur leda till ett annorlunda beteende (t.ex. Castor m.fl., 2003). Så kallade självskattningar av SA, vilket innebär att operatören helt enkelt tillfrågas om hur bra SA han/hon har visat sig ha god korrelation med övriga metoder. Ett problem är dock att försökspersoner som inte har förstått situationen, felaktigt kan tro att de har bra SA (Berggren, 2005).

Situationsmedvetande kan även mätas med psykofysiologiska metoder (se nedan Avsnitt Dynamiska mätmetoder – psykofysiologiska mätmetoder) och bedömas genom observation.

Mental arbetsbelastning

Mental arbetsbelastning är ett begrepp som kan relateras till skillnaden mellan tillgängliga mentala resurser och de mentala resurser som krävs för att utföra en uppgift (Sanders & McCormick, 1993). Att bedöma mental arbetsbelastning vid systemutveckling är huvudsakligen intressant i tre avseenden (Wickens & Hollands, 2000):

- *Förutsäga arbetsbelastning* – Före systemdesign undersöka olika interaktionsprincipers påverkan på användares förmåga.
- *Bedöma systemets påverkan på användarens arbetsbelastning* – Utgående från prototyp eller system, undersöka hur systemet kan optimeras.
- *Bedöma användarens individuella arbetsbelastning* – För att utvälja bäst lämpade operatörer, eller för träning.

Mental arbetsbelastning kan mätas med både psykofysiologiska metoder och subjektiva självskattningar (se Nählinder, 2009). För subjektiva självskattningar finns ett antal vedertagna metoder:

- *Modifierad Cooper-Harper skala* (Wierwille & Casali, 1983) är en utveckling av den ursprungliga Cooper-Harper skalan (Cooper &

Harper, 1969) som utvecklades för piloter. Förutom psykomotorrelaterad arbetsbelastning inom flyget kan generell arbetsbelastning vid ett flertal andra typer av arbetsuppgifter mätas (Charlton, 1996a).

- *BFRS (BedFord Rating Scale)* har också utvecklats från Cooper Harper skalan. De som studeras gör, enligt Svensson m.fl. (1999), subjektiva bedömningar av mental arbetsbelastning baserat på deras förmåga att slutföra arbetsuppgifter och bedömd mängd kvarvarande mental kapacitet.
- *FOI PPS (Pilot Performance Scale)* har utvecklats på FOI för värdering av mental arbetsbelastning vid militära flyguppdrag. Frågorna besvaras på en bipolar skala från 1 – 7. Metoden utviner information om sex dimensioner av mental arbetsbelastning, vilka extraheras med faktoranalys.
- *NASA-TLX (Task Load Index)* (Hart & Staveland, 1988) är en subjektiv metod där försökspersoner själva skattar följande sex olika dimensioner av sin mentala belastning på en skala från 0–100: mentala krav, fysiska krav, tidspress, prestation, ansträngning och frustration. Beräkningen av den mentala arbetsbelastningen baseras på viktade medelvärden av respektive dimension, där vikten för varje enskild dimension utgår från den studerade personens subjektiva skattning av respektive dimensions betydelse för den specifika situationen (Hart & Staveland, 1988). Genom att använda samma vikt för samtliga dimensioner vinner man dock i enkelhet och tydlighet och förlorar inte mycket i exakthet. Dessutom kan dimensionen prestation utlämnas, med motiveringen att den inte är relaterad till övriga dimensioner (Castor m.fl., 2003; Alfredson, Oskarsson, Castor & Svensson, 2003, 2004b).
- *Självskattat mental arbetsbelastning* innebär att försökspersonerna helt enkelt skattar hur mentalt belastade de känner sig på en skala 1 – 7. Denna metod är enkel och har god korrelation med övriga metoder och används därför ofta på FOI (Nählinger & Berggren, 2002).

Att använda psykofysiologiska mätmetoder för att bedöma mental arbetsbelastning har fördelen att de ger mått som är jämförelsevis kontinuerliga över tiden, och att de inte påverkar prestationen för den primära uppgiften. En nackdel är dock att det ofta krävs att elektroder eller annan mätutrustning fästs på försökspersoner, vilket negativt kan påverka deras fysiska handlingsutrymme och acceptans (Wickens & Hollands, 2000). För vidare beskrivning se nedan Avsnitt Dynamiska mätmetoder – psykofysiologiska mätmetoder).

Dynamiska mätmetoder

Dynamiska mått innebär kontinuerlig mätning över tiden. Dynamiska mått kan därmed beskriva utvecklingen av en händelse, vilket är en stor fördel i dynamiska situationer. Utan dynamiska mått kan man bara ge en beskrivning av läget före och efter försöket. Dynamiska mått kan användas till att koppla prestation till olika skeenden i scenarier, samt att studera samvariation över tid och situation. Detta är synnerligen värdefullt för att identifiera kritiska faser i händelseförlopp som utgör ”flaskhalsar” med avseende på

informationshantering och beslutsfattande (se t.ex. Magnusson & Berggren, 2002).

Psykofysiologiska mätmetoder

Psykofysiologiska mått har fördelen av att de är dynamiska, det vill säga man kan mäta signalen kontinuerligt. När väl utrustningen är på plats och inspelningen är påbörjad, fortsätter registreringen kontinuerligt under testperioden utan att det krävs någon inblandning från varken försöksledare eller försöksdeltagare. Subjektiva skattningar och frågeformulär, å andra sidan, kräver mentala processer från deltagaren och kan endast genomföras vid specifika, diskreta tillfällen.

Det finns en rad olika psykofysiologiska mått som används i olika sammanhang. Oftast mäter de reaktioner på hjärtat (t.ex. puls, pulsvariabilitet, blodflöde, blodtryck, syremättnad i blodet), på ögat (t.ex. ögonrörelser, blickriktning, sackader, blinkfrekvens), på hjärnan (t.ex. EEG, NIRS [Near Infrared Spectroscopy], olika typer av s.k. brain imaging) samt på huden (t.ex. hudkonduktans, yttemperatur). Ibland mäter man också djupare in i kroppen, till exempel hormoner och kroppstemperatur (kärntemperatur).

Vanligtvis används psykofysiologiska metoder för att bedöma mental arbetsbelastning. När den mentala arbetsbelastningen stiger har det visats att situationsmedvetandet sjunker, varför även de psykofysiologiska metoder som används för att mäta mental arbetsbelastning även säger något om situationsmedvetandet (se t.ex. Svensson & Wilsson, 2002; Castor 2009; Nählinder, 2009). Sambandet är dock inte helt linjärt (Endsley, 1996b), varför sådana mått bör kompletteras med subjektiva mätmetoder för detta ändamål (se ovan, Avsnitt Mätning av indirekta begrepp – Situationsmedvetande).

Analys av blickbeteende kan bland annat avslöja om försökspersoner tittar på de kontroller och reglage som de förväntas hålla koll på. Blickbeteendet kan även visa tidsåtgång, samt svårigheter med att hitta relevant information. Mätningar av piloters blickbeteende kan exempelvis användas vid design av cockpit (t.ex. Svensson, Angelborg-Thanderz, Sjöberg & Olsson, 1997; Svensson, Angelborg-Thanderz & van Avermaete, 1997). Blinkningsfrekvens kan användas som mått på mental arbetsbelastning, eftersom både fördröjningstid och blinkduration anses öka med stigande visuella krav i uppgiften.

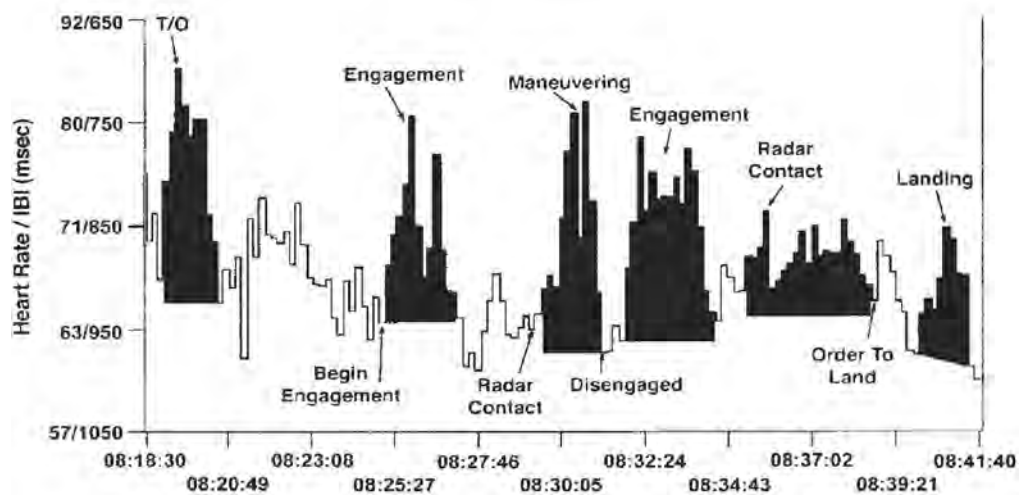
Här ges en kort beskrivning av de vanligaste psykofysiologiska måtten:

- *Hjärtfrekvens* ökar med ökad mental arbetsbelastning (Tsang & Wilson, 1997). Fysisk aktivitet och andningsfrekvens kan dock påverka måttens tillförlitlighet (Svensson m.fl., 1999).
- *Variation i hjärtfrekvens (HRV, Heart Rate Variability)* avtar i allmänhet vid högre mental arbetsbelastning (t.ex. Tsang & Wilson, 1997; Wickens & Hollands 2000). Genom statistisk analys kan man få 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.

- *Ögonrörelser (EOG, Electro-oculography)* kan användas som mått på visuell uppmärksamhet, och beroende på arbetsuppgift och visuella krav som mått på mental arbetsbelastning. Analys av ögonrörelser kan göras genom EOG utrustning som via elektroder mäter aktiviteten hos de muskler som kontrollerar ögonglobens rörelser, eller genom videoinspelning och huvudmonterad utrustning som använder optisk teknik (Castor m.fl., 2003).
- *Ögonblinkningar* har visat sig avta vid ett flertal visuellt krävande uppgifter (Tsang & Wilson, 1997). Ett problem är att frekvensen för ögonblinkningar även kan öka vid snabb avsökning av exempelvis en instrumentpanel, beroende på att en sekvens av informationsinhämtning ofta avslutas med en ögonblinkning (Tsang och Wilson, 1997). Ett annat problem är att vissa forskare anser att kopplingen mellan mental arbetsbelastning och ögonblinkningar är något svag (Krebs, Wingert & Cunningham, 1977; Casali & Wierwille, 1983). Detta trots att empiriska bevis finns för att blinkfrekvensen minskar vid förhöjd uppgiftsbelastning. Det är dock troligt att blinkfrekvens i större utsträckning är användbart vid mätning av trötthet och utmattning (Castor m.fl., 2003; Alfredson, Nählinder & Castor, 2004a).
- *Blickpunktsregistrering (EPOG, Eye Point of Gaze)* innebär dynamisk registrering av vad försökspersonen tittar på. Vid studier av operatörers interaktion med tekniska system kan information om vad på displayer, manöverbord etc. som de tittar på vid respektive fas av olika arbetsuppgifter ge värdefull information om mental arbetsbelastning och situationsmedvetande (Alfredson m.fl., 2004b). Om blicken fixeras längre tid på ett objekt antas det exempelvis innehålla mer information, eller vara mer krävande att förstå (t.ex. VINTHEC, 1997; Wickens & Hollands, 2000).
- *Pupillstorlek* ökar med ökande krav vid ett flertal kognitiva aktiviteter (Wickens & Hollands, 2000). Hänsyn måste dock tas till belysning som även påverkar pupillens storlek (Castor m.fl., 2003).

Vid användning av psykofysiologiska mätmetoder bör utrustning användas som märks och påverkar försökspersonen så lite som möjligt. Han eller hon måste kunna utföra sina uppgifter precis som vanligt, utan att behöva tänka på, eller ens märka av utrustningen. Psykofysiologin har den stora fördelen att den fungerar ungefär likadant på alla människor oavsett språk och kön. Psykofysiologiska mått är ofta korrelerade med skattningar av mental arbetsbelastning och kan därför användas som ett komplement till dessa.

Figur 4 illustrerar hur hjärtfrekvensen hos en stridsflygare förändras under olika delar av ett simulerat flyguppdrag. Varje stapel motsvarar medelvärdet för hjärtfrekvensen under tio sekunder (Svensson & Wilsson, 2002).



Figur 4. Illustration av hur en stridsflygares hjärtfrekvens förändras under olika faser av ett flyguppdrag. Särskilt intressanta perioder är markerade med svart. Varje stapel motsvarar tio sekunder (från Svensson & Wilson, 2002).

GARTEUR (Castor m.fl., 2003) beskriver ett stort antal psykofysiologiska mätmetoder (se ovan). En sammanställning av olika metoder för mätning av blickbeteende och blickpunktsregistrering finns i Alfredson, Nählinder & Castor (2004a). Ett antal studier där psykofysiologisk mätning har använts presenteras i Nählinder (2009).

Nackdelen med psykofysiologisk mätning är att det i regel är nödvändigt att försökspersonerna har mätutrustning på kroppen. Det finns dock numera utrustning som är smidig att bära och därmed ej är speciellt störande. Bland annat finns utrustning för registrering av ögonrörelser som endast använder kameror, vilket innebär att försökspersonerna inte behöver bära någon utrustning.

Systemloggar

Systemloggning innebär att systemet kontinuerligt registrerar status för vissa utvalda parametrar. Det kan också innebära att användarens interaktioner med systemet registreras. Detta innebär att man kan skapa inspelningar av systemet och/eller användarens interaktion med systemet. Idag är det ofta tekniskt möjligt att kontinuerligt logga tusentals variabler per delsystem. Dessa inspelningar måste dock göras meningsfulla för att vara användbara.

Loggning av militära flygförarens interaktioner under flygning görs exempelvis regelmässigt i tränings- och utbildningssyften, vilket även kan ha betydelse för systemutveckling. Fördelen med loggningen är att det vid diskussioner efter uppdrag går att se exakt vad var och en har gjort under flygningen. FOI har exempelvis genomfört studier där UTB (utbildningsbandspelare) har använts för datainhämtning, till exempel registrering av vad flygföraren gjort vid en viss

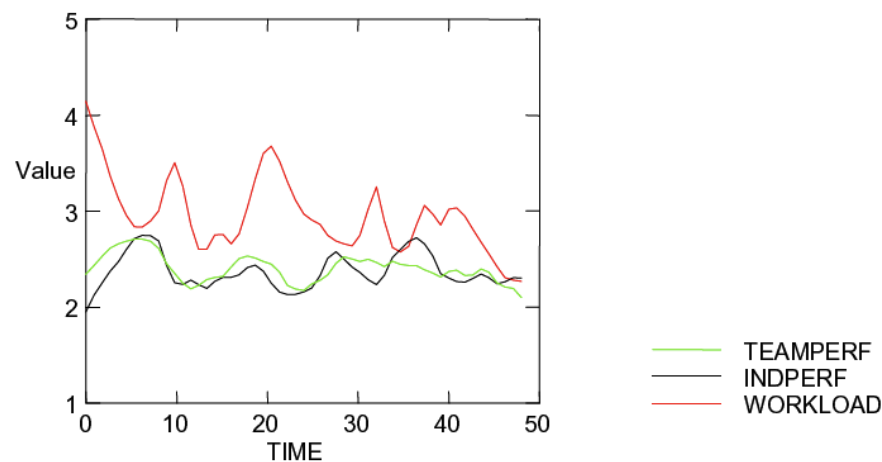
manöver, vilket sedan korrelerats med psykofysiologiska mätdata (Magnusson, 2002).

