Department of Computer Science

Departmental Papers

University of Texas at El Paso Year 2001

Accounting for Domain Context in Evaluation

Meriem Chater, Université Toulouse 1
David Novick, University of Texas at El Paso

Proceedings of the Conference on Computer-Aided Learning in Engineering Education (CALIE 01), Tunis, Tunisia, November 8-10, 2001.

This paper is posted at DigitalCommons@UTEP
http://digitalcommons.utep.edu/cs_papers/6

Accounting for domain context in evaluation

MERIEM CHATER
Human-System Interaction Laboratory (LIHS), Université Toulouse 1

DAVID NOVICK
Department of Computer Science, University of Texas at El Paso

ABSTRACT

Work is situated activity. Taking into account human factors in evaluation involves considering not only users but also their contexts of use. Consequently, the evaluation of systems—from video-games to safety-critical interfaces—requires analysis of context to understand not only the effect of context on usability but also the impact of artifacts' usability on users' environments. In the case of safety-critical systems (SCS), errors (by users or designers) may threaten human lives.

To assess the degree to which interface evaluation methods currently account for context, we have used the research strategy taxonomy of McGrath as a framework for classifying existing evaluation methods of aviation domain and general HCI interactive systems. This framework enabled us to describe common grounds and key differences of methods used in HCI and SCS, and to highlight aspects of context that could be analyzed using each strategy.

For instance, characteristics of SCS, such as time-criticality, unpredictability and dynamics, emphasize the leading role of operational context on the remaining work context including physical or technical constraints defined by organizational, social, cultural and technical contexts which is not the case for general HCI.

KEY WORDS

Evaluation, context, usability, domain context, evaluation methods classification

Meriem Chater (meriemchater@yahoo.fr)

LIHS, Université Toulouse 1
1, Place Anatole France, F-31042 Toulouse cedex, France

David Novick (novick@cs.utep.edu)

Department of Computer Science, The University of Texas at El Paso
El Paso, Texas, 79968-0518

Introduction

Best practice in user-centered design of interactive systems involves iterative evaluation. In conducting these evaluations, it is necessary but not sufficient to find representative users. In fact, taking into account human factors in evaluation involves considering not only users but also their contexts of use. So to implement their evaluations, developers must design experimental protocols that cover as many of the relevant contexts of use as possible. These contexts may vary enormously. For example, depending on the domain application of the artifact to be evaluated, the context may be more or less safety-critical. But developers seeking to employ best practice in evaluation face a problem: how to characterize contexts of use in a way that is systematic enough to enable them to design appropriate evaluations. First, we discuss the importance of accounting for context in evaluation. Second, we propose a set of dimensions of context that would be useful for evaluation. Then using these dimensions, we review the extent to which interface evaluation methods in human-computer interaction, both generally and then specifically in the safety-critical domain of aviation, currently account for context.

1. Context relevance in evaluation

1.1 Why is it important to account for context in evaluation?

Context can be seen as a frame of reference, a space of shared knowledge (Brézillon, Pomerol, & Saker, 1998), explored and exploited by participants in the interaction. So to be usable, interfaces cannot be divorced from contexts as they depend on situations of use and furthermore, they may affect social context.

Situational validity. Usability qualifies user-system interaction in a context of use (ISO, 1998; Karat, 1997). Winograd & Flores (1986) emphasized that context, including social and linguistic environment, shapes interpretation and gives meaning to action. From the standpoint of Suchman (1987), context can be seen as a resource upon which users can draw. Therefore, the study of context is essential because users' actions are necessarily situated within particular spatial and temporal contexts that are crucial to the user' s interpretation of computer systems (Cooper, 1991). As context of use shapes usability, many authors, including Beyer & Holtzblatt (1999) and Bevan & Macleod (1994), recommended representative evaluations in context (choice of representative tasks, users; real world environment or, if not possible, a very close simulation) as well as context-oriented analysis methods.

