activités
revue électronique

Towards a theory and method for usability evaluation of complex human-technology systems

Leena Norros & Paula Savioja

Systems Research Centre, VTT Technical Research Centre of Finland, P.O. Box 1000, 02044 VTT, Finland

leena.norros@vtt.fi, simo.savioja@poliisi.fi

Abstract

One of the central roles of ergonomic research is to provide a normative basis for the evaluation of the appropriateness of artefacts used in various human activities. In this paper a new activity theory based approach, the Core-Task Analysis is introduced. Inspired by that, an evaluation method, Contextual Assessment of Systems Usability, was developed for the analysis of the appropriateness of complex information and control systems. The basic structure of the method is described and the indicators for assessment briefly described.

Keywords:

Activity theory, Core-Task Analysis, systems usability, habits of action, functional modeling, complex systems

1.- Ergonomic evaluation of complex systems

Ergonomics is an interdisciplinary science of human conduct. It takes into account the fact that human beings use tools in interacting with the world, and, is especially focused to shape technologies to fit human purposes, conditions and values. Hence, ergonomics has constructive intentions and faces the difficult task of bridging between theory and practice. Scientific methodology and methods are therefore needed that acknowledge the epistemic value of practice and accept new types of interventive and participative methods. The approach that we advocate in this paper aims at these objectives, and draws on theories that share them, i.e. the cultural historical theory of activity (Vygotski, 1978) and the pragmatist theory of habit (Peirce, 1998).

One of the central roles of ergonomic research is to provide a normative basis for the evaluation of the appropriateness of artefacts that are used in various human activities. The rapidly increasing implementation of information and communication technologies (ICT) into all spheres of human activity has made this technology a universal medium that deeply effects people's interaction with their environments. In this connection also the process of design has changed in many ways. One of the new features is the much deeper involvement of the user in the design. In this situation there is a great challenge to the ergonomics science to be better integrated in the construction of the ICT-based intelligent objects and environments, but at the same time to facilitate reflection what is good development of technology in specific cases.

In this paper a theoretical approach to analysis of human conduct, the Core-Task Analysis, and a method for an integrated evaluation of ICT-based artefacts is introduced. These conceptual tools have been developed to meet the assessment needs of information and control systems used in complex work with high usability and reliability requirements, for example work in modern process and transport industry. It is claimed, however, that the approach could also be helpful in the evaluation of the appropriateness and acceptability of knowledge society infrastructures and intelligent environments.

Ergonomics aims both to minimise negative effects of technology and to maximise the creative role of the human. But it reaches even further as ergonomics in design strives for creating new possibilities and uses of living. The deeper the influence of ergonomics is on shaping the human and his environment, the more important it becomes that the ergonomic research and practice is contextually oriented. It must appreciate the content of behavior and the meaning of technologies for the users if it intends to succeed in its role.

2.- The core task analysis in ergonomic research

2.1.-Taking the context into account

Our way of accomplishing contextual ergonomics research has developed gradually in field studies in the industry, fortified with experimental studies on full-scope training simulators (Norros, 2004). We use the concept of Core Task to define the content and context of the studied activity, normally a complex and technologically mediated work activity. The core task denotes the objectives, goals and result critical intrinsic constrains of the work *of* an organisation, or that *in* an organisation. The analysis of the core content and intrinsic constraints of work is important because, by definition, they have to be taken into account in all situations, the content is not necessarily evident to the actors, and because the content may change.

When defining the content of the core task we start from the methodological notion that the human-environment interaction (transaction) should be conceived as one functional system. Such a system explains behavior as being structured according to its results and with regard to the constraints and possibilities of maintaining appropriate action. Instead of comprehending actions as simple linear causal relationships of transforming information across the border of the two separate systems (human and environment) – as the prevailing information processing approach and prescriptive task analyses assume – researchers aim to discover processes and phenomena of entrainment or resonance within the human-environment system. Such a change of framing the target would necessitate, further, that not only events as causes for action, but also reasons for action and their meaning could considered as adequate explanations for human conduct.

The ecological psychology of Gibson is one of the prominent alternatives to the cognitivist information processing approach. The concept of *affordance* has been adopted and used widely in describing the possibilities that the environment and media provide for meaningful usage. We make use of this notion that defines the environment in its connection to the human user. We complement '*affordance*' with the symmetric English notion of *prehensility.* This concept – known also in philosophy – defines the human subject in connection to the environment and indicates his capability to grasp the environment.