Loggning kan ge både kvalitativ och kvantitativ data. Loggning av en operatörs/försökspersons interaktioner med ett system kan ge kvalitativ information, motsvarande observationsdata. Om däremot det som loggas är antal tryckningar på en viss knapp, eller antal användningar av en viss funktion, så får man kvantitativ mätdata. Kvantitativ data är enkelt att analysera med statistiska metoder, men har nackdelen att det ofta kan vara svårt att tolka betydelsen, av exempelvis ett visst antal knapptryckningar.

Kvasidynamiska mått

Om man inte kan använda fullständigt dynamiska mått, kan man ofta använda kvasidynamiska mått. Detta innebär att mätningar görs med jämna mellanrum under ett försök. Kvasidynamiska mätningar ger därmed en approximativ beskrivning av dynamiken och kan därför användas för att urskilja hur operatörernas arbetsbelastning förändras över tiden. I Figur 5 visas resultat från på en kvasidynamisk mätning av individuell prestation och mental arbetsbelastning. Mätningarna gjordes genom att operatörerna besvarade en elektronisk enkät på en PDA (Personal Digital Assistant) (Svensson, Rencrantz, Lindoff & Norlander, 2006).

Operatörerna besvarade frågorna var tionde minut. Frekvensen av svartillfällen visade sig tillräckligt frekvent för att beskriva dynamiken i förloppen utan att påtagligt störa operatörerna i deras arbete. Scenariot varade i cirka två och en halv dag och varje operatör har besvarat de sex frågorna vid cirka 45 tillfällen.



Figur 5. Förändring av Arbetsbelastning (WORKLOAD), individuell- (INDPERF) och team-prestation (TEAMPERF) som funktion av tid och situationens komplexitet (anpassad från Svensson m.fl., 2006).

Holistisk ansats

Största nyttan finns troligen i att kombinera kvantitativa och kvalitativa ansatser, och därmed dra nytta av båda synsättens fördelar. Kvantitativa data kan ge tydliga svar på specifika frågor, kvalitativa data kan förklara sådant som inte på förhand kan formuleras som en fråga eller mätas på ett enkelt sätt. Detta arbetssätt förespråkas både i Rollenhagen m.fl. (1998) och i Norros m.fl. (2009).

Oavsett vilken typ av metodologisk ansats som används, så är omsorgsfull planering och design nödvändig innan studier genomförs. För att data ska bli användbara till statistiska analyser måste en studie designas med utgångspunkt från vilka resultat som ska visas och vilka statistiska analyser som ska genomföras. Detta har bland annat stor betydelse för vilka variabler som ska konstanthållas och manipuleras, och för urval av försökspersoner och grupper så att resultaten blir representativa. Att samla in data utan förberedelser och omsorgsfull design av studien leder därför ofta till data som är svåra att analysera, och resultat med låg reliabilitet och validitet.

En strikt uppdelning mellan kvantitativa och kvalitativa metoder är inte möjlig. Dessutom skiljer man på objektiva och subjektiva utvärderingsmetoder, där subjektiva metoder vanligen ger kvalitativa data, men även kan ge kvantitativa data, exempelvis vid ifyllande av numeriska skattningsskalor, eller när en observatör registrerar antal händelser av en viss typ. Dessutom skiljer man på experimentella och utvärderande metoder, där experimentella metoder huvudsakligen avser kvantitativa laboratorieexperiment för hypotestestning, medan utvärderande metoder främst avser system- och prototyputvärderingar, som i större utsträckning använder kvalitativa metoder. Att endast skilja på kvantitativa och kvalitativa metoder är därför ett något förenklat synsätt. Mer illustrativt vore istället att placera olika metoder i ett tredimensionellt rum med följande kategoriaxlar: x) kvantitativ – kvalitativ y), objektiv – subjektiv, och z) experimentell – utvärderande.

För att öka resultatens reliabilitet är det en fördel om både kvantitativa och kvalitativa metoder används samtidigt. Det vill säga om metoderna pekar på samma resultat är det sannolikt mer reliabelt. Samtidig användning av olika metoder för att se om de visar på samma resultat, eller problem vid systemutvärdering brukar kallas triangulering.

Integrerade valideringar för kontrollrum inom kärnkraftsområdet genomförs i regel i simulatorer. Detta innebär goda möjligheter att konstanthålla ett stort antal variabler, samtidigt som arbetssituationen kan göras väldigt lik den i naturlig arbetsmiljö i ett riktigt system. Detta gör att man vid väl genomförda simulatorexperiment kan förena fördelarna med både kvantitativa och kvalitativa metoder.

Statistisk analys vid integrerad validering

Vid statistiska analys av data från integrerad validering finns ett stort antal problem att hantera. Ofta är det relativt få försökspersoner, vilket medför få datapunkter. Detta gör att inferensstatistiska analysmetoder kräver stora skillnader för att visa signifikanta resultat.

Viss variation av försökspersoner kan inte undvikas. Som påpekats tidigare i rapporten har det vanligen gått en viss tid mellan t.ex. en baslinjemätning och validering, varför förutsättningarna för mätningar inte är identiska. Som också har påpekats är så kallade takeffekter ett problem. Försökspersonerna ofta har mycket stor vana – varför det ofta är svårt att hitta statistiskt signifikanta (eller praktiskt signifikanta) skillnader vid analys

För statistisk bearbetning av data insamlade under mätsituationer med lägre kontroll, exempelvis observationsstudier, har speciella analysmetoder utvecklats, som kausala modeller, väganalys (path analysis), strukturella ekvationsmodeller (Cook & Campbell, 1979) och linjära strukturella ekvationsmodeller (LISREL) (Jöreskog & Sörbom, 1993). Vid studier av flygförarens mentala arbetsbelastning, situationsmedvetande och prestation har FOI vid ett flertal studier använt denna typ av metoder för att analysera bland annat prestationsdata, psykofysiologiska mätdata och subjektiva självskattningar av situationsmedvetande och mental arbetsbelastning (t.ex., Angelborg-Thanderz, 1990; Svensson m.fl., 1997; Svensson m.fl., 1999).

Här presenteras ett antal alternativa statistiska metoder för hantering av osäkerheter och att studera orsakssamband. Syftet med att ta upp metoderna är inte att vi tycker att de ska användas i stället för de analyser som brukar användas vid integrerad validering. Syftet är snarast att skapa diskussion om alternativa metoder vid statistisk analys. Exempelvis om det finns behov av att komplettera de analysmetoder som brukar användas.

Kovariansanalys

Om det går viss tid mellan baslinjemätning och integrerad validering kommer det oundvikligen att finnas skillnader mellan försökspersonerna. Vissa har nyanställts, andra har slutat, eller befordrats till en ny position i kontrollrummet. Dessutom har operatörerna ackumulerat erfarenheter genom att arbeta i kontrollrummet under den tid som gått mellan mättillfällena. Att genom kovariansanalys statistiskt kontrollera för dessa skillnader är därför önskvärt. Kovariansanalys är i princip likvärdig med vanlig variansanalys, men med tillägget att en, eller flera variabler används som kovariat för att justera för olikheter mellan studerade grupper. Detta görs genom att inomgruppsvariansen, även kallad felvariansen, reduceras med den varians som kan hänföras till kovariatet.

Kovariansanalys kan således kompensera för brister i experimentell kontroll genom ökad statistisk kontroll. Detta är särskilt värdefullt vid fältstudier då det ofta inte finns möjlighet att slumpmässigt skapa helt jämförbara grupper för

olika betingelser. Då kan egenskaper hos gruppmedlemmar, eller testsituationer, användas som kovariat och därmed justera variansanalysen för gruppskillnader mellan betingelserna.

Det finns dock vissa krav för att kovariansanalys ska kunna genomföras. En variabel som användas som kovariat ska vara normalfördelad och innehålla data på intervall- eller kvotskala, dvs. samma krav som för beroende variabler i variansanalys. För att ett kovariat ska få användas finns även kravet att kovariatet är linjärt relaterat till den beroende variabeln (att det finns ett linjärt samband mellan dem, t.ex. att kovariatet ökar, eller minskar när den beroende variabeln ökar). Dessutom finns kravet att regressionslinjerna för de studerade grupperna har samma lutning (homogenitet hos regressionen).

I praktiken är det ofta svårt att hitta variabler som uppfyller kravet för kovariansanalys. Men problemet med att jämföra försökspersonernas prestation vid baslinjemätning och integrerad validering så att säga är inbyggt i studiens design, och därmed uppenbart från början. Den som genomför studien kan därför redan när baslinjemätning genomförs överväga om det finns lämpliga data som skulle kunna användas som kovariat. Exempelvis, om data samlas in om hur länge respektive operatör har tjänstgjort på nuvarande position kan man undersöka om det finns någon relation mellan tjänstetid och de beroende mått som används vid respektive mättillfälle. Eventuellt går det också att utforma någon typ av prestationstest som avseglar allmän skicklighet i kontrollrummet.

Metodik för att hitta orsakssamband

Multivariata statistiska tekniker är mycket viktiga verktyg för att analysera multipla orsakssamband, där ett flertal faktorer påverkar varandra. De möjliggör därmed förenklade beskrivningar av komplexa samband, exempelvis mellan fysiologiska och psykologiska reaktioner. Dessa metoder är nödvändiga vid modellering av mänskligt beteende och en förutsättning för att förstå komplexa orsakssamband.

Metoderna bygger på korrelationsstatistik, dvs. det linjära sambandet mellan variabler. Metoderna ger därmed ett mått på relationen mellan variablerna, i form av mängd förklarad varians (hur mycket av variansen de har gemensam). Detta gör dessa tekniker mer kraftfulla än traditionella metoder, som variansanalys och t-test, vilka bygger på jämförelser av gruppmedelvärden. Metoderna kallas därför ofta för "andra generationens" statistiska metoder.

Korrelationsanalys

Korrelationsanalys används för att påvisa samband mellan variabler. Korrelationsanalys är därför synnerligen användbart vid fältmässiga studier i naturlig miljö med begränsad möjlighet att manipulera och konstanthålla variabler. Storleken på ett korrelations samband visas genom korrelationskoefficientens värde, vilket kan variera mellan +1 och -1. Om korrelationskoefficienten är lika med ± 1 finns ett fullständigt samband. I praktiken förekommer dock i stort sett aldrig så starka samband.

Korrelations sambandet säger inte något om orsakssambandet mellan variablerna. Det vill säga om det finns ett korrelations samband mellan A och B kan orsakssambandet lika gärna vara $A \rightarrow B$ som $B \rightarrow A$. Sambandet behöver inte ens orsakas av någon av de uppmätta variablerna A eller B, utan det kan mycket väl vara en tredje oidentifierad variabel C som påverkar båda A och B.

Korrelationskoefficientens kvadrerade värde (r^2) ger ett mått på hur stor andel av variansen som är gemensam för de båda variablerna. Det vill säga om korrelationen är $r = 0,80$ så har variablerna 64 procent gemensam varians ($0,80^2 = 0,64$).

Datareduktion

Faktoranalys är en analytisk metod som används för att strukturera och reducera ett stort antal relaterade manifesterade (manifest variables) (de variabler man har uppmätt) till ett mindre antal dolda "bakomliggande" variabler (latent variables) eller faktorer. Analysen baseras på samvarians (kovarians) mellan variablerna. Faktoranalys kan vara antingen explorativ eller hypotestestande. Vid explorativ Faktoranalys utforskas, i princip förutsättningslöst, strukturen bakom ett större antal variabler. Vid hypotestestande Faktoranalys definieras före analysen ett antal faktorer och de variabler som antas tillhöra respektive faktor.

Faktoranalys är, utan jämförelse, den mest använda metoden för datareduktion, och utgör grund för besläktade metoder som, *principalkomponentanalys*, *klusteranalys* och *multidimensionell skalning*. Något förenklat kan man säga att faktoranalys visar strukturen hos sambanden (korrelationerna) mellan ett stort antal variabler genom att definiera och indela dessa i ett mindre antal bakomliggande dimensioner eller faktorer. De grupperingar som blir kvar är unika och har högre validitet än de ingående (ursprungliga) variablerna. Till exempel kan det visa sig att vissa olika variabler som man samlat in under en datainsamling i själva verket är mycket snarlika. De kan då med fördel slås ihop till en (övergripande) variabel. Denna övergripande variabel är ofta mer rättvisande och stabilare än de ingående variablerna.

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. Metoden utgör ett kraftfullt verktyg i analys av olika variablers inbördes förhållanden och den ger uppslag i modelleringsarbetet.

Faktoranalys har exempelvis används av FOI analys av en militär ledningsövning med syftet att studera förutsättningarna för interaktion och samverkan (Rencrantz, Lindoff, Svensson, Norlander & Berggren, 2006). Vid en pilotstudie användes kvasidynamisk mätning, genom PDA-enkät med sex frågor om mental arbetsbelastning, situationens komplexitet, situationsmedvetande, operatörens prestation koordinering inom teamet och fördelning av information. Med faktoranalys skapades tre faktorer:

- *Mental arbetsbelastning* – bestående av variablerna mental arbetsbelastning och situationens komplexitet
- *Individuell prestation* – bestående av variablerna individuell prestation och situationsmedvetande
- *Teamets prestation* - bestående av variablerna koordinering inom teamet och fördelning av information)

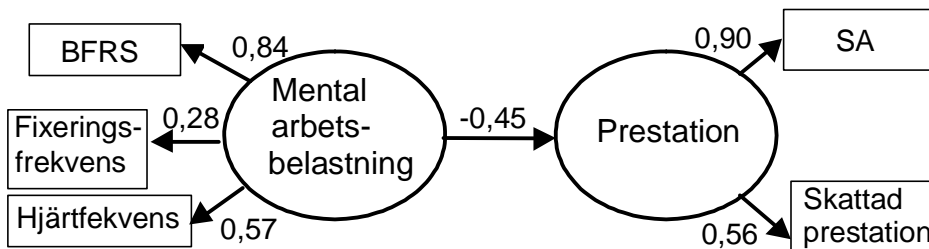
Observera de två första faktorerna har döpts till samma namn som en av sina ingående variabler. Detta kan möjligen verka förvirrande, men beror på att dessa variablers namn bedömdes som bäst passande för respektive övergripande faktor. Det kausala sambandet mellan dessa faktorer analyserades med linjär strukturell ekvationsmodellering (LISREL) (se nedan, Avsnitt Kausalanalys – LISREL)

För vidare läsning om faktoranalys, se exempelvis Svensson (2003) eller Pett, Lackey & Sullivan (2003).

Kausalanalys – LISREL

Kausalanalyser, som strukturell ekvationsmodellering och linjär strukturell ekvationsmodellering (LISREL), används för att förstå hur olika variabler påverkar varandra. LISREL är en mycket användbar metod för att bekräfta och dra generella slutsatser om strukturer och orsakssamband i 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 manifesta (synliga) och latent (bakomliggande) variabler.

