

# Grundsätzliche Überlegungen für die Planung von Evaluationsmaßnahmen im eigenen Projekt- was, wie und wann evaluieren?

Prof. Dr. Annette Kluge

*Fachgebiet Wirtschafts- und Organisationspsychologie  
Abteilung für Informatik und Angewandte Kognitionswissenschaft  
Fakultät für Ingenieurwissenschaften  
Universität Duisburg-Essen, Lotharstraße 65, 47057 Duisburg  
(e-mail: annette.kluge@uni-due.de)*

---

**Abstract:** In diesem Beitrag werden grundsätzliche Überlegungen für die Planung von Evaluationsmaßnahmen für ein Forschungs- und Entwicklungsprojekt, z.B. den technischen Teil eines kognitiven Systems für Entwickler/-innen mit ingenieurwissenschaftlichem oder informatischen Hintergrund, dargestellt. Zunächst wird kurz erläutert, was Evaluation nicht ist und „Evaluation“ als Begriff definiert und erläutert. Ein Schwerpunkt liegt auf der Human Factors-basierten Erläuterung, was unter Wirkung und Nutzen als Evaluationskriterien zu verstehen ist, welche Methoden zur Wirkungsmessung eingesetzt werden können sowie wann man mit der Planung der Evaluation im Projektablauf beginnen sollte. Der Beitrag schließt mit Anregungen dazu, wie man die Evaluationsergebnisse verwenden sollte ab und endet mit einer zusammenfassenden benutzerfreundlichen Checkliste.

*Keywords:* Interventionswirkung, Interventionsnutzen, Cognitive Ergonomics, Mental Workload, Situation Awareness, Effektgrößen

---

## 1. DER PERSÖNLICHE HINTERGRUND FÜR DIESEN BEITRAG

Dieser Beitrag ist eine Reaktion auf einen Ausspruch eines etwas verzweifelten Teilnehmers des letzten Workshops. „Die Psychologen im Projekt sagen uns dann immer: ‚Ihr müsst das auch evaluieren‘, aber wir wissen gar nicht was die meinen!?“

Also- man „muss“ etwas evaluieren, das scheint heute „en vogue“. Projekte und Anträge scheinen nur noch dann erfolgreich zu sein, wenn man eine sozialwissenschaftliche Disziplin mit im Projekt- oder Antragsboot hat, die dem Geldgeber hinter her sagen kann, was das alles denn gebracht hat.

Ich will der Leserschaft dieses Beitrags eine psychologische Perspektive auf das, „was das denn alles gebracht hat“ verdeutlichen. Dabei greife ich auch auf Erfahrungen aus meinem ersten EU-Projekt mit 10 Unterprojekten an der RWTH Aachen in Zusammenarbeit mit Ingenieuren/-innen und Informatikern/-innen aus dem Jahr 1996-99 zurück, als interdisziplinäre Projekte noch ganz jung waren und keine/r so richtig wusste, wie man diese Welten jemals zusammenbringen sollte.

In dem EU-Projekt ging es damals um die Entwicklung von 10 verschiedenen web-basierten e-learning Applikationen. Beim e-learning kommt es darauf an, dass man es Lernenden mit einer Art von computerbasierter Instruktion ermöglicht, wenn nicht sogar erleichtert, Wissen und Fertigkeiten zu erwerben. Das Ziel ist also, dass jemand, der sich mit dem e-learning Produkt auseinandergesetzt hat, hinterher mehr kann

oder weiß oder vorher. Ausgehend von diesen Erfahrungen mit diesem EU-Projekt, die nun schon 16 Jahre zurück liegen, lassen Sie mich zuerst sagen, was Evaluation nicht ist. Evaluation ist KEINE Funktionsprüfung!

Es geht nicht darum sich am Ende der 2-3 jährigen Entwicklungsarbeit zu freuen, dass das System fehlerfrei und ohne Abstürze läuft. Das setzen wir (Psychologen/innen) (vielleicht [?] leichtgläubigerweise) voraus.

Evaluieren heißt z.B. in dem obengenannten EU-Projekt, dass man sich anschaut, ob man es den Lernenden tatsächlich ermöglichen hat, anschließend, nach der Auseinandersetzung mit dem e-learning Programm, mehr zu wissen oder zu können.