2.2.-Modelling the potential to act in a situation

Both affordance and prehensility denote existing *potential* in the system to become involved in a dynamic process of activity. We developed concepts and concrete measures to define this potential.

2.2.1.- Objectives and intrinsic constraints of the work domain

It was mentioned in the introduction that the cultural historical theory of activity is one of the two main theories on which we draw. One well known adaptation of this theory is the activity system approach developed by Engeström and his colleagues (Engeström, 1999). We take advantage of the model of activity system that conceives the work domain as societal, culturally and instrumentally mediated activity. The further advantages of the model are that it facilitates discovery of the tensions

within of the work domain and thereby draws attention not only to the history, but also to the future of the system.

The activity system model is culturally oriented and neglects the material and physical constraints of work. For comprehending these we utilise another modelling technique. It was developed in a human factors engineering environment and focuses on the intrinsic constraints of coping with complex domains. The constraints are defined with the aid of a functional domain modelling (FDM) approach that produces a functional breakdown of the system. The functions are related to the safety, efficiency, and health-relevant aspects of the domain. FDMs describe the domain on a general level.

Modelling should be carried out in multidisciplinary workshops to which all the possible stakeholders are invited to take part in. The team should include experts in operations, training, process design and human factors.

2.2.2- Modelling the core-task demands

As opposed to domain modelling, the core-task modelling (CTM) takes an operational view to the domain. The activity produces a similar deliverable to the FDM but in this case the functions necessary to maintain the activity within the result-critical boundaries of safety, efficiency or health are modelled from an operative perspective. Core-task model describes the functional task of the operators and provides a set of (psychological) core-task demands. In the analysis we utilise a tool that translates the intrinsic constraints into the categories of dynamicity, complexity and uncertainty, main attributes that effect the controllability of the domain. The idea is that these attributes are interdependent and that human skills, knowledge and collaboration are resources to cope with them. With this conceptual tool we are able to produce a set of human resource oriented core-task demands that relate to the particular dynamicity, complexity, and uncertainty features of the domain. It should be noted that the modelling of the core-task demands makes use of performance data but still deals with the potential aspect of activity by abstracting the demands that coping with the environment puts. Analysis of core-task demands has been tested in several domains including nuclear power operations, see piloting, anaesthesia and barley growing.

2.2.3.- Habits

The final form of modelling included in the core-task analysis is the performance-based definition of the capabilities and ways that experts grasp the environmental constraints and possibilities in their work - which may be perceived as demands of the work - and make use of them. We consider, to what extent and according to which logic, the operators take the constraints and possibilities into account. Again, data of actual performance is used to define the potential for action.

The concept of *habit* is used as the theoretical construct in the modelling. The more detailed arguments for utilising this concept are given elsewhere (Norros, 2004). In this context it should suffice to state that the pragmatist notion of habit refers to the potential not the actual aspect of action and crystallises well the idea of generalised psychological structure to cope, both reflectively and pre-reflectively, with the uncertainties of the environment. The repetitive nature of habit has its origin not only in the function of habit in creating continuity but also in expressing meaning and style (Peirce, 1958).

The aim of the analysis of habits is to construct a set of behavioural markers, indicators of practice, that thanks to the analysis may be considered as interpretants of meaningful signs in the environment. There are always different possible interpretations of the situations. We make use of the variance and grade different responses with regard to how well and according to which logic the affordances of the work, the core task, appear to be taken into account. So far we have created and published indicators for practice for nuclear power plant operator work, anaesthesia and sea piloting (Klemola, & Norros 2001; Norros, & Nuutinen, 2005; Nuutinen, & Norros, 2007).

2.3.- Analysis and assessment of the actual behaviour in a situation: Activity, Action, Operation

The ergonomic analysis is, of course, aiming at understanding the actual real activity of people. The analysis of the environmental and human potential for action is a prerequisite to do this, and provides a reference to evaluate activity, as we indicated above.

As in the modelling phase, also here we make use of the cultural historical theory of activity. A.N. Leont'ev was interested in the psychological structure and construction of human activity both in a historical perspective and in actual situations. He proposed the three level concept of activity, including the levels of the activity, action and operations (Leont'ev, 1978). These levels are indicated in the model below (see Figure 1).