För att illustrera användningen av LISREL ges ett par exempel från forskning som FOI genomfört. Figur 6 visar en LISREL-modell av sambandet mellan flygförarens Mentala arbetsbelastning och prestation, baserat på skattad mental arbetsbelastning (Bedford Rating Scale) uppmätt fixeringsfrekvens, hjärtfrekvens, samt skattning av situationsmedvetande och prestation.

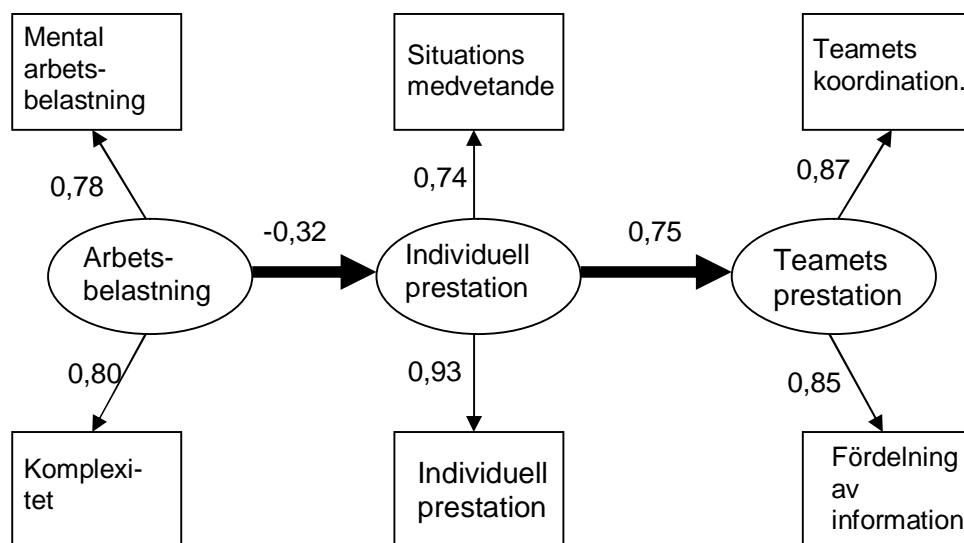


Figur 6. Strukturell ekvationsmodellering (LISREL) av sambandet mellan flygförarens mentala arbetsbelastning och prestation (anpassad från Svensson & Wilson, 2002).

Som Figur 6 visar så är skattad arbetsbelastning, fixerings- och hjärtfrekvens signifikanta markörer (manifesta variabler) för den övergripande faktorn mental

arbetsbelastning (latent variabel). Det innebär att ökad skattning av mental arbetsbelastning, samt högre mått på fixerings- och hjärtfrekvens, visar på ökad generell mental arbetsbelastning. På motsvarande sätt är skattningar av situationsmedvetande och prestation signifikanta markörer för den latent variabeln prestation. Särskilt intressant är att den latent variabeln mental arbetsbelastning utgörs av två psykofysiologiska variabler (fixerings- och hjärtfrekvens) och en psykologisk variabel (BFRS). Det negativa sambandet mellan de latent variablerna visar att ökad mental arbetsbelastning leder till försämrad prestation (Svensson & Wilson, 2002).

Figur 7 visar en LISREL modell av sambandet mellan arbetsbelastning, individuell prestation och teamets prestation i en militär stab. Analysen är gjord på de faktorer som redovisats ovan (se Avsnitt Datareduktion). Som framgår av Figur 7 påverkar arbetsbelastningen individens prestation negativt, dvs. en ökad mental arbetsbelastning försämrar individens prestation. Den individuella prestationen påverkar, i sin tur, gruppens eller teamets prestation positivt, dvs. om individen presterar bra så påverkar det teamets prestation positivt. Det kvadrerade värdet av respektive samband ger mängd förklarad varians, dvs. individuell prestation förklarar 56 procent av teamets prestation ($0,75^2 \approx 0,56$) (Rencrantz m.fl., 2006).



Figur 7. Strukturell ekvationsmodellering (LISREL) av sambandet mellan operatörernas mentala arbetsbelastning individuella prestation och teamets prestation (anpassad från Rencrantz m.fl., 2006).

FOI har även visat att Mental arbetsbelastning påverkar situationsmedvetandet och att situationsmedvetandet, i sin tur, påverkar prestationen (se t.ex. Svensson, 2003, 2004). För vidare läsning och för referenser kring datareduktion och modellering, se t.ex. Svensson (2003, 2004).

3. Fokusgrupp

I projektet genomfördes en så kallad fokusgrupp med deltagare från samtliga kärnkraftsanläggningar i Sverige. Syftet med fokusgruppen var att fånga upp relevant personals tankar och föreställningar om integrerad validering samt att få nya uppslag till förbättringar av valideringsprocessen.

Fokusgrupper används till en rad olika saker, från grundforskning inom samhällsvetenskaperna till marknadsföringsundersökningar. Kortfattat kan en fokusgrupp beskrivas som en modererad gruppdiskussion där ett antal deltagare förväntas diskutera ett eller flera ämnen. Gruppens sammansättning beror på frågeställningen som ligger bakom fokusgruppen. En viktig skillnad från en intervju eller en gruppintervju är att metoden förespråkar just diskussion snarare än svar på givna frågor. Moderatorns uppgift är att skapa diskussionsunderlag snarare än att ställa frågor. För att göra det har moderatorn ett i förväg förberett "stimulimaterial", vilket kan variera kraftigt beroende på vilket område som ska diskuteras. I en mycket strukturerad fokusgrupp kan det vara ett frågebatteri, i en helt ostrukturerad kan det vara ett föremål eller helt enkelt en uppmaning att diskutera en företeelse. Förhoppningen är att kunna studera "människors föreställningar, attityder och värderingar inför ett visst ämne" (Wibeck, 2000, sid. 10). Normalt sett spelas diskussionen in, transkriberas och analyseras i efterhand. En fokusgrupp består typiskt av fyra till åtta personer och pågår en till två timmar.

Fokusgrupper har använts inom forskningen sedan 1930-talet. Efter andra världskriget minskade det akademiska intresset för fokusgruppen som metod, även om den levde kvar som en viktig metod inom marknadsundersökningar. Dock har metoden åter kommit till bruk inom akademiska kretsar sedan mitten på 1980-talet. Ett flertal böcker och publikationer har sedan dess producerats samt ett stort antal forskningsprojekt (se t.ex. Basch, DeCiccio & Malfetti, 1989; Kitzinger, 1994; Agar & McDonald, 1995; Kidd & Parshall, 2000; Bloor, Frankland, Thomas & Robson, 2001; Granström, 2002; Hammarlund, Lundgren & Nyström, 2008).

Forskningsprocessen kan grovt delas upp i fyra steg: *Planering – Rekrytering, Diskussionsledning* och *Analys* (Wibeck, 2000). I denna studie har planeringen påbörjats under hösten 2009 och har bestått i en dialog mellan FOI och SSM rörande syftet med studien, planering av underlaget (stimulimaterialet) på FOI. Rekryteringen har skett från FOI, baserat på rekommendationer av lämpliga personer från SSM. Diskussionsledningen utfördes av FOI:s personal. Analys har skett i samverkan mellan FOI:s personal och Natalia Gonzales, Linköpings Universitet.

Metod

Fokusgrupper kan användas för en mängd olika syften. Vanligast är troligen att använda fokusgruppen för att försöka fånga attityder, åsikter och värderingar. Fokusgrupper kan dock även användas för att studera social interaktion i

grupper, samkonstruktion av mening eller t.o.m. för att tydliggöra konflikter kring vissa ämnen. För det första är dock fokusgruppen en forskningsteknik (Wibeck, 2000), det vill säga en metod som används för att besvara en forskningsfråga av någon typ. Gruppdiskussionen uppstår på forskarens initiativ till skillnad från spontana diskussioner (Wibeck, 2000; Morgan, 1996) och styrs i en viss riktning.

I detta projekt användes fokusgruppen främst för att kunna studera tillståndsinnehavarnas syn på valideringsprocessen, dess syfte och dess faktiska genomslag i verksamheten på anläggningarna. En viktig anledning till att vi valt att inte använda intervjuer eller gruppintervjuer är att den frågeställning som vi har utgått från i stort sett varit explorativ. Syftet är att finna nya perspektiv och innovativa grepp rörande valideringsprocessen. Genom att använda fokusgrupper och därmed inte medvetet lägga frågeställningar på våra informanter hoppas vi på att få tillgång till de möjligheter och problem rörande valideringsprocessen som informanterna själva ser – inte de vi antar finns där. Fokusgruppen ger sådana möjligheter, emedan en traditionell intervju lätt leder informanten/informanterna in på ett givet spår.

Vidare är ämnet som sådant svårt att fånga ”i flykten”. Valideringar av kontrollrum sker inte ofta, är i regel utsträckta över en viss tid och utförs på olika sätt på olika anläggningar. Fokusgruppen, där representanter för de olika tillståndshavarna samlas, är därför en lämplig metod för att explicitgöra tankar om valideringsprocessen och att samtidigt få en chans att diskutera ämnet med några av de få andra intressenter som finns i landet. Wibeck (2000) påpekar att sådana diskussioner ofta uppskattas av informanterna just eftersom de får en möjlighet att utbyta erfarenheter och på så sätt också kan få nya uppslag kring det berörda ämnet. Morgan (1998, citerad i Wibeck 2000) menar att fokusgruppsmetoden är lämplig i fyra fall:

- 1.) När det finns stora olikheter mellan människor
- 2.) När handlande och motivation ska undersökas
- 3.) När olikheter ska förstås
- 4.) När det finns behov av en vänlig och respektfull metod

I denna studie är det svårt att säga att informanterna uppvisar stora olikheter i den bemärkelsen att alla har ungefär samma bakgrund och samma roller på de respektive anläggningarna. Däremot gör vi antagandet att det finns olikheter i hur genomförandet av en ny kontrollrumsimplementering och validering går till. Samtliga anläggningar har varit i drift under lång tid och utvecklat sina egna praktiker och kulturer, även om de i grunden styrs av samma regelverk. Det har sedan projektets början varit en uttalad del av frågeställningen att belysa hur de olika anläggningarna skiljer sig när det gäller dessa processer.

På samma sätt är det klart att det är just handlande och motivation för detta handlande som ska undersökas – informanternas motivation eller rationalitet för arbetet kring valideringen är just det som vi önska belysa. Vi vill också finna och förstå dessa olikheter genom att studera informanternas egna redogörelser

och attityder till dessa. Populationen av informanter är samtidigt väldigt liten, varför det är nödvändigt med en ”vänlig och respektfull” metod som inte avskräcker informanterna. En fri diskussion kring ett ämne som berör samtliga deltagare bör vara acceptabel för samtliga deltagare.

Genomförande

Fokusgruppen genomfördes i ett konferensrum på Arlanda flygplats. Fördelen med detta är att det är ”Neutral” mark. Om fokusgruppen skulle ha förlagts till någon av anläggningarna, FOI eller SSM finns det en risk att en ”maktförskjutning” skulle kunna uppstå. På en anläggning kan den som är ”hemma” där indirekt eller direkt ges en roll som ”värd” vilket kan ge en obalans i diskussionerna. På FOI eller SSM finns en risk att deltagarna upplever sig som ”försökspersoner” eller forskningsobjekt i högre grad än vad de annars skulle göra.

I fokusgruppen användes ett antal påstående och frågor som diskussionsledaren initierade diskussionerna med och kunde tillgripa i händelse av att diskussionen avklingade. Ämnen rangordnades också av forskarna redan under planeringsfasen. Denna ansats kan sägas vara i linje med det av Wibeck (2000) föreslagna sättet att jobba med ”breda områden”

Datinsamlingen bedrevs i detta fall genom inspelning av diskussionen (endast ljudet) på digital inspelare samt anteckningar som gjordes av assistenterna.

Fokusgruppen tog totalt ca 3 timmar, med två korta pauser.

Analys

Den inspelade diskussionen transkriberades i enlighet med Linell nivå tre (Linell, 1994). Linell nivå tre syftar endast till att återspegla innehållet och omfattar därför inte exakta angivelser av pauser, deras längd, överlappande tal eller annan detaljerad information rörande kvalitativa aspekter av själva samtalet. I de fall det varit svårt att tyda vad som sades i det inspelade materialet har notationen [svårtytt?] använts. Det bör poängteras att talat språk som renskrivs kan upplevas forcerat eller upphugget. Detta är helt naturligt och kännetecknar i stort sett all talad interaktion som renskrivs. Hela transkriptionen har gjorts med hjälp av mjukvaran Transcriber vilken distribueras gratis från <http://trans.sourceforge.net/>.

Fältanteckningarna som fördes har renskrivits, men har i övrigt inte bearbetats.

Därefter utfördes en innehållsanalys på materialet oberoende av två personer. Syftet med en innehållsanalys är att utifrån de frågeställningar som finns i studien koda materialet, dela upp det i enheter och sedan försöka hitta trender och mönster (Strauss & Corbin, 1998; Wibeck, 2000). Denna analys bedrivs enligt att antal i förväg definierade steg, nämligen grov kodning, och förfining/mätning och utvecklande av mer generella analytiska ramverk. I denna studie har vi framförallt fokuserat på de första två stegen då syftet inte har

varit att skapa nya ramverk. Istället har vi syftat till att lyfta fram centrala kategorier i materialet. Därtill utgår vi i detta fall från en s.k. huvudkategori, nämligen integrerad validering – denna är given i syftet med fokusgrupperna och behöver inte identifieras i materialet. Vad som däremot är syftet är att lyfta fram anläggningsrepresentanternas perspektiv på processens, samt eventuella nya aspekter på integrerad validering. I analysen har vi utgått från de frågor som låg till grund för fokusgruppsdiskussionen, men även skapat nya kategorier för sådant som inte förekom i frågorna. I fokusgrupper av mer strukturerad karaktär sammanfaller diskussionsämnena givetvis mer med frågorna i intervjuguiden och gör då att forskaren får svar på sina forskningsfrågor. Dock kan ett strukturerat tillvägagångssätt även belysa ämnen som inte tagits upp, och därmed exempelvis indikera att ämnet är känsligt, eller ämnen som introducerats av deltagarna själva.

Deltagare

I fokusgruppen deltog fyra personer (män) från olika kärnkraftverk i Sverige. Samtliga personer har arbetat på de respektive anläggningarna i minst tio år och har arbetat/arbetar med människa-systemaspekter av valideringsfrågor. Tre av de fyra deltagare har en bakgrund som stationstekniker och de har arbetat sig igenom diverse tjänster i kontrollrummet. Vid diskussionstillfället fanns också en moderator och två observatörer.

Resultat

Nedan följer resultatet från fokusgruppsdiskussionen. De olika ämnena presenteras ej i strikt kronologisk ordning då deltagarna ofta hoppade mellan ämnen. Däremot är yttrandena inom respektive ämne ordnade kronologiskt. Varje excerpt (utdrag) är numrerat efter sin förekomst i originaltranskriptionen samt kopplat till respektive talare. Notationen är som följer:

Deltagare från kärnkraftverken: I1 – I4
Moderator: M
Observatörer: O1 – O2

Excerpten (utdragen) från fokusgruppen refereras med ett nr kopplat till varje yttrande vilket i sin tur kopplas till talaren enligt följande:

Radnr.talare: Excerpt

Före varje excerpt finns en kort förklaring. Varje avsnitt avslutas med en kort diskussion om de ämnen som berörts utifrån rapportens syfte och frågeställning.