Ich bin, wie schon gesagt, damals bereits davon ausgegangen, dass die beteiligten ingenieurwissenschaftlichen Projektpartner/innen ihr Fach beherrschen und hatte das Kriterium, dass „es“ abschließend auch „funktioniert“, als gegeben vorausgesetzt. Ich habe in der Zwischenzeit dazu gelernt, nichtsdestotrotz kann ich sagen, dass „es funktioniert“ ist aus psychologischer Perspektive die notwendige Voraussetzung dafür, dass man „etwas“ evaluieren kann. Ich kann den „Fahrspaß“ eines Fahrzeugs auch nur evaluieren, wenn das Fahrzeug überhaupt fährt.

## 2. EVALUATION IST .....?

Nachdem ich gesagt habe, was Evaluation nicht ist, nähern wir uns nun dem psychologischen Evaluationsbegriff. Dieser ist meines Wissens nach für die Betrachtung von kognitiven Systemen bisher nicht definiert.

Die meisten Personen, die den Evaluationsbegriff nutzen, meinen mit Evaluation eine Art „Bewertung“.

Aus psychologischer Perspektive ist mit Evaluation die Erfassung von „Wirkungen“ gemeint. Das können positive Wirkungen sein, z.B. ein Operator wird in der Entscheidungsfindung entlastet, oder auch negative, wie z.B. der Mental Workload steigt an.

Evaluation auf den Bereich der kognitiven Systeme angewendet, definiere ich im Folgenden in Anlehnung an Mittag und Hager (2000, S. 103) **als die wissenschaftlich fundierte, empirische und hypothesenorientierte Erforschung der Wirkung und des Nutzens einer technischen Entwicklung mit dem Ziel die Evaluationsergebnisse zur (Neu-)konzeption, Ausgestaltung und Umsetzung in einem konkreten soziotechnischen Kontext zu verwenden.**

Die **Wirkung** (die Effektivität) umfasst dabei die kognitiven, motivationalen, emotionalen und sozial-interaktiven Prozesse, die mit einer technischen Entwicklung ausgelöst oder die bei ihrer Anwendung mehr oder weniger direkt beobachtet werden können.

Der **Nutzen** (die Effizienz) beschreibt den Aufwand der technischen Entwicklung in Relation zur Wirkung. Es wird hier die Frage gestellt, ob sich der Aufwand im Vergleich zur festgestellten Wirkung lohnt.

Im Idealfall lassen sich aus kognitiven Modellen oder Theorien konkrete Hypothesen über die Wirkungen z.B. auf die Situation Awareness oder den Mental Workload ableiten, um **hypothesenorientiert** vorzugehen. **Wissenschaftlich** fundiert umfasst die Aspekte der Validität, Reliabilität und Objektivität, mit dem Ziel gültige und nachvollziehbare Aussagen zu treffen. **Empirisch** bedeutet dabei, dass man systematisch konkrete Beobachtungen durchführt, um die eigenen Hypothesen zu prüfen. Man schaut konkret „nach“, ob die eigenen Hypothesen auch zutreffen. Der Begriff „Beobachtungen“ schließt dabei verschiedene Formen der Datenerhebung ein, tatsächliche Beobachtungen, wie jemand etwas tut, auch psychophysiologische Daten, wie Augenbewegungen, oder subjektive Daten wie fragebogenbasierte Einschätzungen des Workloads.

Gerade unter Human Factors- und ergonomischen Gesichtspunkten, und vor allem unter Aspekten der sog. Cognitive Ergonomics, ist Evaluation überaus spannend, weil jede technische Weiter- oder Neuentwicklung meistens SOWOHL positive ALS AUCH negative Wirkungen entfaltet.

“**Ergonomics (or human factors)** is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data, and other methods to design in order to optimize human well-being and overall system performance”  
[http://www.iea.cc/01\\_what/What\\_is\\_Ergonomics.html](http://www.iea.cc/01_what/What_is_Ergonomics.html),  
retrieved August, 26<sup>th</sup>, 2012

Im Cockpit haben sich z.B. Anzeigen, die die Situation Awareness (SA) erhöhen sollen, als negativ für den Mental Workload herausgestellt. Automation kann zu einer durchaus sinnvollen Entlastung der Operateure führen, aber auch zu einem Verlust der SA, einem Fertigkeitenverlust, zu Ermüdung und zu Monotonieerleben.