Social context. Context study not only helps determining the effect artifact' s usability in its context of use, but also enables to identify the impact of the artifact on social, cultural, and organizational contexts (Brown & Duguid, 1994), and especially on user praxis (Sachs, 1995). Brown & Duguid (1994) introduced three dimensions for the study of context: the center (the artifact in use), the periphery (the context) and the border which is distinguishable if it plays a socially recognized role. They argued that designers need to understand the role of border resources and to negotiate their change with users. Sachs (1995) emphasized the need to study work practices before the design of a new artifact instead of relying on the organizational view; otherwise, practice will be the result of prescriptions plus workarounds.

Thus context and usability have a two-way relationship. Developers rely on context to help users interact with the system, and use of the system shapes the users' contexts. Understanding the effect of context of use on usability and the impact of usability on context creates a basis for systematic iteration of evaluations, including traceability, assessment and reuse.

1.2 How do contexts vary across domains?

As domains vary, they provide different contexts for interaction. For example, the definition of what is an "effective" human-computer interface is not necessarily the same in aviation as in the office. While major goals of interface usability include minimizing human information processing, minimizing cognitive demands on the user and avoiding errors, the relative importance of these goals differs greatly between the safety-critical domain of aviation and the non-safety critical domain of the office. In office automation, the goal is to avoid costly rework and schedule delays (Butler, 1996), errors or poor performance, lest unusable software "result in employee dissatisfaction, high staff turnover, absenteeism and tardiness" (Henderson, Podd, Smith, & Varela-Alvarez, 1995, p. 412). In contrast, in safety-critical domains the key issue is to avoid three classes of risk: vital, ecological and economic (Amalberti, 1995; McCarthy, Healey, Wright, & Harrison, 1997). Safety-critical systems (SCS) include nuclear power plants, aviation, air-traffic control and space missions. For SCS, the human performance that leads to incidents is significantly shaped by the

context (Woods, 1994). Moreover, context should be interpreted broadly in conducting accident investigations (McCarthy et al., 1997); in this sort of case, the notion of context should be extended to include factors such as deficiencies in training, lack of attention to the human-computer interface, and ignorance of work routines and practices. And evaluation of SCS requires greater attention to context than in the case of non-safety-critical systems because of the risks incurred in the event of error. Therefore, the evaluation should concern not only the isolated usability of the tested interface, but also the integrated effect of interaction with the interface on overall user activity.

2. Which context?

Given that domain contexts have significant variations, how can these contexts of use be characterized in a way that is systematic enough to enable them to design appropriate evaluations? We address this question by employing a decomposition of the concept of context. Chater (2000) highlighted the main categories of context to be considered in evaluation: (1) work context, (2) organizational, social, cultural, technical (OSCT) context, and (3) evaluation context (see Figure 1). We examine each part of this multidimensional structure of context in turn.

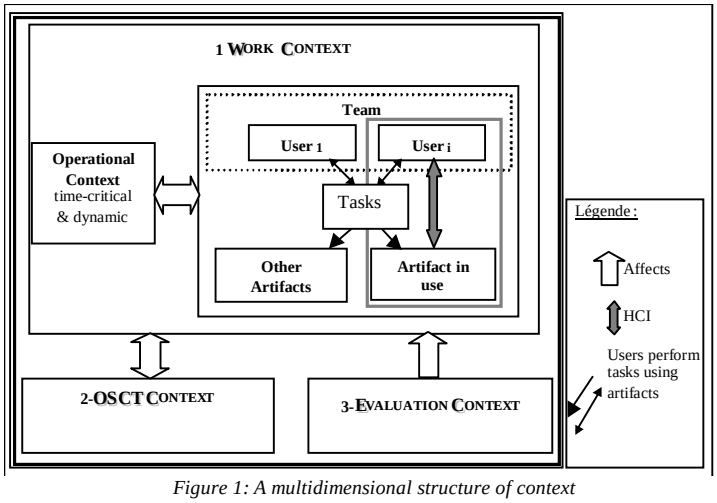

Figure 1: A multidimensional structure of context

Work context can be either *static*, including users, tasks, the artifact in use, the other artifacts (such as for instance operational procedures, or other interfaces not currently in use). It may also be *operational* described by the activity history, ongoing activity, work phase, systems states, occurred and occurring events, and the kind of situation in progress, such as normal, abnormal, and emergency.