Referensdokument och begreppsdefinitioner

Kärnan i många av diskussionerna i fokusgruppen handlade om att informera varandra om hur ens eget verk arbetar, vilka dokument de arbetar efter och vad begreppen betyder för dem.

11.I1: Så vi har ju en anvisning och eller som vi använder oss av i vår projektmall som heter MTO i anläggningsutvecklingsprocessen och den ser till precis nästan ut som er XXXX och er HFE i projekt och den grundar ju sig helt nästan på NUREG0711

46.I3: Så vi har ju övergripande dokument som talar om vilka vitala funktioner man vill ha i kontrollrumsfunktioner och vad man ska innehålla när det gäller larm och händelser, omhänderta de fasta störningar som finns och så vidare. (...)

49.I2: det jag kallar ISV det är en avslutande validering.

I 11 förtydligar I1 vilket dokument han arbetar efter och tydliggör samtidigt att han vet hur vissa av de andras dokument ser ut. Han nämner även ett NUREG-dokument som alla känner till och till viss del arbetar efter. 46 är ytterligare ett exempel på ett uttalande som syftar till att informera de andra i fokusgruppen om deltagarens riktlinjer i arbetet. Begreppsdefinitioner förekommer vid flera olika tillfällen och I2 berättar för de andra om hur han definierar ISV i 49.

I fokusgruppen diskuterades även svårigheter med nuvarande referensdokument. En del av problematiken tycks ligga i att dokumenten inte är tydliga nog.

89.I3: Det är ingen enkel grej, det är någonting som vi önskar att man skulle göra ett arbete på att få en process kring vår TIGER som är enklare att jobba emot. Den består av för mycket skall och kanske och bör (...)

164.I3: Jag kan nog känna att vi inte själva inte vet det, i XXXX (...) Ja, jag tycker det finns så mycket frågetecken i de här processerna fortfarande.

165.I2: Ja visst är det så, det finns mycket utrymme för tolkning

I 89 uttrycker I3 önskan om ett enklare dokument att arbeta utifrån. De nuvarande riktlinjerna i referensdokumentet tycks vara för olikartade i hans tycke. I 164 och 165 konstaterar I3 och I2 att det finns för mycket tolkningsutrymme i referensdokumenten.

I deltagarnas diskussioner om referensdokument och begreppsdefinitioner ligger också en avsikt i att lära sig mer om de andras arbetsätt och riktlinjer.

259.I4: Jag har en fråga är det liksom ni refererar till guidelines och normer och sådant då har ni inte instruktioner som har uttolkat de? Vår TIGER, den är ju uttolkningar av de här normerna och guidelines

I4 frågar de andra deltagare om hur de arbetar och vad de anser om den nämnda arbetsmetoden.

Diskussion

Arbetet med validering i de olika verken är på många sätt lika, målet och syftet är ofta det samma men tillvägagångssättet är inte alltid det. Deltagarna i fokusgruppen är väl medvetna om detta och de vet att det råder viss skillnad i arbetsmetod mellan verken. De vet också på ett ungefär hur de andra arbetar och men det är tydligt att de inte alltid gör det. Det tycks ligga i alla deltagarnas intresse att lära sig mer om de andras arbetsmetoder.

Samarbete mellan verken

Under fokusgruppsdiskussionerna blev det tydligt att utbytet av erfarenhet var värdefullt för deltagarna. Deltagarna uttryckte att ett enhetligt arbetssätt hos verken skulle kunna vara bra för verksamheten.

91.I1: Det där, ja det är ju där lite det här kommande arbetet som jag ser det, den här workshopen vi kan ha och så vi kan göra valideringar ihop. Alltså på samma sätt på alla block, på alla verk då (...)

I1 lyfter fram ”den här workshopen” och poängterar därmed att gemensamma träffar skulle gynna arbetet och att verken skulle kunna lära av varandras valideringar eller samarbeta vid valideringar.

554.I1: Där kan vi lära av varandra också. Ni var och tittade på, ni fick massa svar när ni var och tittade på XXXX:s simulator

555.I2: Det är så vi får göra på något sätt

I 554 pratar I1 om att verken kan lära av varandra. I just detta fall handlar det om implementering och modifiering av nya system från utomstående leverantör och han får medhåll av I2 i 555.

Frågorna som ställdes under fokusgruppsintervjun gav deltagarna möjligheten att diskutera och jämföra sina arbetsmetoder. Deltagarna poängterade att tydliga gemensamma referensdokument och mål skulle gynna deras arbete. Detta diskuteras vidare även i avsnittet ”Designprocessen” nedan.

528.I2: (...) branschen har ju enats om gemensamma krav på både elektrisk konstruktion och mekanisk konstruktion i så kallade TBE:er och KBE:er, men när det gäller det området vi representerar så finns det ingen gemensam kravbild så det är något som vi mycket gärna skulle vilja ha

I2 drar paralleller till andra delar av verksamheten där det arbetas efter gemensamma kravbilder, något som han saknar i valideringsarbetet.

Diskussion

Deltagarna i fokusgruppen tillhör en relativt begränsad verksamhet och möjligheten att med jämna mellanrum kunna samlas för att utbyta erfarenhet och diskutera detaljer tycks inte finnas i nuläget. Under fokusgruppsintervjun visade deltagarna varandra sina referensdokument och verket med minst erfarenhet av mer omfattande valideringar fick med sig ett referensdokument efteråt. För deltagarna själva var nog den största nyttan med fokusgruppen själva erfarenhetsutbytet. Efter att fokusgruppsintervjun var över uttryckte de alla att det hade varit en värdefull träff.

Samarbete med externa parter

Verken samarbetar mycket med externa parter, allt ifrån konsulter, leverantörer, myndigheter samt forskare. Deltagarna i fokusgruppen tog upp både positiva samt negativa kvaliteter hos sina samarbetspartners.

78.I1: (...) Men om man har en leverantör som är kritisk och oberoende så ser inte jag nått hinder i att ha med en mix av leverantör och beställare, jag skulle gärna se att en tre-, ja vi hade ju en tredjepart som kom in och tittade också då va? I form av XXXX kom det folk från så, nå jag tyckte det gick bra, bara de är, det viktigaste är att de är oberoende

I 78 tar I1 upp vikten av att vara oberoende i valideringen. Oavsett vilken organisation som står bakom handlar det om att göra ett bra valideringsarbete. Exempelvis får inte en leverantör se valideringsprocessen enbart som en möjlighet att göra en ekonomisk vinst för sitt företag.

Deltagarna var relativt överrens om att en utomstående leverantör kan vara med och driva en validering, men de poängterade även att den egna kompetensen är viktigt resurs som måste få vara med.

80.I3: Men en leverantör kan ju få vara en som driver själva valideringen, men

resurserna består ju fortfarande utav en kompetens ifrån verket, så en leverantör även om de driver valideringen så kan han ju liksom inte styra det åt sitt håll i alla fall, så jag ser inga hinder till att en leverantör är den som driver valideringen, så länge man har resurser, så länge resurserna består utav den egna kompetensen

89.I3: (...) Jag tycker också man tar, ska ta med biten som du säger I2 att det är ganska komplext, även för en leverantör som kanske har kört det här på ett läkemedelsverk eller någon annan industri att komma ner till kärnkraftsverket och köra den här valideringsprocessen, utifrån den instruktionen som verket har, alla fall så som jag ser det så är det, det kräver en ganska djup kompetens att driva den här modellen framåt (...)

I3 lyfter fram i 80 att en leverantör kan vara med och driva en validering men understryker att de egna resurserna gör så att leverantören inte kan ”styra åt sitt håll i alla fall”, vilket tycks vara viktigt. I 89 argumenterar I3 för att det är komplicerad process och att leverantörer inte kan överföra sina kunskaper från andra domäner hur som helst, och ställer sig då tveksam till om de kan driva en valideringsprocess.

Forskningsvärlden är något som alla verk har kontakt med och även om de förstår vikten av samarbetet så kan de olika bakgrunderna krocka ibland.

206.I2: det jag funderade på det är ju på det sättet XXXX validerar då, det är ju många för oss tekniker, lite okända begrepp. Och jag tror det skrämmer många, kanske till och med SSM när dom läser valideringsrapporten. När du, och jag tycker att jag hör här från mina kollegor att liksom, lite back to basic, vad tycker egentligen operatörerna?

I 206 tar I2 upp skillnad mellan forskningsvärlden och verksamheten på verken. De flesta av dem som arbetar med valideringsprocessen har en annan bakgrund än de som arbetar på exempelvis Halden och I2 menar då att det akademiska språket och begreppsanvändning kan göra att viktiga grundläggande delar av valideringen kan missas.

Diskussion

Verken samarbetar dagligen med externa parter och det är viktigt att dessa relationer fungerar väl. Eftersom leverantörer har expertkunskap om de nya systemen som ska implementeras och verken har expertkunskap om den egna verksamheten och valideringsprocessen så är det viktigt att de utnyttjar de kompletterande egenskaperna på ett lämpligt sätt. Som alltid vid samarbeten så gäller det att hålla en kontinuerlig dialog och vara tydlig med vad som parterna kan förvänta av varandra. Det samma gäller verkens samarbete med forskningsvärlden. För att kunna dra nytta av varandra är det viktigt att båda parter förmedlar vilket format på samarbetet de behöver. Det är också viktigt att de ger varandra feedback för att kunna utvecklas.

SSM

Det fanns många åsikter bland deltagarna om hur SSM:s inblandning i valideringsprocessen ska se ut. När SSM:s granskning ska ske, hur den ska utföras samt hur rapporteringen till SSM ska göras.

619.I1: (...) Men jag, något som jag har tänkt på de här NUREG-guiderna. Vi förlitar oss på som egentligen är NRC verktyg för de ska använda i sin granskning. SSM, varför tar de inte fram några egna guidelines för vad de tänker titta på i sin granskning mot oss.

I 619 tar I1 upp problematiken med att SSM inte verkar ha egna guidelines att arbeta efter. Verken arbetar huvudsakligen efter utländska referensdokument och det tycks inte alltid stämma överens med vad de svenska myndigheterna begär och kontrollerar.

Deltagarna diskuterar också en upplevd ovisshet. Verken är inte helt klara med hur SSM vill sköta sin granskning.

155.I1: Men det är ju det som varför vi sitter här också är ju att SSM, vad jag tror att de förstår, de vet inte riktigt själva vad en-, hur en validering ska gå till, hur en rapport ska se ut. Så det är väl det vi ska utveckla lite, de kallade ju till den här workshopen i Halden, det var ju de som var [chairmen?], där också att där också

I1 hävdar att SSM inte är riktigt säkra på hur en validering ska gå till, eller i alla fall att SSM:s arbetssätt inte förmedlar det. Han tar även upp bekymret med att SSM inte tycks ge klara besked för hur en rapportering till dem ska se ut. Verken tycktes inte helt nöjda med hur SSM kommer in i valideringsprocessen. Genom att ha klarare arbetssätt och guidelines skulle SSM kunna förutse vad

verken kommer att göra och därmed kunna ägna sig åt en bevakande roll, inte en tillrättavisaende och styrande roll.

627.I1 (...) jag tycker att dom ska ha en granskande roll, precis som dom har haft tidigare

637.I3: Ni är välkomna att komma in och göra stickprov på att vi följer den här modellen och om ni kommer till resultatet att den här stora förändringen som ni har gjort med experter och allt som involverar vill säga, den här är fullgod, den här är bra. Då anser inte jag att SSM ska kunna komma in och ta bort den värderingen

I 627 argumenterar I1 för att SSM ska hålla sig till en bevakande roll och att de ska vara så pass insatta att de inte ska behöva komma med korrigerande eller kompletterande direktiv sent i valideringsprocessen.

Deltagarna nämner vid olika tillfällen vikten av att SSM måste kunna ge tydliga signaler om när en validering är "good enough".

161.I1: (...) jag känner att dom inte riktigt själva vet, vad som kommer att, dom måste ju kunna säga nån gång att det här är good enough liksom, dom känns som dom inte törs det riktigt än?

641.I1: Nu ska vi inte störa er mer, det är väl därför vi sitter här idag. För jag tror att SSM är lite osäkra på hur vi ska komma till det handslaget

643.I1: . Så att jag hoppas ju att någon snart när allt det här är över så ska vi komma dit så att man vet exakt när dom är tillräckligt nöjda. Det kan vara så att vi säger till dom att, vi är nöjda med valideringen men tänker fortsätta och utbilda i larmhantering. Det är vi också nöjda med, så ni kommer få köra upp. Vi kommer inte liksom be er, vi litar så mycket på att ni kommer själva att ta tag i den frågan.

645I1: Så att man får ett handslag. Nu är det good enough

I 641 vill I1 se en förändring i SSM:s arbete, där verken tydligt vet vad SSM önskar av en validering och i 643 vill den samme att det ska vara tydligt när en överenskommelse har gjorts. Samtidigt lyfter både han fram i 161 och 645 att de inte tror att SSM vet, i nuläget, hur de ska komma fram till att det är ”good enough”.

Även om verken vill att SSM:s roll ska vara mer av det bevakande slaget så anser de att det är viktigt att detta påbörjas så tidigt som möjligt. På så vis kan verken lättare hinna med att rätta till om något behöver korrigeras eller kompletteras.

630.I1: Och sen så är det viktigt att de kommer in, som i ett sådant stort projekt, att de kommer in tidigt i granskningen. För om de låter oss hålla på hela vägen till valideringen. Det har vi märkt nu när de kommer upp med de här kompletterande frågorna, det är ju jättejobbigt för oss att skapa, efter en validering, ja ni måste ha. Alla skiftlag måste få en veckas utbildning i larmhantering. Bara att få till det liksom.

635.I3: Ja men det känns ju nästan som att det skulle vara de som är med hela tiden och utvärderar och, i valideringen. (...)

I 630 tar I1 upp lite av logistikproblemen som uppstår om SSM kommer sent med sina kompletterande frågor. De menar att om SSM är med från ett tidigt stadium så kan de både få bättre insyn i processen och därmed fortare se när korrigeringar behöver göras.

Diskussion

SSM:s del i verkens arbete är så omfattande att det inte mer än rimligt att deltagarna har många åsikter om hur SSM:s roll ska se ut. Diskussionerna om SSM handlade mycket om att verken ville ha tydligare referensramar, framförallt för att båda parterna exakt ska veta vad som krävs och vilket ansvar de har gentemot varandra.

Metoder för att samla in data

Mätmetoder togs upp vid ett flertal tillfällen under fokusgruppen. Flera olika ämnen diskuterades, bland annat typen av mätmetod, sätt att mäta, samt kriterier för mätningar. En aspekt som togs upp var huruvida olika typer av instrumentering för psykofysiologiska mätningar (t.ex. ögonrörelser) av operatören skulle kunna störa operatörerna vid deras testkörningar.