“**Cognitive ergonomics** is concerned with mental processes, such as perception, memory, reasoning, and motor response, as they affect interactions among humans and other elements of a system. The relevant topics include mental workload, decision-making, skilled performance, human-computer interaction, human reliability, work stress and training as these may relate to human-system design.”  
[http://www.iea.cc/01\\_what/What\\_is\\_Ergonomics.html](http://www.iea.cc/01_what/What_is_Ergonomics.html),  
retrieved August, 26<sup>th</sup>, 2012

Manchmal kann man auch beobachten, dass man mit einer technischen Entwicklung eine Art von menschlichen Fehlern zwar „ausmerzt“, dafür aber eine andere Art „herauf beschwört“. Man kommt also vom „Regen in die Traufe“ oder „treibt den Teufel mit dem Beelzebub aus“. Der Ausbildungsleiter einer großen Organisation für Flugsicherung formulierte es bei der Einführung einer neuen Software einmal so: „Wir ersetzen den Human Error des Air Traffic Controllers mit dem Human Error des Ingenieurs der neuen Software“

Auch müssen die positiven und negativen Wirkungen nicht zeitgleich auftreten. Der Fertigkeitenverlust (Skill decay) oder auch der sog. Over-trust (auch Complacency genannt), als unerwünschte Nebenwirkungen bei hohen Automationsgraden, stellen sich erst viel später ein als z.B. der positive Effekt der kognitiven Entlastung der Operateure (Parasuraman, Sheridan & Wickens, 2000).

Insofern hat Evaluation auch etwas von einer „Technologiefolgeabschätzung“ und versucht in die Zukunft gerichtet zu erforschen, was man im Best Case und Worst Case zu erwarten hat, wenn man die neue technische Entwicklung einsetzt.

### 3. WAS WILL ODER SOLL MAN DENN EVALUIEREN?

Aus psychologischer Perspektive dient die Technik dem Menschen. Für die Psychologie steht der Mensch im Mittelpunkt der Betrachtung. Das technische System, die technische Erfindung, das Interface etc. sollen dem Mensch in irgendeiner Weise dienen, z.B. eine Form der Erleichterung bringen.

#### 3.1 Welche Art von Wirkungen kann man bei kognitiven Systemen betrachten?

Wenn man sich die Definition zu Cognitive Ergonomics anschaut, werden die Evaluationskriterien bereits schon angesprochen. Wie wirkt sich die technische Seite eines kognitiven Systems auf Wahrnehmung, Gedächtnis, Schlussfolgern und motorische Reaktionen des Menschen aus? Welche Wirkung hat der technische Teil des kognitiven Systems auf den Mental Workload (MWL), die Situation Awareness (SA), die für Entscheidungsprozesse essentiell ist,

auf fertigkeitbasierte Leistung, Human Reliability, Stresserleben und das für die Aufgabenausführung benötigte Training?

#### *Kognitive Aspekte:*

Auf der *Wahrnehmungsebene* kann man z.B. Aufmerksamkeits- oder Orientierungsreaktionen betrachten. Wird etwas wahrgenommen, gehört, gesehen, gefühlt? Ein Evaluationskriterium kann hier z.B. sein, ob das technische System die Aufmerksamkeit des Operators bestmöglich lenkt?

*Mental Workload* lässt sich nach Vidulich (2003) und Tsang und Vidulich (2006) beschreiben als die mentalen Kosten, die ein Operateur für die kognitive Informationsverarbeitung aufbringen muss, um eine Aufgabe auszuführen. Nach Wickens und McCarley (2008, S. 4) sollten Mensch-Maschine-Systeme generell so gestaltet sein, dass die menschlichen mentalen Ressourcen, die zur Bedienung des Systems benötigt werden, niedriger sind als die insgesamt verfügbaren, damit noch Reserven für unerwartete Aufgaben verfügbar bleiben. Ein Evaluationskriterium kann diesbezüglich die Frage sein, ob das technische System den MWL optimiert.