OSCT (Organizational, Social, Cultural, Technical) context is defined by international, and national safety rules, national and organizational culture, organizational knowledge, memory, strategies and goals (safety, efficiency, economy, performance), the task domain, work, product standards, rules, standard operating procedures, and latent errors (Reason, 1993). Organizational latent errors may include, for instance, decision errors at the management level, training problems; technical latent errors may include for instance artifacts design errors or inconsistencies and equipment maintenance failures. Organizational, social and cultural context may involve people' s judgment at these different levels as well as users' accountability. (See, e.g., the analysis of McCarthy, Healey, Wright, & Harrison (1997) of the relationships between work activity and accountability). These factors may influence users' behaviors accordingly.

Evaluation context is described by the evaluation characteristics, including evaluation objectives, experimental protocol (such as scenarios, evaluation methods, evaluation criteria, simulation tools and environment, and test users). These data can be used to assess the realism (the gap between evaluation implementation and work context), the generalizability and the precision of the evaluation and its results.

3. Context and evaluation methods

Most of the evaluation methods used in the safety-critical domain of aviation, are methods adapted from general HCI practice (e.g. Irving, Polson, & Irving, 1994; Palmer, Rogers, Press, Latorella, & Abbott, 1995). Irving et al. (1994) suggested that modern automated offices and advanced-technology cockpits are

comparable in the sense that both office workers and pilots supervise complex automated systems. They concluded, therefore, that evaluation techniques developed for human-computer interaction could be used and adapted to the aviation domain. Similary, the general-purpose cognitive walkthrough evaluation technique has been adapted to operating procedures for commercial aircraft (Novick, 1999; Novick & Chater, 1999) The adaptation of methods lies essentially in their implementation, in the means used and, most important, in the data they analyze. These data include domain and context knowledge as well as critical issues (Chater & de Brito, 1999).

In order to assess to what extent evaluation methods account for context, we have classified HCI and SCS evaluation techniques (see figure 2) using McGrath' s (1995)framework (see Chater, 2000). McGrath distinguished quadrants corresponding to four research strategies: field strategies, experimental strategies, respondant strategies, and theoretical strategies. McGrath' s taxonomy sets he limits for each research strategy of: (a) the generalizability of the evaluation method results (b) the precision of measurement of the behaviours being studied (c) the realism of the situation within which the evidence is gathered.