714.I2: Ja visst. Sen kommer det här in med man instrumenterar operatören och då blir det lite artificiellt så att säga eller, ja man försöker ju att dom ska va-uppträda normalt och i sin normala miljö då va. Och jag är lite rädd att det blir en någon negativ från operatören. Man hänger på dom någon slags grej men. I Halden som sagt i forskning i sånt där så använder de sånt här men

I2 uttrycker en oro för att olika typer av mätinstrument skulle kunna störa genomförandet av själva simuleringen då det upplevs som störande av operatörerna eller gör det svårt för dem att uppträda normalt. Något som också diskuterades var möjligheten att använda olika former av mått som hämtas direkt från de tekniska systemen. Det framkom att sådana mått redan används till viss del.

785.O1: Men vi säger att flödet någonstans, att vattennivån blir fel i turbinen eller något sådant. Om det vore relevant att använda måttet på hur mycket fel det är under viss tidsperiod och bedöma hur väl dom lyckas med uppgiften Men då förutsätter det att det simuleras i simuleringen

786.I1: Ja men det gör det

787.O1: Men måttet var bara ett exempel

788.I1: Men sådana scenarier körde vi liksom temperaturen som stiger eller sjunker och så ska dom fatta beslut innan en viss tidsperiod sådär. Då loggar man hur lång tid det tog innan man fattade beslutet. Från det att dom såg, upptäckte larmet till att dom fattade

Själva kriteriet för vad som ska bedömas som en framgångsrik handling problematiserades också. Utan domänkunskap kan det vara svårt att förstå vad som är ett rimligt mått. Vid ett tillfälle diskuterades t.ex. användandet av tid som ett mått på framgång. Ibland är det t.ex. viktigare att fatta rätt beslut än att fatta ett snabbt beslut.

791.I1: Ja, vi hade väl ofta. Det går från 0 till 100, så att någon stans där, om dom tog det på 70 % eller 40 eller 30 det spelar ingen roll utan bara att man klarar av uppgiften innan en nivå mäts. För som ofta så har man den här rådrumsregeln på 30 minuter så att, alla skiften klarar det ju

på olika tid olika scenarier men allting var- hade någon acceptansnivå. Så det var inte fel för ett skiftlag komma fram på 50 minuter medan ett kom fram på 30 minuter, för det är ändå dom här PSA-analyserna till när härden friläggs som man bedömer. Alltså är man 12 timmar ifrån den eller 13 timmar ifrån den så är man tillräckligt långt ifrån. Det är ofta det där när man tappar härdkylningen som är relevant egentligen.

933.I3: Ja men det kan du ju mäta med just det här att "hur snabbt såg du informationen", "hur snabbt tryckte du". Jag sa det här med återträning här att vi gör intervjuer med de som är instruktörer uppe på KSU och då kan man lätt lägga ut till exempel en blåsande 3.14 ventil och inget skiftlag upptäckte det på den tiden som de skulle upptäcka det på. Vi gjorde om det så satte vi in en röd diod på blåsning av 3.14 ventil, alla tog det långt långt innan de skulle ha gjort det.

934.I1: Och sen inom vår business är det inte alltid helt rätt att fatta ett beslut fort heller. Det är mycket viktigare att fatta rätt beslut och så fort du har haft ett reaktorsnabbstopp så, du har ju en del ganska långsamma förlopp ändå så du har tid att fatta beslut. Så det är väldigt svårt att säga rätt att fatta det här om man ska mäta då hur fort gick det och så där. Så får man ha det till hur lång tid är det kvar till att härden friläggs också.

Ett centralt problem är alltså som synes ovan att mätningar måste tolkas av någon med expertis inom domänen för att vara meningsfulla. I1 föreslår till och med expertutlåtandet/rapport som en möjlig form av enklare validering (se excerpt 811 nedan). Huvudsakligen diskuterades dock expertisens roll i tolkningen av det som sker under en validering.

811.I1: Jag vet inte hur ni hade det I2 men vi hade i XXXX, så hade vi ett team som satt och förde loggar och så hade vi tre ett expertteam då som följde skiftlagets mer handlingar. Tog dom fram instruktioner, läste dom färdigt stegen i instruktionerna, så att man kunde titta på sådana saker och där skulle man kunna titta på en sån sak.

Varför gjorde dom inte det, varför använde dom sig inte av bilden i den datorn till exempel som dom har lärt sig. Men då är det mera en, för oss blev det ett expertutlåtande då. Så vi hade två team som, vi hade två olika rapporter också. En expertrapport och en, vad heter den, för loggbok. Vali- ja. Men det, vissa valideringar kanske bara har ett team som för loggbok då.

Mått av lägesförståelse/situationsmedvetenhet är en central del av valideringsförfarandet, speciellt i de fall då stora förändringar sker i de displayer som visar central processinformation. Deltagarna i fokusgruppen hade olika erfarenheter av detta, både ur perspektivet av metoden, i vilken det är vanligt att stoppa körningar för att kunna mäta, och i det att resultaten kan bli missvisande.

938.I1: Men det där frångår man väl lite situation awereness nu va, det där att man ska stoppa vända sig om och berätta massa värden

939.I2: Och det är just den mätmetod, man tycker att överblicken blir så mycket bättre när man verkligen kan lägga upp den information man vill ha på storbildskärmar och de tavlor, och sen tycker ändå operatören att enligt mätmetoden att det blir sämre. Men pratar man med dom så tycker de att det är bättre. Så det är min mättningsmetod som är alltså osäker

Diskussion

Deltagarna i fokusgruppen tar upp ett flertal praktiska problem rörande mätningar i samband med en validering. Behovet av att ha domänexpertis för att kunna tolka data och skapa meningsfulla mätpunkter lyfts fram. Att ha "rena" körningar i den bemärkelsen att mätmetoderna ska störa deltagarna i försöken så lite som möjligt betonas också. Fokusgruppsdeltagare I2 uttrycker en skepsis till att använda utrustning som på något sätt kan hindra eller störa deltagarna i en utvärdering. Realism anses troligen som viktigare än att kunna samla in exakta mått, vilket ligger i linje med olika utlåtande under fokusgruppen som betonar vikten av operatörernas åsikter och experters bedömning snarare än regelrätta kvantitativa mätningar.

Praktiska problem

Att utföra en validering av ett kontrollrum är en mycket omfattande verksamhet. En diskussion som fördes rörde de praktiska problem som uppstår, framförallt i termer av att det tar mycket lång tid att genomföra ändringar och att det krävs omfattande personalresurser.

280.I3: Jo, men det var lite det jag sa tidigare, jag förstår absolut att vi måste ha forskarvärlden med oss och beteendevetarbiten och allting med i det här men när det blir så här mycket så är det svårt för den normala människan och köra en valideringsprocess, det blir, det blir så brett, det blir så stort. För trots allt så har vi ett, vi står inför modernisering av hela svenska kärnkraftsflottan, vi har anläggningsändringar som är utan dess like som påverkar kontrollrummen, skulle vi köra sådana här analyser på varje ändring som påverkar kontrollrummet så skulle varje ändring ta två tre år och genomföra för vi har inte resurser och genomföra det, det är helt klart finns det inte sådan bemanning i kontrollrummet

I3 påpekar också att valideringen kräver olika typer av expertis som inte alltid finns på anläggningen som ska valideras. Detta anses som oroväckande då kärnkraftsindustrin samtidigt troligen står inför en mängd stora förändringar. Diskussionen återkom senare under fokusgruppen. Önskemål om en förenklad form av valideringsprocess lyftes fram.

749.I3: Nä det jag skulle vilja göra, det är att göra det enklare

750.I2: Ja

751.I3: Det är det i sådana fall. Lättare att arbeta mot, följa

Rapporten har tidigare tagit upp fenomenet ”envisioned world problem”, dvs. att det i en design av ett komplext system är mycket svårt att förutse alla tänkbara interaktioner mellan komponenter och regelverk. Detta togs upp under diskussionen. I2 redogör ned för ett sådant exempel.

282.I2: ...man åker dit och tittar att dom har, det är mer en verifiering dom sätter upp sin dokumentation och dom visar att vi har testat allt finns ingen avvikelser, två meter hyllpärmar men så fort jag tar fram min instruktion och försöker köra det där systemet, då fastnar jag på första punkten, för då går inte den att genomföra för att det är, det trots att tekniska specen är 10 000 sidor, allt checkat, men hellå det går inte att använda

I2 belyser i 282 behovet av praktiska prov under designprocessen. Att anta att vissa funktioner kommer att fungera utifrån en teknisk beskrivning, oavsett hur detaljerad den är, kan vara riskabelt. Verifiering och validering har i detta fall en viktig roll att fylla.

Diskussion

Även om behovet av validering/verifiering kan ses som givet i ljuset av denna rapport är det inte självklart hur dessa moment ska bedrivas. Både teori och praktik varnar för en övertro på "skrivbordstester" och påtalar behovet av iterativ utveckling. Relationen mellan verifiering och validering är särskilt intressant med tanke på detta, då en sund process bör innehålla utrymme för omdesign av kritiska funktioner innan själva slutvalideringen sker. Rent praktiska problem uppstår också om processen är mycket utdragen, något deltagarna ger uttryck för när de diskuterar att implementering kan ta mycket lång tid "vi har inte resurser och genomföra det".

Baslinjemätning

Ett av de centrala ämnena under diskussionen var aspekter av baslinjemätning. Behovet av att ha en baslinjemätning eller någon annan form av referens att jämföra olika förändringar i systemen på anläggningen mot diskuterades upprepade gånger under de tre timmarna. En frågeställning som togs upp spontant var hur ofta baslinjemätningar bör genomföras. På en av anläggningarna har de inte genomfört stora genomgripande förändringar under senare tid, utan istället ofta jobbat med små ingrepp. I ljuset av kommande större uppgraderingar sågs behovet av att ha en baslinjemätning att jämföra mot.

202.I3: Och vi sitter ju i den sitsen att vi har gjort många små förändringar så vi har ju liksom kört det här successivt och tagit det steg för steg, om det är bra eller dåligt det vet jag inte, men, jag tycker det är en styrka, men när man kommer till stora förändringar då känner jag att då behöver man nog ha den här, ett fingerprint för att kunna jämföra mot, så här, så här är det åt 2009

Senare under diskussionen kom ämnet tillbaka och preciserades i det att tankar lyftes rörande hur ofta det vore lämpligt att genomföra baslinjemätningar i kombination med ett iterativt designarbete.

452.I3: Vi har ju fått tänka till i det här nu igenom att vi vet att det ska göra en stor förändring som kommer och slå igenom fjorton femton, där i kring. Och vi har ju pratat om baslinjemätning samtidigt som det kommer ske otroligt mycket andra förändringar i kontrollrummet, vi kommer ju

lägga all vår kraft på att genomföra modernisering av fem noll fem fem tjugoett plattformen, där kommer alla våra resurser gå till det, men hur vi gör med allting annat som kommer, som kommer att rulla in. Och då har vi ju tänkt i dom här banorna att man skulle köra regelbundna baslinjemätningar, man tar fingerprint regelbundet, var tredje år, eller vart [ohörbart] år, det får någon annan värdera, men, och där i mellan så skulle man ha någon form av tolerans till att inte göra fullskaliga analyser ner i djupaste detalj, utan man kanske säger att vi ska ha medverkan av operatörer, vi ska itereras det nya gränssnittet och sen kör man återträning utan man intervjuar och ser att det här fungerar, och sen kommer en återkommande baslinjemätning som faktiskt tittar går vi tillbaka eller går vi åt rätt håll. Står vi kvar, vi kanske till och med blir bättre vi har en inlärningseffekt som man säger. Det skulle avlasta oss otroligt för jag tror inte vi skulle sänka vår kontrollrumsfunktion

459.I3: Det här är bara en tanke som vi sitter och spånar på hur ska vi klara av det här framöver, vi måste hitta någon lösning för allting som är mindre som rullar på också

460.I1: Ja i och med att ni har gjort projekt på så här små delar varje år så förstår jag att ni känner ni vill göra någon slags sätta ner foten då någon gång så ni har om fem tio år och mäta emot

461.I3: Ja

Även i de fall där det finns tidigare baslinjemätningar kan svårigheter uppstå. Dessa beror ofta på att ändringarna som ska genomföras i sig medför att de gamla baslinjemätningarna inte kan genomföras på samma sätt som ursprungsmätningarna. Detta kan t.ex. bero på att anläggningen förändrats i sin konstruktion så att vissa delar av de scenarier som användes vid ursprungsmätningen helt enkelt inte finns längre, eller att arbetsuppgifter numer sköts av automation eller av andra operatörer än i det ursprungliga scenariot.

385.I1: Ja vi körde i XXXX så körde vi en vecka med gamla operatörer och sånt, tog ihop ett team man körde igenom scenarierna

[app?] i och med att moderniseringen bort en del av problemen ja det var kanaler som skulle fela och så där i den gamla, för tio år sen så kunde en kanal fela nu i det nya så gick ju inte det för att vi hade ju byggt bort många av problemen va, det kom ju upp sånt sådant under valideringen, vi fick ändra lite för att vissa scenarior gick inte att köra fullt ut för att, det hade vart liksom en negativ inlärning att säga till att det här det här kan egentligen inte hända, men vi måste köra det för att, vi gjorde det på baslinjemätningen, det blir ju. Så vi hade, men det blev ingen rapport av testveckan där, utan det var ju mer ett sätt o testa

403.I2: Det är helt annorlunda, ja då blir det en arbetsuppgiften kvarstår ändå va att ta hand om larmen men det blir totalt annorlunda va och men det kan tillkomma nya helt nya arbetsuppgifter att till exempel ställa av anläggningen från ett nytt nödkontrollrum till exempel som inte fanns innan

Baslinjemätning vid nybyggnation lyftes fram på samma sätt fram som ett moment med särskilda svårigheter. Det tycks inte finnas någon aktuell praxis som kan användas för detta och deltagarna påpekade att utvecklingen av acceptanskriterier i stort sett blir en tankeövning vid nybyggnation.

113.I1: För nybyggnation är jättesvårt där kan man inte göra en baslinjemätning där till exempel

120.I1: Ja, det, det har vi haft workshop i Halden med hela världen där, och det var en av frågorna som kom upp alltså, hur gör man vid, dom håller på och forskar i det här Halden, jag vet inte hur långt dom har kommit, men det var ju för ett år sen vi träffades precis.

121.I2: Jag kan ju ta ett exempel hur vi gjorde efter XXXX-valideringen då, vi fick, vi tyckte vi hade gjort en tillräcklig validering, använt våra baslinje data som jämförelse, men, (...) fick vi ju då konstruera våra acceptanskriterier. Som, till exempel, ja

operatören han ska upptäcka det här larmet i så god tid att han hinner justera processen innan han löser ut på TSSS eller någonting annat. Det var liksom, då får man ju, som jag förstår det, som vi har gjort ta fram sina acceptanskriterier genom och sitta och fundera ut dom. Vad är det som är rimligt då

Ett intressant yttrande som fölls rörde kriteriet för att använda baslinjemätningar och var gränsen skulle gå för en ”godkänd” mätning. I dagsläget bygger principen på att en ny design inte ska medföra några försämringar i prestation. I2 uttryckte en viss frustration.