SA lässt sich nach Endsley (1995) definieren als die raumzeitliche Wahrnehmung der Elemente in der eigenen Umgebung (Level 1), das Verstehen ihrer Bedeutung (Level 2) und die Projektion ihrer Veränderung in der Zukunft (Level 3, S. 36, auch Wickens, 2008). Die SA wird maßgeblich beeinflusst durch selektive Aufmerksamkeitsprozesse, um die erforderlichen Informationen für Level 1 zu erwerben sowie durch das Long-Term Working Memory, das am SA-up-dating auf allen drei Level beteiligt ist (Ericsson & Kintsch, 1995; Vidulich, 2003; Wickens, 2008). Kognitive Systeme sollten so gestaltet sein, dass sie die SA optimal unterstützen, jedoch nicht den MWL erhöhen. Die Evaluationsfrage lautet dementsprechend: Welche Auswirkungen hat mein technisches System auf die einzelnen Stufen der SA?

Als *Human Error* gelten nach Reason (2008) nicht-intentionale Handlungen mit unerwünschten Konsequenzen (sog. Slips, Lapses, Mistakes). Fehler basieren vor allem auf Fertigungsdefiziten, aktuellen Aufmerksamkeitsbeeinträchtigungen und Ablenkungen und damit auf kognitiven Prozessen der Informationsverarbeitung. Kognitive Systeme sollten so gestaltet sein, dass man dem Human Error möglichst entgegenwirkt. Ein Evaluationskriterium wäre hier z.B. die Anzahl der Fehler, bzw. die Reduktion der Anzahl der Fehler im Vergleich zu einem bisherigen System. Ein Beispiel dafür findet sich z.B. bei Weyers, Burkolter, Luther und Kluge (2012)

Auf Team- oder Gruppenebene lassen sich zudem Aspekte der *Makrokognition* evaluieren.

Cognitive Processes at a **macro-level** = “knowledge interoperability development, team shared understanding development, convergence of individual mental models to team mental model” (Letsky, Warner, Fiore & Smith, 2008)

Evaluationskriterien können z.B. sein: Inwieweit fördert das technische System die SA von verteilt operierenden Teams? Wie werden Abstimmungsprozesse auf einem makrokognitiven Level durch das System unterstützt?

#### **Macrocognition:**

„Is defined as the internalized and externalized high-level mental process employed by teams to create new knowledge during **complex, one-of-a-kind, collaborative problem-solving**“ (Letsky, Warner, Fiore & Smith, 2008)

#### *Motivationale und emotionale Aspekte:*

Zusätzlich zu diesen rein kognitiven Aspekten der Leistung im Zusammenspiel des kognitiven Systems, lassen sich darüber hinaus auch *motivationale Aspekte* evaluieren.

Wie ist die Akzeptanz des neuen Systems? Welche Valenz hat die Nutzung für den Operateur?

Auch *emotionale* Prozesse können eine Rolle spielen wie z.B. das Gefallen eines technischen Systems, die „User Experience“ aber auch Ängste im Umgang mit technischen Systemen.

#### *Weitere Aspekte:*

Um noch eine weitere „Schwierigkeitsstufe“ einzubauen, sollte man dabei die Zielgruppenadäquatheit als Evaluationskriterium aufnehmen. Wenn Sie z.B. einen „Carebot“ für die Unterstützung älterer Menschen in ihren Wohnungen entwickeln wollen, dann müssen die Evaluationskriterien die Besonderheiten älterer Menschen mitberücksichtigen. Es ist nicht ausreichend zu wissen, dass eine studentische Stichprobe mit dem Carebot zielführend, z.B. beim Tischdecken, interagieren kann. Eine ältere Person muss dazu in der Lage sein.

Die Liste an Evaluationskriterien kann noch weiter fortgeführt werden. Im Rahmen dieses Beitrags beschränke ich mich jedoch auf die oben genannten.

Letztendlich muss der Entwickelnde oder der/die Auftraggebende/r zumindest identifizieren/ vorgeben, welches Ziel mit dem technischen System erreicht werden soll.

### 3.2 Wann hat etwas einen Nutzen?

Der Nutzen einer technischen Entwicklung ergibt sich rückblickend durch eine Betrachtung über die Größe oder Intensität einer Wirkung und die dafür eingesetzten Ressourcen. Diesbezüglich ist es ratsam statistisch signifikante Ergebnisse, z.B. eine signifikante Erhöhung der SA, UND die Effektgröße zu betrachten (Kluge & Burkolter, 2012). Hier ist die zentrale Frage, ob das statistisch signifikante Ergebnis auch eine praktische Bedeutsamkeit hat. Kleine Effekte können zwar statistisch signifikant sein, gleichzeitig aber auch praktisch bedeutungslos, weil man z.B. die Operateure des technischen Systems erst 12 Monate

darauf hin trainieren muss, damit dieser kleine Effekt auch eintritt. Auf der anderen Seite kann man sehr wohl auch einen kleinen Effekt, der statistisch signifikant wird als praktisch bedeutsam bezeichnen, wenn es z.B. um die Rettung von Menschenleben geht. Man ist dann auch ggf. gewillt, mehr Ressourcen zu investieren, als bei einem „Gimmick“, mit deutlich geringeren dramatischen Konsequenzen.