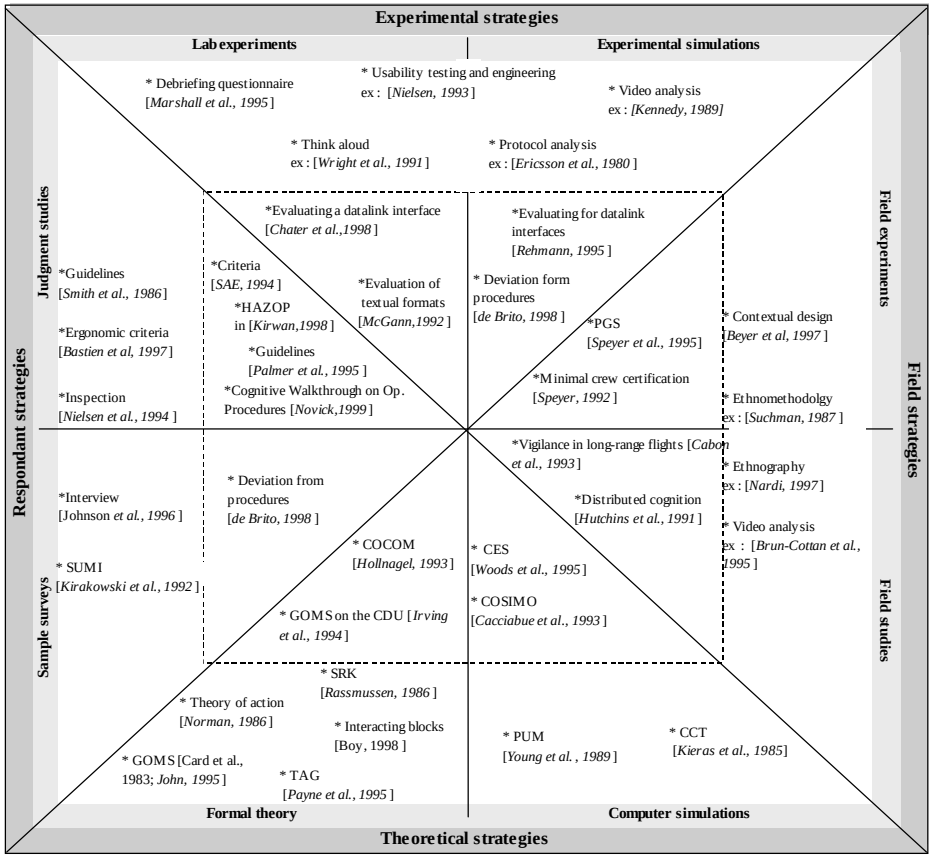

Figure 2. Classifying instances of HCI and SCS evaluation methods in McGrath's adapted research taxonomy (methods used in HCI are in the periphery, methods used in SCS are in the center)

These methods are drawn from the general HCI literature (Beyer & Holtzblatt, 1999; Boy, 1998; Brun-Cottan & Wall, 1995; Card, Moran, & Newell, 1983; Ericsson & Simon, 1980; John & Packer, 1995; Johnson & Nardi, 1996; Kennedy, 1989; Kieras & Polson, 1985; Kirakowski, Porteous, & Corbett, 1992; Marshall & Novick, 1995; Nardi, 1997; Nielsen, 1993; Nielsen & Mack, 1994; Norman, 1986; Payne & Green, 1986; Rassmussen, 1986; Scapin & Bastien, 1997; Smith & Mosier, 1986; Suchman, 1987; Wright & Monk, 1991; Young, Green, & Simon, 1989) and the SCS literature (Cabon, Coblentz, Mollard, & Fouillot, 1993; Cacciabue & Kjaer-Hansen, 1993; Chater, Sikorski, & Boy, 1998; de Brito, 1998; Hollnagel, 1993; Hutchins & Klausen, 1991; Irving, Polson, & Irving, 1994; Kirwan, 1998; McGann, Morrow, Rodvold, & Mackintosh, 1998; Novick, 1999; Palmer, Rogers, Press, Latorella, & Abbott, 1995; Rehmann, 1995; SAE, 1994; Speyer, 1992; Speyer & Elsey, 1995; Woods & Roth, 1995; Wynn, 1991). McGrath' s categories can be in turn classified according to the place of evaluation: real world context (field strategies), simulated context (experimental strategies), and out-of-context (theoretical and respondant strategies).

4. Context perspectives and representation

Given the kinds of context and the kinds of context-based evaluation techniques presented in Sections 3 and 4, how can these classification schemes be used? Perspectives of context depend on evaluation objectives, on the expertise, and skills of the evaluators, on data-collection methods used, on domains in which they are studied, and on the range of contexts provided by the place of evaluation (real-work contexts, simulated contexts, and out-of contexts). We contrast each of these three kinds of context in terms of their effects on evaluation, particularly in terms of how context is represented.

More complete evaluations of work in context can be made in a real-world context, as the work and OSCT contexts already exist and do not need to be reproduced. However, it is essential to take into account bias derived from subjective interpretations and behavior translations from the perspectives of users, analysts and designers. Experimental simulations involve a subset of real-world contexts and need to be reproduced, and sometimes this kind of bias will be introduced. Controlled experimental simulations recreate users' environments and enable the elicitation of certain behaviors, attitudes, stresses, errors, and actions with respect to working conditions, events, and artifacts in use. Baars (1980), cited in (Reason, 1993) emphasized the interrelationship between field studies and experimental simulations, noting that "Without naturalistic methods, experimental research may become narrow and blind; but without experimental research, the naturalistic approach is in danger of being superficial and uncertain." (p. 39). Real-world and simulated context strategies emphasize the analysis in the dynamic, time-critical, complex operational environment, and its effect on user's activity situated within social, cultural, and organizational contexts background. While the focus of field studies in HCI generally is primarily organizational and social, their primary focus in safety-critical systems is more operationally oriented; the dynamic, time-critical, complex features of the operational domain drive the activity, which is constrained by organizational procedures and rules. Community issues such as team coordination, task division are essential features of activity in these settings. Representations of context are usually informal.