622.I2: Jag ser det så här, egentligen så baserar vi hela den här valideringen på något som inte är väl underbyggt. Vi baserar det bara på dagens kontrollrum och vi säger att idag har vi tillgång att köra vår anläggning med det här kontrollrummet. Och vi är så fega så vi säger inte att det här måste bli mycket bättre, utan vi säger bara vi får inte göra det sämre. Men var vi är någonstans idag, om det per definition är det tillräckligt bra eftersom kraftverken har tillstånd att köra.

En av frågeställning var möjligheten att skapa eller använda en referensanläggning av någon typ som skulle kunna användas för baslinjemätningar som en form av nationell eller internationell referens.

564.I3: Nä det är därför vi behöver referensanläggning som verkligen kan tala om att det här är bra

Diskussion

Baslinjemätningen ses som en nödvändig och viktig del av valideringen av samtliga deltagare. Däremot råder en viss förvirring kring när den ska genomföras och vad den ska omfatta. Deltagarna belyser själva flera praktiska problem, som avsaknaden av gemensamma riktlinjer för baslinjemätning, avsaknaden av referensanläggning, problem som uppstår på grund av förändringar i tekniska system eller procedurer etc. Rent praktiskt diskuteras det att kriterier ofta måste skapas utifrån vad man känner till om nödvändiga procedurer, riktlinjer och föreskrifter. En annan viktig frågeställning som lyftes fram var vilka kriterier som ska användas för att definiera en ”godkänd” mätning. Den kapacitet/säkerhetsnivå som gäller idag tycks vara normen för en ”godkänd” mätning, vilket leder till problemet att säkerheten i alla fall på pappret inte förbättras, även om det kanske är så i realiteten.

Att tolka valideringsrapporter

Resultatet av en validering är i regel en eller flera rapporter, ofta skrivna av tredje part. Deltagarna påpekade att sådana texter många gånger inte har en tydlig målgrupp, utan av naturliga skäl är skrivna ur författarnas perspektiv.

206.I2: det jag funderade på det är ju på det sättet Halden validerar då, det är ju många för oss tekniker, lite okända begrepp. Och jag tror det skrämmer många, kanske till och med SSM när dom läser valideringsrapporten.

Ofta lämnar valideringsrapporten också tolkningsutrymme. Beroende på vem som läser den och med vilket syfte kan olika slutsatser dras. Detta gäller även för dem som arbetat med valideringsprocessen.

753.I4 Färre sätt att tolka det på
754.I3: Ja. Den är för tung
755.M: Vad menar du med att tolka?
756.I4: Ja, det blir hela tiden att resultaten
757.I3: Resultaten tolkas ju på något vis, värderas på något vis
758.I4: Har du, utför du väldigt många olika och ska värdera och tolka ja då blir det väldigt mycket att tolka och utvärdera. [QVX] det är ju hop[xx]. Har du färre då blir det mera konkret än om du har många. Kan det finns det bara två möjliga svar då, det är det absolut enklaste

Diskussion

Valideringsrapporten som leverabel bör vara utformad på ett enhetligt sätt så att mottagarna kan tolka den. Om det i senare undersökningar skulle framkomma att olika mottagare tolkar den på olika sätt bör rapporten i så fall innehålla flera avsnitt med utsagor riktade specifikt till t.ex. genomföraren, tillsynsmyndigheten samt leverantören/entreprenören av förändringen.

Designprocessen

Under fokusgruppen diskuterades relationen mellan designprocessen och validering vid flera tillfällen. Hur designprocessen ska utformas för att ge bra förutsättningar för validering togs upp vid flera tillfällen. Som synes nedan betonades vikten av att operatörer medverkar i designprocessen/validering vid flera tillfällen.

285.I3: vi har faktiskt genomfört ett antal projekt vid XXXX där vi har haft

operatörer med i hela
konstruktionsprocessen

286.Unkown: hmm

287.I3: från start till mål

288.I1: det är jättebra

289.I3: och, det har varit vägen för
dom projekten

290.I1: ja

291.I2: Det måste va så

292.Unkown: ja

296.I3: Det kan inte bli bättre, men
det kan vi ju inte göra, så vi måste hitta
någon form av väg där ja, iterera med
operatörer under konstruktionsprocessen,
det är så vi försöker jobba, vi kallar det
för verifiering där vi försöker vara en
itererande del mot konstruktionsprocessen i
steg ett till tre och sen kommer vi in i
validering

544.I1: Ja, så är det på XXXX också.
Dom två duktigaste operatörerna har ju, ja
dom kommer aldrig gå tillbaka. Dom kommer
att jobba på teknikkontoret nu, dom har
blivit designers av, dom har fått lära sig
och givetvis fått hjälp av kunnigt folk
ibland.

698.I1: I och med att det kan gå en tid
efter ens validering och tills dess att man
ska ta handläggningen i drift beroende på
att man ofta utför det under revision, så
revision kan bli väldigt lång då på grund
av en stor modernisering. Så tycker jag att
dom här personliga samtalen skulle kunna
vara ett slags avslut. [...] Intervjun med
operatörerna som då kanske ska bekräfta att
det är en bra anläggning vi tar i bruk då.
Det ingår inte idag vad jag vet.

919.I1: Men hade man kopplat bort all
driftpersonal och inte gjort någon
verifiering och tagit en leverantör, "gör
ett nytt kontrollrum", så då tror jag att
du hade haft stora problem att ta det i
drift alltså. Om du inte haft med kunskap
under tiden.

Deltagarna i diskussionen lyfter alltså fram värdet av att ha operatörer inblandade i designarbetet. Åsikten tycks vara att dessa ger viktig återkoppling på vad som är bra lösningar. Framförallt anses det att acceptansen för förändringar blir god med detta förfarande.

436.I3: Jag måste säga att vi har ju faktiskt tagit många saker ifrån utvärderingar utav utförd återträning i simulator till att förändra vår, vårt verkliga kontrollrum

485.I2: det här är problemet, man får också skriva förslag till hur man vill ha det löst. Sen processas den genom XXXX kallas det en ändringsprocess, så den här frågan går upp till ett möte där den bedöms är det här en bra fråga, är det något vi ska jobba vidare med eller är det något som vi kan kasta, och sen är det här bra va så rullar det på och så kan man se att jaha, det här det hänger ihop med det här så kan man paketera det här

Deltagarna efterlyser också gemensamma riktlinjer för design. I dagsläget finns det inte sådana, i varje fall inte några som är anpassade till dagens tekniska förutsättningar.

523.I2: Men det är ju egentligen söker efter

524.I3: Ja

525.I2: Ja

526.I2: Att ha gemensamma referensramar, att ha gemensamma kravbilder på

527.I3: Gränssnittet

528.I2: På det här området vi har ju, branschen har ju enats om gemensamma krav på både elektrisk konstruktion och mekanisk konstruktion i så kallade TBE:er och KBE:er, men när det gäller det området vi representerar så finns det inget gemensamt kravbild så det är något som vi mycket gärna skulle vilja ha

571.I2: Det bästa som finns nu är ju, inom ett visst område som vi använder det nu är NUREG700.

572.I1: Ja

573.I2: Men den är baserad på
amerikansk teknologi från
574.I1: Från 79 eller sådant
575.I2: Ja 60-talet tänkte jag säga men

Bra riktlinjer anses också kunna underlätta baslinjemätningar och deras användningspotential då alla system i framtiden skulle ha mer lika förutsättningar.

594.I3: Jag skulle vilja säga att det finns ett större förutsättningar för att få ett bättre resultat i en baslinjemätning om du använder dig av referensbitarna
595.O1: Ja
596.I3: Men när du väl börjar jobba mot en baslinjemätning så har du större förutsättningar att få goda resultat. Du har inte dåligt gränssnitt utan du har faktiskt ett väl inarbetat gränssnitt som inte innehåller buggar och fel och sådant där. Så tänk om det kunde vara ett (enhetligt?) larm- och händelsesystem i Sverige

Diskussion

Vikten av operatörsmedverkan under designprocessen tas upp som en grundläggande sak av deltagarna. De efterlyser också att detta ska synas tydligare i de steg som ingår i förändringsarbetet, även valideringen. Operatörernas reflektioner lyfts fram som en möjlig avslutning på en förändringsprocess. Acceptans för förändring kan också främjas av att operatörerna är med under hela processen.

Behovet av riktlinjer för design av MSI-komponenter tas också upp. De riktlinjer som idag finns tillgängliga anses som ålderdomliga och saknar relevans. Detta kan delvis bero på den ”digitala revolutionen”. De gränssytor som skapas idag består i regel av skärmbaserade lösningar, medan tidigare system bestått av paneler och pulpeter. De möjligheter och problem som digitala system innebär omfattas inte av äldre riktlinjer. Deltagarna anser också att tydliga riktlinjer skulle kunna underlätta valideringen då vissa osäkerhetsmoment elimineras.

Sammanfattning

Fokusgruppsdiskussionen betraktades som givande av både deltagare och forskare. De ämnen som togs upp var delvis initierade av moderatorn, men de flesta togs upp spontant av deltagarna. Ursprungligen fanns farhågor om att tre timmar inte skulle räcka för att täcka alla de områden som från början planerats, men diskussionen kom att omfatta större delen av dessa.

Deltagarna uttryckte också att de tyckte det var givande att träffas och diskutera ämnet (se t.ex. excerpt 554 ovan). Ett konkret förslag är att stärka samordningen mellan verkens sätt att validera genom att organisera flera workshops där deltagarna under fria former kan diskutera valideringsförfarandet.

Metodiskt tycks fokusgruppsmetoden ha fungerat väl. Samtliga deltagare var aktiva i diskussionen, även om I4 bitvis var mer tystlåten än övriga deltagare. Detta kan bero på att I3 och I4 kom från samma anläggning och att I3 därför blev ”talesman” för den anläggningen. Vi bedömer att de diskussioner som fördes bidrog på ett bra sätt till denna rapport. Framtida studier kan mycket väl innehålla fokusgrupper om forskningsfrågan syftar till att fånga uppfattningar och åsikter i löst definierade frågeställningar.

4. Förslag och synpunkter

Valideringar av militära ledningscentraler pågår ofta under lång tid, ibland under flera dagar, med stor osäkerhet avseende en händelseutveckling som till stor del bestäms av motdragen hos en verklig eller spelad fiende. Dessutom är det vid militära valideringar vanligt att deltagarna samtidigt genomför träning, och därmed har bristande erfarenhet av sin befattning i ledningscentralen.

Att genomföra en integrerad validering av kontrollrum i kärnkraftverk är därför i flera avseenden betydligt enklare än de utvärderingar som genomförs av militära ledningssystem. Vid integrerad validering har man tillgång till operatörer med stor vana som kan sina arbetsuppgifter. Dessutom används ofta skiftlag som till vardags arbetar tillsammans med varandra. Händelseutvecklingen i scenarierna kan i större utsträckning förutbestämmas. Det är därmed mindre osäkerhetsfaktorer avseende händelseförloppet. Scenarierna är också kortare. Trots detta innebär validering av kontrollrumsförändringar i kärnkraftverk en stor utmaning, teoretiskt såväl som praktiskt. Nedan följer det som vi anser är de viktigaste synpunkterna på valideringsprocessen samt eventuella förslag till förändring eller fortsatt forskning.

Synpunkter på gemensamma riktlinjer för valideringsprocessen

Något som framförallt tydliggjorts under fokusgruppsdiskussionen var avsaknaden av gemensamma riktlinjer för valideringsprocessen. Arbetet med validering i de olika verken är på många sätt lika i termer av mål och syften, men tillvägagångssättet är inte alltid det. Deltagarna i fokusgruppen är väl medvetna om detta och de vet att det råder viss skillnad i arbetsmetod mellan verken. De har en grov uppfattning om hur de andra arbetar, men känner inte till ”detaljer” som vilka riktlinjer som tillämpas eller till hur stor del andra anläggningar tillåter inblandning av leverantörer i valideringsprocessen. Det tycks ligga i alla deltagarnas intresse att lära sig mer om de andras arbetsmetoder.

Deltagarna i fokusgruppen tillhör en relativ begränsad verksamhet och möjligheten att med jämna mellanrum kunna samlas för att utbyta erfarenhet och diskutera detaljer tycks inte finnas i nuläget. Under fokusgruppsintervjun visade deltagarna varandra sina referensdokument och deltagarna från anläggningen med minst erfarenhet av mer omfattande valideringar fick med sig ett referensdokument efteråt. För deltagarna själva var troligen den största nyttan med fokusgruppen själva erfarenhetsutbytet.

Förslag: Handböcker

1.) I rapportens bakgrund tas några handböcker för valideringsprocesser upp, bl.a. Rollenhagen m.fl. (1998). Detta visar att mycket arbete redan har utförts för att skapa användbara handböcker. I dagsläget verkar det som tillståndshavarna

saknar ett utlåtande från SSM om vilken av dessa handböcker som bör användas. Vi bedömer att det visserligen är viktigt att välja ”rätt” handbok, men att det kanske är ännu viktigare att SSM tar ställning till vilken eller vilka som ska anses vara ”godkända” för användning. Ett sådant uttalande från SSM:s sida skulle minska osäkerheten hos tillståndshavarna. En tänkbar lösning är att SSM tillsätter en utredning för att utvärdera de olika handböcker som finns idag.

2.) I dagsläget tycks det som samtliga verk har skapat sina egna riktlinjer och handböcker. Även om samtliga delvis verkar vara baserade på NUREG förekommer naturligtvis skillnader. En tänkbar väg framåt kan vara att samla de handböcker som används idag och i samråd med representanter från tillståndshavarna göra ett försök att skapa en gemensam handbok. Detta skulle vara värdefullt av flera orsaker – dels skulle det minska osäkerheten hos deltagarna kring vad som är tillräckligt bra, vilket efterfrågades i fokusgruppen, och dels skulle gemensamma riktlinjer ge en bättre grund för jämförbara baslinjemätningar, samt bidra till att samtliga verk genomför valideringar på rimligt liknande sätt.

Förslag: Workshops

Deltagarna i fokusgruppen upplevde möjligheten att träffas och diskutera ämnet integrerad validering som mycket givande. SSM skulle kunna stötta verken genom att organisera modererade workshops där deltagare från de olika verken träffas för att byta erfarenheter. På så sätt skulle interaktionen mellan SSM och tillståndshavarna också bli mer frekvent och oklarheter och missförstånd kring valideringsprocessen skulle kunna upptäckas och korrigeras i ett tidigt stadium. Erfarenhetsutbytet kan också vara till gagn för SSM då kunskap om hur det praktiska arbetet bedrivs underlättar utvecklingen av kommande föreskrifter.

Synpunkter på baslinjemätning

Hur och när en baslinjemätning skall genomföras är en stor fråga. Behovet av att ha en ”referens”, särskilt vid nybyggnation, är stort eftersom det annars är svårt att uttala sig om ifall förändringen är av godo. Samtidigt säger både teorin (”the envisioned world problem” och ”Iterativ design”, se ovan) och fokusgruppdiskussionerna att det finns ett antal tydliga problem kopplade till själva mätningarna, både när det gäller metoderna i sig (hur mäta), hur ofta, och när det gäller relevansen i att mäta på ett system som kan vara fundamentalt annorlunda än ursprungssystemet. Grundläggande frågor som vad en ”godkänd” mätning/validering ska vara är inte självklart. I dagsläget tycks ”i alla fall inte sämre än nu” vara normen. En deltagare i fokusgruppen påpekar att detta kan synas vara ”fegt” då en förändring också borde innebära en förbättring (se excerpt 622 ovan).