#### 4. WIE EVALUIERT MAN?

Mit wissenschaftlich fundierten Methoden der Evaluation sind in erster Linie psychologische und sozialwissenschaftliche Methoden gemeint wie Skalierungen, standardisierte Tests, experimentelle und quasi-experimentelle Versuchspläne, Interviewtechniken sowie ökonomische Kosten-Nutzen-Rechnungen (Westermann, 2002, S. 4).

In der Human Factors-Literatur werden vor allem subjektive (z.B. Workload Ratings/Judgements), objektive (z.B. Primary & Secondary Task Performance) und physiologische Maße (Elektronkardiogramm/EKG, Herzrate, Elektroencephalogramm/EEG Messung bzw. Event Related Potentials/ERP, Positron Emission Tomography/PET) z.B. zur Messung des MW besprochen (Kramer & Weber, 2005; Marshall, 2007; Tsang & Vidulich, 2006).

Bei den subjektiven Verfahren handelt es sich meistens um Selbsteinschätzungen der Personen, die im Rahmen der Evaluationen als Probanden (Pbn) eingesetzt werden. Der NASA-TLX ist solch ein subjektives Verfahren. Die Pbn schätzen z.B. die Höhe ihres MWL bei der Ausführung einer Aufgabe auf einer Skala zwischen 1-10 ein.

Bei den objektiven Verfahren wird die Leistung der Pbn erhoben. Man misst z.B. wie schnell jemand auf eine Warnung reagiert, man protokolliert die Mausclicks, man wertet Logfiles aus, man zählt die produzierte Menge von etwas.

Zu den psychophysiologischen Verfahren zählen z.B. das Eye Tracking, die Messung der Herzrate oder der Hautleitfähigkeit. Mit den psychophysiologischen Verfahren wird nicht die Leistung an sich gemessen, sondern die physiologischen Bedingungen unter denen sie erbracht werden und zustande kommen.

Um zu entscheiden, welches Verfahren das sinnvolle ist, muss jedoch klar sein, was man evaluieren will. Dann empfiehlt sich als nächstes ein Blick in eines der führenden Human Factors-Handbücher wie z.B. das von Salvendy (2006). Dort finden sich neben der Definition von Begriffen wie MWL oder SA auch bewährte und gängige Meßverfahren.

#### 5. WANN IM PROZESS SOLLTE ICH ÜBER EVALUATION NACHDENKEN?

In dem einleitend berichteten EU-Projekt habe ich die Erfahrung gemacht, dass die meisten Projektpartner dachten, dass Evaluation etwas ist, was „am Schluss kommt“. Zuerst müsse man das Produkt fertig haben, dann evaluiert man.

Sie ahnen es schon- so ist es natürlich nicht. Die Evaluation eines kognitiven Systems beginnt mit der Festlegung dessen was ich damit erreichen will. D.h. wie einleitend schon

genannt: Was soll es bewirken und was soll es nutzen? Sobald ich weiß, was ich erreichen will, kann ich festlegen, an welchen Evaluationskriterien ich die Zielerreichung feststellen kann. D.h. evaluieren Sie anhand der Ziele, die Sie vorab festlegen.

Gerade bei technischen Systemen ist das besonders relevant, da man sich hier bereits darüber Gedanken machen kann, welche Daten das System für meine Evaluationszwecke mit protokollieren kann oder muss. Das können z.B. Reaktionszeiten auf Warnungen sein, das können implementierte Sekundäraufgaben sein, aber auch Fehler, die das System entdeckt.

Fragebogen zur Selbsteinschätzung können ggf. auch später entwickelt oder ausgesucht werden, ihr technisches System lässt aber gerade eine Fülle von Möglichkeiten für die direkte Leistungsmessung zu, die man nicht „verschenken“ sollte.