In out-of-context strategies, the kind of context analyzed (organizational, social, cultural, technical, work context) depends on the evaluation objectives and on the evaluators. In general, evaluations are performed based on fairly representative scenarios of use that restore part of the context focused on the tested artifact and tasks. Representation of context is thus informal and tacit in the case of respondent strategies, and in the case of theoretical strategies, models can represent part of contexts formally, depending on evaluation objectives, and the experience and skills of models' designers. Respondent strategies are based on analysts' and users' representations of context. In contrast, theoretical strategies rely on designers' representations of context. In general HCI, represented contexts in the case of respondent strategies and theoretical strategies deal more with work context in terms of artifact, tasks, and users characteristics. In the case of SCS, these techniques emphasize the leading role of the dynamic, time-critical, and risky operational context as constrained by organizational factors such as rules and operating procedures. Concerning the evaluation context, we have to consider not only the gaps between the pictured model of context and the real world but also the gaps between contexts' interpretation.

The use of redundant or complementary methods to gather more reliable data (Mackay & Fayard, 1997; McGrath, 1995) is useful in the case of office-like systems and required in the case of safety-critical domains. In order to maximize the validity of research results, Mackay and Fayard recommend using a triangulation approach accross the disciplines that make up human-computer interaction: psychology, sociology, anthropology, ergonomy and computer science.

5. Conclusion

As Suchman (1987) observed, work is situated activity. Consequently, the evaluation of systems—from video-games to safety-critical interfaces—requires analysis of context to understand not only the effect of context on usability but also the impact of artifacts' usability on users' environments, especially in the case of SCS where errors (by users or designers) may threaten human lives. In this paper, we classified representative techniques from general HCI and the aviation domain using McGrath's research taxonomy framework. This framework enabled us not only to describe common grounds and key differences of methods used in HCI and SCS, but also to highlight aspects of context that could be analyzed using each strategy. Domain knowledge determines at least part of the context to be considered in evaluating a user

interface. For instance, characteristics of SCS, such as time-criticality, unpredictability and dynamics, emphasize the leading role of operational context on the remaining work context including physical or technical constraints defined by organizational, social, cultural and technical contexts. In contrast, complexity and avoidance of and recovery from errors, more connected with sub-context of artifacts and their integration in the working environment, point up the importance of organizational, social, cultural and technical contexts as moral or psychological constraints.

References

Amalberti, R. (1995). Paradoxes de la sécurité des grands systèmes à risques, le cas de l' aéronautique.*Performances Humaines & Techniques, 78*(Sept-Oct), 45-55.

Baars, B.J. (1980). Eliciting predictable speech errors in the laboratory. In V. Fromkin (Ed.), *Errors in linguistic performance: slips of the tongue, ear, pen, and hand.* New York: Academic Press.

Bevan, N., & Macleod, M. (1994). Usability measurement in context. *Behaviour & Information Technology, 13*(1-2), 132-145.

Beyer, H., & Holtzblatt, K. (1999). Contextual design. *ACM interactions, 6*(1), 32-49.

Boy, G. (1998). *Cognitive Function Analysis*. Stamford, CT: Ablex Publishing Corporation.

Brézillon, P., Pomerol, J.-C., & Saker, I. (1998). Contextual and contextualized knowledge: an application in subway control. *Int. J. Human-Computer Studies, 48*(3), 357-373.

Brown, J.S., & Duguid, P. (1994). Borderline issues: social and material aspects of design. *Human-Computer Interaction, 9*(1), 3-36.

Brun-Cottan, F., & Wall, P. (1995). Using video to re-present the user. *Communications of the ACM, 38*(5), 61-71.

Butler, K.A. (1996). Usability Engineering Turns 10. *ACM interactions, 3*(1), 58-75.

Cabon, P., Coblentz, A., Mollard, R., & Fouillot, J.P. (1993). Human vigilance in railway and long haul flight operations. *Ergonomics, 36*(9), 1019-1033.

Cacciabue, P.C., & Kjaer-Hansen, J. (1993). Cognitive modelling and human machine interactions in dynamics environments. *Le Travail Humain, 56*(1), 1-26.

Card, S.K., Moran, T.P., & Newell, A. (1983). *The psychology of Human Computer Interaction*: Lawrence Erlbaum.