Det är dock inte alltid data från en adekvat baslinjemätning finns att tillgå. Även om data finns att tillgå kan det vara så att det gått en viss tid sedan baslinjemätningen genomfördes och som tidigare diskuterats finns risken att scenarier som inte är helt relevanta för uppgraderingen har använts. Ett annat problem vid jämförelsen är att operatörerna har stor vana vid det ”gamla”

systemet som testas vid baslinjemätningen, men liten vana vid det ”nya” systemet som testas vid den integrerade valideringen. Detta normaliseras i bästa fall av att operatörerna har mycket stor vana och därmed stor förståelse för de processer de kontrollerar. Vid omfattande nybyggnation finns det dock inga garantier för att detta stämmer.

Hur ofta baslinjemätningar ska genomföras bör också utredas. I dagsläget genomförs inte baslinjemätningar med något specificerat intervall. Olika anläggningar har genomfört olika många baslinjemätningar och har sålunda olika stora underlag för historiska jämförelser. Omfattningen på en baslinjemätning är inte heller tydligt beskriven.

Förslag: Acceptanskriterier

I nuläget skapar anläggningarna själva acceptanskriterier inför en validering. Att arbetet bedrivs på detta sätt är troligen ofrånkomligt då det på förhand inte går att veta vad som måste testas. Ny teknik ger upphov till helt nya problem och scenarier. Därför kommer det inte att gå att skapa på förhand fastställda kriterier för acceptans. Däremot bör SSM utreda om det är möjligt att skapa generella riktlinjer för acceptanskriterier. Nationella ”referensmått” för acceptanskriterier skulle bidra till att samtliga verk ställer samma krav vid integrerad validering. Detta arbete bör bedrivas i dialog med tillståndshavarna och forskarvärlden.

Synpunkter på SSM:s roll i valideringsprocessen

SSM:s del i verkens arbete är så omfattande att det inte mer än rimligt att deltagarna har många åsikter om vad SSM:s roll bör vara. Diskussionerna om SSM handlade mycket om att verken ville ha tydligare referensramar, både gällande valideringsprocessen och designprocessen. Detta efterfrågas framförallt för att båda parterna (SSM och tillståndshavarna) exakt ska veta vad som krävs och vilket ansvar de har gentemot varandra.

I fokusgruppen efterfrågas en större medverkan från SSM:s sida i både designarbetet och valideringsprocessen. Tydligare riktlinjer från SSM på valideringsprocessen efterfrågas. Vissa deltagare anser att SSM bör göra utlåtande om en pågående förändring före själva valideringen. De menar att det skulle ge SSM större möjlighet att fokusera sin bevakande roll och mindre behov av en tillrättavisande och styrande roll. De önskade också riktlinjer för hur validering av mindre uppdateringar ska hanteras.

Förslag: Avstämningspunkter

En möjlighet är att utreda om det går att införa ”avstämningspunkter” i valideringsprocessen, dvs. på förhand givna punkter i förändringsarbetet där SSM och tillståndshavaren stämmer av att inriktningen på förändringsarbetet lever upp till de krav som myndigheten och gällande föreskrifter ställer. Detta kräver självfallet en större arbetsinsats än i dagsläget, men bör totalt sett höja kvalitén på förändringsarbetet samt minska risken för förseningar vid

ombyggnation, eftersom det minskar risken för stora anmärkningar vid slutvalideringen.

Synpunkter på externa aktörers roll i valideringsprocessen

Verken samarbetar dagligen med externa parter och det är viktigt att dessa relationer fungerar väl. Eftersom leverantörer har expertkunskap om de nya systemen som ska implementeras och verken har expertkunskap om den egna verksamheten och valideringsprocesserna så är det viktigt att de utnyttjar de kompletterande egenskaperna på ett lämpligt sätt. Som alltid vid samarbeten så gäller det att hålla en kontinuerlig dialog och vara tydlig i vad som förväntas av varandra. Det samma gäller verkens samarbete med forskningsvärlden. För att kunna dra nytta av varandra är det viktigt att båda parter förmedlar vilket format på samarbetet de behöver. Det är också viktigt att de ger varandra feedback för att kunna utvecklas. I fokusgruppen finns det yttranden som tyder på att t.ex. resultaten som redovisas i valideringsrapporter utförda av forskare upplevs som svårtolkade. Detta är självfallet inte önskvärt; en valideringsrapport måste vara tydlig och begriplig för mottagaren.

En annan aspekt av samarbete med externa aktörer som togs upp under fokusgruppen är till vilken grad leverantörer bör vara inblandade i valideringsprocessen. Representanter från verk där detta skett har inte framfört några negativa erfarenheter. En viss försiktighet bör dock iaktas när det gäller att överlämna valideringsansvar till leverantörer. Tydliga riktlinjer kring detta bör tas fram.

Förslag: Gemensamma möten

För att öka förståelsen mellan parterna (tillståndshavarna, forskarvärlden, SSM och leverantörer) i en validering bör gemensamma möten hållas regelbundet (t.ex. en gång per år). Dessa möten bör användas till att diskutera begrepp och metoder som används i valideringen för att öka förståelsen mellan parterna. En fördel är om mötena kan hållas på en plats där det finns tillgång till utrustning som underlättar diskussionerna i de fall exempel behövs (t.ex. Halden eller en av anläggningarna).

Förslag: Sammansättning av valideringsteam

SSM bör utreda vilka parter som bör ingå i ett valideringsteam för att kunna ge en rekommendation eller riktlinje till tillståndshavarna. Leverantörernas roll i dessa bör belysas.

Synpunkter på tidpunkt och konsekvenser av integrerad validering

När integrerad validering genomförs är det ofta väldigt kort tid innan planerad implementering. Detta kan medföra att det i princip blir omöjligt att genomföra förändringar av kontrollrummet med avseende på resultat som framkommit i

valideringen. Detta väcker frågan om hur stora problem en integrerad validering egentligen skulle behöva identifiera för att stoppa, eller medföra omedelbara förändringar av, den planerade implementeringen.

Förslag: Tidpunkt för integrerad validering

SSM bör utreda, och bestämma riktlinjer för, hur långt innan planerad implementering integrerad validering bör genomföras.

Förslag: Åtgärder efter integrerad validering

SSM bör utreda, och bestämma riktlinjer för, vilka konsekvenser identifierade brister vid integrerad validering ska medföra för planerad implementering

Synpunkter på designprocessen

Design handlar ur ett Human Factors-perspektiv om att se till att anläggningen, systemkonstruktionen, uppgifter samt arbetsmiljö är förenliga med personalens sensoriska, perceptuella och kognitiva uppfattningar. Dessutom är det viktigt att inkorporera HF-principer så tidigt som möjligt i utformningen av ett kontrollrum. I bakgrunden till rapporten påpekas vikten av att arbeta iterativt med designen av kontrollrummet. I fokusgruppen diskuterades vikten av att ha operatörsmedverkan redan på ett tidigt stadium i designen. Även i designlitteraturen lyfts behovet av användarmedverkan i designprocessen fram som en viktig förutsättning för att skapa ett bra system som accepteras av sina brukare.

Designprocessen saknar också en gemensam handbok/riktlinjer. De riktlinjer som finns idag betraktas som ålderdomliga av deltagarna i fokusgruppen. Detta beror troligen på den ”digitala revolution” som pågått en längre tid i kontrollrumssammanhang. De riktlinjer som togs fram för analoga komponenter är inte helt giltiga i skärmbaserade miljöer. Flexibiliteten som erbjuds i skärmbaserade system skapar stora möjligheter, men medför också nya problem då informationen organiseras och presenteras på ett nytt sätt.

Förslag: Designrekommendationer

Då det är troligt att ett flertal stora moderniseringar kommer att äga rum i närtid vore det lämpligt att SSM i samråd med forskare och erfaren personal från anläggningarna skapar designrekommendationer. Erfarenheter från de anläggningar som nyligen genomfört större förändringar bör samlas i ett sådant dokument, tillsammans med ”state-of-the-art” inom design och designmetod gällande kontrollrum. Sådana rekommendationer kan även ge återkoppling på vad som bör beaktas vid en validering. De bör också ha en positiv effekt på utfallet av en validering om de följs. Eventuellt borde SSM verka för att det inom EU skapas någon typ av referensanläggning för att man vid utformning av ett kontrollrum ska veta vad som är bra design och vad som redan har visat sig vara dåliga lösningar.

Synpunkter på mät- och beräkningsmetoder

Den litteratur som gått igenom pekar på att de mät- och beräkningsmetoder som används i stort sett kan klassificeras som "traditionell" Human Factors. Prestationsbedömning, situationsförståelse och mental arbetsbelastning är centrala mått. Jämförelse med historiska data förekommer i de fall det är möjligt. Valideringar använder ofta en stor mängd mätmetoder, som oftast endast visar små skillnader mellan baslinjemätning och uppgraderat system. Separata analyser kan då påvisa enstaka skillnader för enstaka mått för olika operatörer vid olika scenarier. Detta kan göra det svårt att få en helhetsbild av resultatet.

Förslag: Alternativa mätmetoder

I denna rapport diskuteras möjligheten att i större utsträckning använda t.ex. psykofysiologiska och dynamiska mätmetoder. Psykofysiologiska mått ger reliabla mätningar, men kan vara svåra att tolka. Rapporten tar dock upp lovande exempel på psykofysiologiska mätningar som visar på kopplingen mellan psykofysiologi och beteende.

Dynamiska mätmetoder kan ge ett mått på prestationen över tiden kopplat till olika skeenden i scenarierna. Detta är en stor fördel i den typ av dynamiska situationer som utspelar sig i ett kontrollrum, och är synnerligen värdefullt för att identifiera kritiska faser i händelseförlopp som utgör "flaskhalsar" med avseende på informationshantering och beslutsfattande.

En utredning skulle kunna genomföras för att undersöka om psykofysiologiska och dynamiska mätmetoder, och i så fall vilka, kan vara användbara vid validering av kontrollrum.

Förslag: Alternativa statistiska metoder

I denna rapport presenteras vissa statistiska metoder som vi inte har kunnat hitta i de rapporter som vi har undersökt. Dessa är t.ex. kovariansanalys, datareduktion (t.ex. faktoranalys) och kausalanalys (LISREL). Kovariansanalys borde kunna användas för att normalisera för skillnader mellan mättillfällen. Datareduktion borde kunna användas för instrumentutveckling och för förenklad resultatpresentation, och tillsammans med kausalanalys i valideringsarbetet för att identifiera orsakssamband. En utredning skulle kunna genomföras för att undersöka om, och hur, denna typ av metoder skulle kunna användas vid validering.

5. Sammanfattning och slutsats

I rapporten har ett stort antal frågor berörts. Projektets omfattning har varit av betydande karaktär och har genererat förslag på ytterligare arbete snarare än lösningar. Arbetet har varit värdefullt i det att ett antal problemställningar kring valideringsprocessen har kunnat formuleras. Dessa har formulerats som synpunkter på valideringsprocessen och omfattar:

- a. Behovet av gemensamma riktlinjer för validering
- b. Baslinjemätningar: hur och när
- c. Externa aktörers roll
- d. SSM:s roll
- e. Tidpunkt och konsekvens av integrerad validering
- f. Behovet av gemensamma riktlinjer för design
- g. Möjligheter med alternativa mått- och beräkningsmetoder

Till samtliga av dessa problemställningar har ett antal förslag på åtgärder givits. Dessa finns utförligt formulerade ovan, och sammanfattas i punktform nedan.

Den empiriska delen av projektet har omfattat en litteraturöversikt och genomförande av en så kallad fokusgruppsdiskussion med deltagare från de olika kärnkraftverken i Sverige. Fokusgruppen betraktades som givande av båda parter (forskarna och deltagarna). Det bör dock klarläggas att utsagan från en enskild fokusgrupp bör tolkas med försiktighet. Ytterligare fokusgrupper och intervjuer bör genomföras för att kunna dra mer långtgående slutsatser.

De kanske viktigaste slutsatserna av arbetet är att det för närvarande råder en osäkerhet bland tillståndshavarna. De anser sig sakna tydliga riktlinjer både gällande själva valideringsprocessen och det övergripande designarbetet. Idag hanteras dessa problem till stor del lokalt, vilket resulterar i liknande men inte identiska handböcker. Detta är i sig inte fel så länge handböckerna håller en tillräcklig kvalitet, men det oklart om och hur dessa kvalitetssäkrats. Tillståndshavarna är inte heller helt klara över SSM:s roll i valideringsprocessen och söker dialog.

Sammanfattning av förslag:

- Utvärdera de olika handböcker som finns idag.
- Skapa en gemensam handbok med riktlinjer för validering.

- Organisera modererade workshops där deltagare från de olika verken kan träffas för att utbyta erfarenheter.
- Utred möjligheten att skapa generella riktlinjer för acceptanskriterier.
- Utred möjligheten att införa ”avstämningpunkter” i valideringsprocessen.
- Genomför gemensamma möten med SSM, tillståndshavare, forskarvärlden och leverantörer).
- Utred vilka parter som bör ingå i ett valideringsteam.
- Utred hur lång tid innan planerad implementering integrerad validering bör genomföras.
- Utred vilka konsekvenser identifierade brister vid integrerad validering ska medföra för planerad implementering.
- Skapa ett dokument med rekommendationer för design.
- Utred möjligheter till användning av dynamiska och psykofysiologiska mätmetoder.
- Utred möjligheter till användning av alternativa statistiska metoder för datareduktion och kausalanalys.

6. Referenser

Agar, M. & MacDonald, J. (1995). Focus groups and ethnography. *Human Organizations*, Vol. 54 ,p. 78-86.

Alberts, D. S., & Hayes, R. E. (2002). *Code of best practice for experimentation.*: CCRP Publication series.

Alfredson, J. (2001). *Aspects of situational awareness and its measures in an aircraft simulation context* (LiU-Tek-Lic-2001:2). Linköping: Graduate school for Human-Machine interaction, Division of Industrial Ergonomics, Institute of Technology, Linköping University.

Alfredson, J., Nählinder, S., & Castor, M. (2004a). *Measuring eye movements in applied psychological research - five different techniques - five different approaches* (FOI-R--1406-SE). Linköping: Swedish Defence Research Agency.

Alfredson, J., Oskarsson, P.-A., Castor, M., & Svensson, J. (2003). Development of a meta instrument for evaluation of man-system interaction in systems engineering. In G. L. Rafnsdóttir & H. Gunnarsdóttir & ó. Sveinsdóttir (Eds.), *Mind and body in a technological world, Nordic Ergonomics Society, 35th Annual Conference* (pp. 77-79). Reykjavik: Nordic Ergonomics Society (NES).

Alfredson, J., Oskarsson, P.-A., Castor, M., & Svensson, J. (2004b). *Metodvalsverktyg - Ett hjälpmedel vid planering av MSI-utvärdering* (FOI-R--1225--SE). Linköping: Totalförsvarets Forskningsinstitut.