Die Evaluationskriterien sollte man dementsprechend vom Projektstart aus mitdenken und entwickeln, auch wenn sie erst zum Einsatz kommen, wenn das technische System bei den Pbn zum ersten Einsatz kommt bzw. gekommen ist.

Gediga und Hamborg (2002, S. 41) empfehlen zudem sich in Bezug auf die Evaluationsziele folgende Fragen zu stellen, die für die Planung des Untersuchungsdesigns bei der konkreten Erfassung der Wirkungen relevant sind:

- „Which is better?“- bezieht sich auf die Evaluation von Alternativen, z.B. wenn es mehrere Möglichkeiten gibt etwas technisch zu realisieren und Sie wissen wollen, welche die bessere ist.
- „How good?“ – bezieht sich auf die Bestimmung der Ausprägung bestimmter gewünschter oder geforderter Systemeigenschaften, z.B. wenn Sie wissen wollen wie schnell eine Person innerhalb eines festgelegten Kriteriums auf eine Warnung reagiert.
- „Why bad?“ – bezieht sich darauf, Schwachstellen als Ausgangspunkt für die Gestaltung oder direkte Gestaltungsvorschläge zu liefern, und Sie z.B. herausfinden wollen, warum es zu Human Errors kommt (Gediga & Hamborg, 2002, S. 41)

#### 6. WAS MACHE ICH MIT DEN EVALUATIONSERGEBNISSEN?

In der einleitend formulierten Definition von Evaluation kognitiver Systeme habe ich beschrieben, dass Evaluation das Ziel beinhaltet, die Evaluationsergebnisse zur (Neu-)konzeption, Ausgestaltung und Umsetzung in einem konkreten soziotechnischen Kontext zu verwenden. Die Idee ist also, die Ergebnisse der Evaluation zu nutzen, um festzustellen, ob mein ursprünglich angedachtes Ziel erreicht wurde – oder leider nicht. Meistens stellt man fest, dass nur ein Teil der Ziele erreicht wurde, oder dass man wie eingangs auch schon berichtet, ein Kriterium erreicht, wie z.B. die SA zu erhöhen, gleichzeitig aber ein anderes Kriterium unterwandert wurde, z.B. bei einem ebenfalls erhöhten MWL.

Diese Evaluationsergebnisse sind eine sehr wertvolle Lerngelegenheit, um daraufhin sein System anzupassen, neu zu entwickeln oder abzuändern, wieder zu evaluieren, dann

wieder auszuwerten, und ggf. weiter zu „tüfteln“ bis man alle seine Ziele gleichermaßen erreicht hat.

Evaluation ist somit eine Art des selbstgewählten Feedbacks, welches eine wissenschaftlich fundierte Quelle des eigenen Lernens und der Weiterentwicklung darstellt, die zum Forschenden-Dasein maßgeblich dazugehört. Manche nennen es „trial-and-error“, oder „Lernen aus Erfahrung“. Wir nennen es Evaluation.

## 7. ZUSAMMENFASSENDE CHECKLISTE

Die obengenannten Gedanken sind im Folgenden in einer Checkliste für Sie zusammengefasst.

### ZIELBESTIMMUNG

- Legen Sie fest, welche Wirkung(en) das von Ihnen zu entwickelnde kognitive System erzielen soll.
- Legen sie fest, wann für Sie ein Nutzen des von Ihnen zu entwickelnden kognitiven Systems gegeben ist.

### KRITERIENENTWICKLUNG

- Legen Sie vorab Kriterien fest, mit denen Sie die Wirkung erfassen wollen (→ Operationalisierung).
- Legen Sie vorab Kriterien für die Nützlichkeit fest.
- Wählen Sie geeignete subjektive, objektive und psychophysiologische Methoden und Verfahren aus.
- Überlegen Sie, inwieweit Ihr zu entwickelndes kognitives System Leistungsdaten im Hinblick auf die Wirkung erfassen kann.

### KONZEPTION UND GESTALTUNG DES SYSTEMS

- Entwickeln Sie das kognitive System so, dass es funktioniert.