Chater, M. (2000). *L'évaluation contextualisée et sa documentation, vers un outil de conception centrée sur l'homme. Application au domaine aéronautique.* Thèse de doctorat en informatique, Université de Toulouse 1 et EURISCO.

Chater, M., & de Brito, G. (1999). Adapting evaluations methods to aeronautics, *CHI'99 workshop "HCI in domains, common ground and key differences"*, May 15-20, 1999, Pittsburgh, PA USA.

Chater, M., Sikorski, S., & Boy, G. (1998). An Experimental Method to Assess the Usability of Flight Deck Interfaces. G. Boy & J.-M. Robert (Eds.), *International Conference on Human-Computer Interaction in Aeronautics HCI-Aero'98* (pp. 235-240), Montréal, Québec.

Cooper, G. (1991). Context and its representation. *Interacting with Computers, 3*(3), 243-252.

de Brito, G. (1998). Study of use of Airbus Flight-deck procedures and perspectives for operational documentation. G. Boy & J.-M. Robert (Eds.), *International Conference on Human-Computer Interaction in Aeronautics HCI-Aero'98* (pp. 195-202), Montréal, Québec.

Ericsson, K.A., & Simon, H.A. (1980). Verbal Reports as Data. *Psychological Review, 87*(3), 215-251.

Henderson, R., Podd, J., Smith, M., & Varela-Alvarez, H. (1995). An examination of four user-based software evaluation methods. *Interacting with Computers, 7*(4), 412-432.

Hollnagel, E. (1993). Models of cognition: procedural prototypes and contextual control. *Le travail humain, 56*(1), 27-51.

Hutchins, E., & Klausen, T. (1991). Distributed Cognition in Airline Cockpit., *Cognition and Communication at Work. Cambridge University Press.*

Irving, S., Polson, P., & Irving, J.E. (1994). A GOMS Analysis of the advanced Automated Cockpit, *CHI94 Celebrating Interdependence* (pp. 344-350).

ISO (1998). ISO 9241-11:1998 (F) : Exigences ergonomiques pour travail de bureau avec terminaux à écrans de visualisation (TEV). Partie 11: Lignes directrices concernant l' utilisabilité. Paris: Norme internationale édité par l' AFNOR.

John, B.E., & Packer, H. (1995). Learning and using the cognitive walkthrough method: a case study approach, *CHI'95* (pp. 429-436), Denver, Colorado, USA.

Johnson, J.E., & Nardi, B.A. (1996). Creating presentation slides: a study of user preferences for task-specific versus generic application software. *ACM Transactions on Computer-Human Interaction, 3*(1), 38-65.

Karat, J. (1997). User-centered software evaluation methodologies. In M.G. Helander & P.V. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (pp. 689-704): North-Holland Elsevier.

Kennedy, S. (1989). Using video in the BNR usability lab. *Sigchi Bulletin, 21*(2), 92-95.

Kieras, D., & Polson, P.G. (1985). An approach to the formal analysis of user complexity. *Int. J. Man-Machine Studies, 22*, 365-394.

Kirakowski, J., Porteous, M., & Corbett, M. (1992). How to use the Software Usability Measurement Inventory: the user' s view of software quality, *Proceedings European Conference on Software Quality*.

Kirwan, B. (1998). Human error identification techniques for risk assessment of high risk systems--Part 1: review and evaluation of techniques. *Applied ergonomics, 29*(3), 157-177.

Mackay, W.E., & Fayard, A.-L. (1997). HCI, natural science and design : a framework for triangulation across disciplines, *Designing Interactive Systems: processes, practices, methods, and techniques (DIS'97)* (pp. 223-234).

Marshall, C.R., & Novick, D.G. (1995). Conversational effectiveness in multimedia communications. *Information Technology & People, 8*(1), 54-79.

McCarthy, J.C., Healey, P.G.T., Wright, P.C., & Harrison, M.D. (1997). Accountability of work activity in high-consequence work systems: human error in context. *Int. J. Human-Computer Studies, 47*(6), 735-766.

McGann, A., Morrow, D., Rodvold, M., & Mackintosh, M.-A. (1998). Mixed-Media Communication on the flight-deck: A comparison of Voice, Datalink, and Mixed ATC environments. *The International Journal of Aviation Psychology, 8*(2), 137-156.