Angelborg-Thanderz, M. (1990). *Prisvärd militär flygning med rimliga risker* (Doktorsavhandling FOA C 500083-5.1). Stockholm: Handelshögskolan / Försvarets Forskningsanstalt.

Angelborg-Thanderz, M. 1997, Military pilot performance – dynamic decision making in its extreme, in *Decision Making Under Stress, Emerging Themes and Applications*, (eds.), Flin, F., Salas, E., Strub, M., & Martin, L. (Ashgate Publishing Company, Aldershot. Hants, England), 225-232.

Annett, J. (2003). Hierarchical task analysis. In E. Hollnagel (Ed.), *Handbook of cognitive task design* (pp. 17-35). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Basch, C., DeCiccio, I., & Malfetti, J. (1989). A focus group study on decision processes of young drivers: reasons that may support a decision to drink and drive. *Health Education Q*, Vol. 16, p. 389-396.

Berggren, P. (2000). *Situational awareness, mental workload, and pilot performance - relationships and conceptual aspects* (FOA-R--00-01438-706--SE). Linköping: Försvarets Forskningsanstalt.

- Berggren, P. (2005). Observing situational awareness: When differences in opinion appear. In H. Montgomery, R. Lipshitz & B. Brehmer (Eds.), *How professionals make decisions* (pp. 233-241). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bloor, M., Frankland, J., Thomas, M. & Robson, K., (2001). *Focus groups in social research*. The Cromwell Press, Wiltshire
- Booher, H. R. (Ed.) (1990). *MANPRINT an approach to systems integration*. New York: Van Nostrand Reinhold.
- Braarud, P. Ø., & Skraaning, G. (2006). Insights from a benchmark integrated system validation of a modernized NPP control room : performance measurements and the comparison to the benchmark system, *Presented at American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Controls, and Human Machine Interface Technology, 5, Albuquerque, NM, 2006-11-12--11-16.*: IFE/HR/E-2006/029.
- Braarud, P. Ø., & Strand, S. (2007a). *O2 Turbic human factors validation - Analysis and results, IFE/HR/F/2007/1317*. Halden, Norway: Institute for Energy Technology
- Braarud, P. Ø., & Strand, S. (2007b). Extended summary and support for interpretation of the O2 Turbic human factors validation, IFE/HR/F/2007/1327. Halden, Norway: Institute for Energy Technology.
- 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. (2009). *The use of structural equation modeling to describe the effect of operator functional state on air-to-air engagement outcomes* (Dissertation No. 1251). Linköping: Graduate school for Human-Machine interaction, National Graduate School in Cognitive Science, Department of Management and Engineering, Linköping University.
- 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. (1996a). Mental workload test and evaluation. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of human factors testing and evaluation* (pp. 181-199). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Charlton, S. G. (1996b). 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.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.

Cooper, G. E., & Harper, R. P. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities*. (NASA TN-D-5153). Moffett Field, CA: NASA Ames Research Center.

Endsley, M. R. (1996a). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications. Human factors in transportation* (pp. 163-181). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Endsley, M. R. (1996b). Situation awareness measurement in test and evaluation. In T. G. O'Brien & S. G. Charlton (Eds.), *Handbook of human factors testing and evaluation* (pp. 159-180). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Endsley, M. R. (1997). The role of situation awareness in naturalistic decision making. In C. E. Zsombok & G. Klein (Eds.), *Naturalistic decision making* (pp. 269-283). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT), In *Proceedings of the IEEE national aerospace and electronics conference, NAECON 1988: held at the Dayton convention Center* (Vol. 3, pp. 789-795). New York: IEEE.

Ericsson, K.A. & Simon, H.A. (1984). *Protocol Analysis – Verbal reports as data*. Cambridge, Mass: The MIT Press.

Eriksson, L., Jansson, B., Ströbeck, E., Gunnarsson, T., & Ståhle, L.-O. (1998). *Handbok - Validering av kontrollrumsändringar.: MTO-gruppen vid Vattenfall Energisystem AB, Vattenfall*.

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

Granström, K. (2002). GÖTEBORGSKRAVALLERNA– Identitetsskapande och attitydförändringar genom deltagande i fredliga demonstrationer och våldsamma upplopp. *Styrelsen för Psykologiskt Försvar, Stockholm*.

Green, M. & Collier, S. (1999). *Verification and Validation of Human Factors Issues in Control Room Design and Upgrades*. SKI Report 99:10.

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

Hammarlund, K., Lundgren, I. & Nyström, M. (2008). In the heat of the night, it is difficult to get it right – teenagers attitudes and values towards sexual risk taking. *International Journal of Qualitative Studies on Health and Well-being*. Vol. 3, No. 2, Pages 103-112

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. Booher (Ed.), *MANPRINT an approach to systems integration* (pp. 433-470). New York: Van Nostrand Reinhold.

Hogg, D., Follesø, K., Torralba, B., & Volden, F. S. (1994). Measurement of the operator's situation awareness for use within process control research: Four methodological studies. (HWR-377). Halden, Norway: OECD Halden Reactor Project.

Hollnagel, E. (1998) *Cognitive Reliability and Error Analysis Method*. Oxford: Elsevier Science Ltd.

Hollnagel, E. & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. Boca Raton, FL: CRC Press / Taylor & Francis.

ISO 7000:2004. *Graphical symbols for use on equipment -- Index and synopsis*. Geneva: ISO.

IEC 80416-3:2002. Basic principles for graphical symbols for use on equipment -- Part 3: Guidelines for the application of graphical symbols. ISO.

ISO 9241-12:1998. *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 12: Presentation of information*. Geneva: ISO.

ISO 9241-13:1998. *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 13: User guidance*. Geneva: ISO.

ISO 9241-14:1997. *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 14: Menu dialogues*. Geneva: ISO.

ISO:11064-4:2004. *Ergonomic design of control centres -- Part 4: Layout and dimensions of workstations*. Geneva: ISO.

ISO 13407:1999. *Human-centred design processes for interactive-systems*. Geneve: ISO.

ISO/IEC 11581-1:2000. *Information technology - User system interfaces - Icon symbols and functions - Part 1: Icons: General*. Geneve: ISO.

ISO/IEC 11581-2:2000. *Information technology - User system interfaces - Icon symbols and functions - Part 2: Object icons*. Geneve: ISO.

ISO/IEC 11581-3:2000. *Information technology - User system interfaces - Icon symbols and functions - Part 3: Pointers*. Geneve: ISO.

ISO/IEC-11581-6:1999. *Information technology - User system interfaces and symbols -- Icon symbols and functions -- Part 6: Action icons*. Geneve: ISO.

ISO/TR 16982:2002. *Ergonomics of human-system interaction -- Usability methods supporting human-centred design*. Geneva: ISO.

ISO/TR 18529:2000. *Ergonomics -- Ergonomics of human-system interaction -- Human-centred lifecycle process descriptions*. Geneve: ISO.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific software international.

Kidd, P. & Parshall, M. (2000). Getting the focus and the group: Enhancing analytical rigor in focus group research. *Qualitative Health Research*, Vol. 10, p. 293-308.

Kirwan, B., & Ainsworth, L. K. (1992). *A guide to task analysis*. London: Taylor & Frances.

Kitzinger, J. (1994). *The methodology of focus groups: the importance of interaction between research participants*, *Sociology of Health*, Vol. 16, p.103-121

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. V. Prabhu (Eds.), *Handbook of human-computer interaction* (Second, completely revised edition ed., pp. 717-732). Amsterdam: Elsevier Science B.V.

Linell, P. (1994). *Transkription av tal och samtal: Teori och praktik. Arbetsrapport från TEMA Kommunikation, 1994:9*, Linköpings Universitet.

MAEVA, (2004). *Validation Guideline Handbook*. MAEVA, EUROCONTROL.

Magnusson, S. (2002). Similarities and differences in psychophysiological reactions between simulated and real air-to-ground missions. *The International Journal of Aviation Psychology*, 12(1), 49-61.

Magnusson, S., & Berggren, P. (2002). Dynamic assessment of pilot mental status, *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, September 30-October 4, 2002, Baltimore, Maryland* (pp. 1997-2001). Santa Monica, Ca: Human Factors and Ergonomics Society.

MANPRINT Handbook (2005). Office of the deputy chief of staff G1, MANPRINT Directorate, Pentagon: Washington, DC. Hämtad 2009-12-17 från: <http://www.manprint.army.mil/manprint/docs/MPTHandbook.pdf>

Morgan, D. (1996). Focus groups. *Annual Review of Sociology*, Vol. 22, p.129-152.

NATO code of best practice for C2 assessment (2002).: CCRP Publication series.

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

Nielsen, J. (1994). Heuristic evaluation. In Nielsen, J., and Mack, R.L. (Eds.), *Usability Inspection Methods*, John Wiley & Sons, New York, NY

Nielsen, J. (2004). *Ten usability heuristics*. Retrieved December 17, 2009, from URL:http://www.useit.com/papers/heuristic/heuristic_list.html

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces, *Proc. ACM CHI'90 Conf.* (Seattle, WA, 1-5 April), 249-256.

Norros, L., & Savioja, P. (2006). Towards a theory and method for usability evaluation of complex human-technology systems. In R. N. Pinar & A. A. P. Koningsveld & P. J. M. Settels (Eds.), *Proceedings of the 16th world congress on ergonomics*.: Elsevier CD.

Norros, L., Savioja, P., & Salo, L. (2009). Approach to integrated system validation of NPP control rooms, *Sixth American Nuclear Society International topical meeting on nuclear plants instrumentation, control, and human-machine interface technologies NPIC&HMIT 2009, Knoxville, Tennessee, April 5-9, 2009*. On CD-ROM, LaGrange Park, IL: American Nuclear Society.

Nählinder, S. (2009). *Flight simulator training: Assessing the potential* (Dissertation No. 1250). Linköping: Graduate school for Human-Machine interaction, National Graduate School in Cognitive Science, Department of Management and Engineering, Linköping University.

Nählinder, S., & Berggren, P. (2002). *Dynamic Assessment of Operator Status*. Paper presented at the Proceedings of the HFES 2002 Conference., Baltimore, MD, USA.

O'Hara, J. M., Higgins, J. C., Perensky, J. J., Lewis, P. M., & Bongarra, J. P. (2004). *Human factors engineering program review model* (NUREG-0711, Rev. 2). Washington, DC: U.S. Nuclear Regulatory Commission.

O'Hara, J., Stubler, W., Higgins, J. C., & Brown, W. (1997). *Integrated system validation: Methodology and review criteria* (NUREG/CR-6393). Washington, DC: U.S. Nuclear Regulatory Commission.

Oskarsson, P.-A., Nählinder, S., & Svensson, E. (2008). *Granskning av Turbic Validering* (Manuskript). Stockholm: Strålsäkerhetsmyndigheten.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis - The use of factor analysis for instrument development in health care research*. London: SAGE.

Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design: Beyond human-computer interaction*. New York: Wiley.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Simon, H., & Tom, C. (1994). *Human-Computer Interaction*. Wokingham: Addison-Wesley.

Rencrantz, C., Lindoff, J., Svensson, E., Norlander, A., & Berggren, P. (2006). Interoperabilitets- och metodstudie i en operativ stridsledningscentral. *FOI-R--2040--SE*. Swedish Defence Research Agency: Linköping, Sweden.

Rollenhagen, C., Bladh, K., Borg, A. & Evénus, P. (1998). *Handbok – Validering av kontrollrumsändringar*.

Sanders, M. S., & McCormick, E. J. (1993). *Human factors in engineering and design* (7th ed.). New York: McGraw-Hill.

Savioja, P., & Norros, L. (2008). Systems usability - Promoting core-task oriented work practices. In E. L.-C. Law & E. T. Hvanberg & G. Cockton (Eds.), *Maturing usability: quality in software, interaction and value* (pp. 123-143). London: Springer.

Savioja, P., Norros, L., & Salo, L. (2008). Evaluation of systems usability, *Proceedings of the 15th European conference on Cognitive ergonomics: the ergonomics of cool interaction* (pp. Article No. 26). New York, NY: ACM Press.

Strauss, A. & Corbin, J. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, London, Sage.

- Svensson, E. (2003). Statistical techniques for data reduction and modelling. In M. Castor & E. Hanson & E. Svensson & S. Nählinder & P. LeBlay & 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. (2004). Statistical Techniques for Data Reduction and Modeling, In G. Wilson, W. Frazer, M. Beaumont, M. Grandt, A. Gundel, G. Varoneckas, H. Veltman, E. Svensson, A. Burow, B. Hockey, G. Edgar, T. Balkin, K. Gilliland, & R. Schlegel, *Operator Functional State Assessment*, RTO-TR-HFM-104.
- Svensson, E., Angelborg-Thanderz, M., & van Avermaete, J. (1997). *Dynamic measures of pilot mental workload, pilot performance, and situational awareness* (Tech. Rep. No. VINTHEC-WP3-TR01). Amsterdam: NLR.
- Svensson, E., Angelborg-Thanderz, M., Sjöberg, L., & Olsson, S. (1997). Information complexity: Mental workload and performance in combat aircraft. *Ergonomics*, 40, 362–380.
- Svensson, E., Angelborg-Thanderz, M., & Wilson, G. F. (1999). *Models of pilot performance for systems and mission evaluation - psychological and psychophysiological aspects*. AFRL-HE-WP-TR-1999-0215.
- Svensson, E., Rencrantz, C., Lindoff, J., Berggren, P., & Norlander, A. (2006). Dynamic measures for performance assessment in complex environments, *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, San Francisco, Ca* (pp. 2585-2589). Santa Monica, Ca: Human Factors and Ergonomics Society.
- Svensson, E., & Wilson, G. F. (2002). Psychological and psychophysical models of pilot performance for systems development and mission evaluation. *The International Journal of Aviation Psychology*, 12(1), 95-110.
- Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool 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. (1995). *CC-SART: The development of an experimental measure of cognitive compatibility in system design*. Report to TTCP UTP-7 Human factors in aircraft environments, Annual meeting. Toronto, Canada: DCIEM.
- The Command and Control Research Program. Hämtad 2009-05-06 från: http://www.dodccrp.org/html4/books_downloads.html

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.

VINTHEC. (1997). *Review of eye point-of-gaze equipment and data analysis (Final report)*. Technical Report: VINTHEC-WP2-TR-02.

Wibeck, V. (2000). *Fokusgrupper: Om fokuserade gruppintervjuer som undersökningsmetod*. Studentlitteratur, Lund.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Waag, W. L., & Bell, H. H. (1997). Situation assessment and decision making in skilled fighter pilots. In C. E. Zsombok & G. Klein (Eds.), *Naturalistic decision making* (pp. 247-254). Mahwah, New Jersey: Lawrence Erlbaum Associates.

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

Zon, R., Neary, C., Taylor, M., Berggren, P., Jacobsson, M., Andersen, H., et al. (2004). *Full-scale simulator experiment - Final report* (VINTHEC II Technical Report No. VINTHEC II-WP8-TR 01).



Strålsäkerhetsmyndigheten
Swedish Radiation Safety Authority

SE-171 16 Stockholm
Solna strandväg 96

Tel: +46 8 799 40 00
Fax: +46 8 799 40 10

E-post: registrator@ssm.se
Webb: stralsakerhetsmyndigheten.se