### UNTERSUCHEN DER WIRKUNG

- Entwickeln Sie einen Untersuchungsplan (dazu auch Hager, 2000) und entscheiden Sie welche und wie viele Personen (Stichprobe) und wie viele Gruppen Sie untersuchen müssen, um eine valide Aussage in Bezug auf „Which is better?“, „How good?“ oder „Why bad“ treffen zu können.
- Setzen Sie die ausgewählten Meßverfahren ein.
- Werten Sie die Daten bezogen auf die Evaluationskriterien und –fragen deskriptiv und inferenzstatistisch aus.

### BEWERTEN VON WIRKUNG UND NUTZEN

- Schlussfolgern Sie, inwieweit die von Ihnen festgelegten Wirkungsabsichten erreicht wurden.
- Schlussfolgern Sie, welcher Nutzen sich aus der statistischen Signifikanz, der praktischen Bedeutsamkeit und der dafür benötigten Ressourcen ergibt.
- Leiten Sie die Implikationen für (Ihre) weitere Forschung und für die Praxis ab.

## LITERATUR

Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32-64.

Ericsson, K.A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.

Gediga G. & Hamborg, K.-C. (2002). Evaluation in der Software-Ergonomie: Methoden und Modelle im Software-Entwicklungsprozess. *Zeitschrift für Psychologie*, 210 (1), 40 - 57.

Hager, W. (2000). Wirksamkeits- und Wirksamkeitsunterschiedshypothesen, Evaluationsparadigmen, Vergleichsgruppen und Kontrolle. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.) *Handbuch Evaluation psychologischer Interventionsmaßnahmen*. Standards und Kriterien (S.180-201). Bern: Huber.

Kluge, A. & Burkolter, D. (2012). Training for Cognitive Readiness: Research Issues and Experimental Designs. *Journal of Cognitive Engineering and Decision Making*. May 10<sup>th</sup>, 2012 DOI: 10.1177/1555343412446483

Kramer, A.F. & Weber, T. (2005, reprint from 2000). Applications of Psychophysiology to Human Factors. In J.T. Cacioppo, L.G. Tassinary & G.G. Berntson (Eds.), *Handbook of Psychophysiology* (S. 794-814). Cambridge University Press: Cambridge.

Letsky, M.P., Warner, N.W., Fiore, S.M., & Smith, C.A.P. (2008), (Eds.). *Macro-cognition in teams. Theories and methodologies*. Aldershot, United Kingdom: Ashgate.

Mittag, W. & Hager, W. (2000). Ein Rahmenkonzept zur Evaluation psychologischer Interventionsmassnahmen. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.) *Handbuch Evaluation psychologischer Interventionsmaßnahmen*. Standards und Kriterien (S.102-128). Bern: Huber

Parasuraman, R., Sheridan, T.B. & Wickens, C.D. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics- Part A: Systems and Humans*, 30 (3), 286-297

Reason, J. (2008). *The Human Contribution. Unsafe Acts, Accidents, and Heroic Recoveries*. Surrey: Ashgate

Salvendy, G. (2006). (Ed), *Handbook of Human Factors and Ergonomics*. Hoboken, NJ: Wiley & Sons.

Tsang, P.S. & Vidulich, M.A. (2006). Mental Workload and Situation Awareness. In G. Salvendy (Ed), *Handbook of Human Factors and Ergonomics* (S. 243-268). Hoboken, NJ: Wiley & Sons.

Vidulich, M.A. (2003). Mental Workload and Situation Awareness. Essential Concepts in Aviation Psychology. In P.S. Tsang & M.A. Vidulich (Eds.), *Principles and Practice of Aviation Psychology* (S. 115 -147). Mahwah, NJ: Lawrence Erlbaum

Westermann, R. (2002). Merkmale und Varianten von Evaluationen. Überblick und Klassifikationen. *Zeitschrift für Psychologie*, Vol. 210, 4-26

Weyers, B., Burkolter, D. Luther, W. & Kluge, A. (2012). Formal Modeling and Reconfiguration of User Interfaces for Reduction of Errors in Failure Handling of Complex Systems, *International Journal of Human-Computer Interaction*. January 11<sup>th</sup>, 2012, DOI: 10.1080/10447318.2011.654199

Wickens, C.D. & McCarley, J.S. (2008). *Applied Attention Theory*. Boca Raton: CRC Press.

Wickens, C.D. (2008). Situation Awareness: Review of Mica Endsley's 1995 Articles on Situation Awareness Theory and Measurement. *Human Factors*, 50, 397-403.