McGrath, J.E. (1995). Methodology matters: doing research in the behavioural and social sciences. In R.M. Baecker, J. Grudin, & S. Greenberg (Eds.), *Readings in Human Computer Interaction: Toward the Year 2000*: Morgan Kaufmann Publishers, Inc.

Nardi, B.A. (1997). The use of ethnographic methods in design and evaluation. In M.G. Helander & P.V. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (pp. 361-366): North-Holland Elsevier.

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

Nielsen, J., & Mack, R.L. (1994). *Usability inspection methods*: John Wiley & Sons, Inc.

Norman, D.A. (1986). Cognitive engineering. In D.A. Norman & S.W. Draper (Eds.), *User-centered system design New perspectives on Human-Computer Interaction* (pp. 31-62): Lawrence Erlbaum Associates.

Novick, D. (1999). Using the cognitive walkthrough for operating procedures. *Interactions, 6*(3), 31-37.

Novick, D., & Chater, M. (1999). Evaluating the design of human-machine cooperation: The cognitive walkthrough for operating procedures., *Proceedings of the Conference on Cognitive Science Approaches to Process Control (CSAPC 99)* (pp. 21-26), Villeneuve d' Ascq, FR.

Palmer, M.T., Rogers, W.H., Press, H.N., Latorella, K.A., & Abbott, T.S. (1995). *A Crew-Centered Flight Deck Design Philosophy for High-Speed Civil Transport (HSCT) Aircraft*: NASA Langley.

Payne, S.J., & Green, T.R.G. (1986). Task-action grammars: a model of the mental representation of task languages. *Human-Computer Interaction, 2*, 93-133.

Rassmussen, J. (1986). *Information processing and human machine interaction: an approach to cognitive engineering*: Elsevier Sciencs Publishers.

Reason, J. (1993). *L'erreur humaine*: Presses Universitaires de France.

Rehmann, A.J. (1995). *Handbook of Human Performance Measures and Crew Requirements for Flightdeck Research*: FAA.

Sachs, P. (1995). Transforming work: collaboration, learning and design. *Communications of the ACM, 38*(9), 36-44.

SAE (1994). Human Engineering: SAE G-10K Subcommittee, Flight Deck Information Management of Committe, Aerospace Behavioral Engineering Technology (ABET).

Scapin, D.L., & Bastien, J.M.C. (1997). Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & Information Technology, 17*(4-5), 220-231.

Smith, S.L., & Mosier, J.N. (1986). Guidelines for designing user interface software. Massachusetts, USA: The MITRE corporation.

Speyer, J.J. (1992). *ATA 102 Minimum crew certification. Cockpit and flight analysis and evaluation*. Toulouse: Airbus Industrie.

Speyer, J.J., & Elsey, A. (1995). Towards the integration of pilot guard systems for monitoring attentiveness in flight, *HMI-AI-AS'95 Fifth International Conference on Human-Machine Interaction and Artificial Intelligence in Aerospace*, Toulouse, France.

Suchman, L.A. (1987). *Plans and situated actions: The problem of human-machine communication*: Cambridge University Press.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition*: Ablex.

Woods, D., & Roth, E. (1995). Symbolic AI computer simulations as tools for investigating the dynamics of joint cognitive systems. In J.M. Hoc & E. Hollnagel (Eds.), *Expertise and technology* (pp. 75-92). New Jersey: LEA.

Woods, D.D. (1994). Observations from Studying Cognitive Systems in Context, *Annual Conference of the Cognitive Science Society*.

Wright, P.C., & Monk, A.F. (1991). The Use of Think-Aloud Evaluation Methods in Design. *Sigchi Bulletin, 23*(1), 55-57.

Wynn, E. (1991). Taking Practice Seriously. In J. Greenbaum & M. Kyng (Eds.), *Design at work: cooperative design of computer systems* (pp. 45-64): Lawrence Erlbaum Associates, Inc.

Young, R.M., Green, T.R.G., & Simon, T. (1989). Programmable User Models for Predictive Evaluation of Interface Designs, *CHI'89* (pp. 15-20), Austin, Texas, USA.