Figure 1: The structure of actual activity in the background of potential for action (Adapted from earlier work, Norros, 2004)

The model in Figure 1 aims at guiding an analysis and evaluation of actions (with bolded borders) embedded in their societal and material context. It may appear that actions in a particular work are evident. This may be true in the sense of *what* people do, but not in the sense that one would understand what is their *meaning*. By observing the environmental potentials and the perceived shared objectives we may clarify the meaningful activity in which the actions are embedded. Furthermore by observing the conditions and the routines of grasping them we may elaborate the meaning of the operations. Via inquiring the personal sense of operations to *particular* subjects, and the personal sense of the activity to them – with reference to the affordances, prehensilities and indicators defined in the modelling phase – we identify the real *habits of action* and *orientation* of these subjects. Hence, we may draw conclusions of the logic on which particular people act. As a result we may define *how* people act, i.e. what is the personally meaningful modus or style of actions that people have appropriated in the community of practice. This provides basis for understanding actions. It also allows explaining the situational courses of action. As people tend to repeat what they have understood for meaningful, the orientations and habits of action also predict behavior in further sufficiently corresponding situations.

3.- Application of Core-task analysis – approach to evaluation of the usability of complex systems

The core-task analysis theory and approach has been used in the construction of an evaluation method for validation of complex information and control systems. The method has been labeled the Contextual Assessment of Systems usability, i.e. the CASU method (Savioja, & Norros, 2004; Norros, & Nuutinen, 2005).

The CASU method is developed for use in the integrated system validation of nuclear power control room modifications. Integrated system validation is an evaluation using different types of performance-based evaluations to ensure that the design is consistent with performance requirements and

acceptably supports safety operation of the plant (O'Hara, Higgins, Persensky, Lewis, & Bongarra, 2002). The essence of CASU method is depicted in Figure 2. It consists of four separate phases.

3.1.-Modelling

The purpose of the modelling phase is, first, to ensure comprehensive and representative evaluation; second, to elicit the critical parts of the domain that need to be considered in the tests and specially in the specific scenarios, and third, to elicit domain information to help form the performance indicators.

The modelling phase is critical for the validity and quality of the evaluation results. It outlines the basis for the evaluation by producing the reference stating what good process control activity in a given operational situation is. In addition to FDMs and CTMs situation specific models of the affordances of the environment are also created. We have developed a particular technique of designing functional situation models (FSM), i.e. scenarios. According to this, the more global domain model is given a situational meaning. Situational models are always conducted with an expert of the domain who has the knowledge of the impacts of basic functions of the system and process (portrayed in FDMs) in that specific situation,

The important outputs of the modelling phase are the measures and criteria used in the control room evaluation.

3.2.-Data Collection

The second phase is the data collection phase. In this the actual simulator runs (or in the future applications probably real work situations) are run with comprehensive operator crews. The activity of the crew is observed and video and other performance and interview data is collected.

Our data collection tool pack consists currently of the following methods:

— Orientation interviews;
— Observations and logs of human-system interactions;
— Task Load and Complexity measurement (TLX);
— Stimulated process tracing interview.

The data collection methods are partly developed by ourselves and partly we utilise established methods.

3.3.-Data Analysis

The analysis of the data collected in the performance evaluation sessions comprises of three intertwined phases:

1– In the human-system interaction analysis the most prominent task is to construct and analyse the course of action. A timeline of the process phases that take place during the operational situation is constructed. The operational events and the events of the process are presented in a chronological order on a timeline. After that the observations, actions and communication of the operators are added to the timeline. The goal is to reconstruct the actualised course of action in this particular situation.

2- In the analysis of the process control performance we are interested in two aspects. In the first case we focus on the evaluation of the adequacy (successfulness and pertinence) of the operators' process control performance. We assess the operators' recognition of the state of the process and their control over the process. To accomplish the evaluation, the process control task was broken up to three sub-tasks: identification of the disturbance in the state of the process, stabilisation of

the process, and recognition of the causes of disturbance.

The second aspect of process control performance relates to the subjective assessment of operators of the task performance. So far we have been exploiting the well-known NASA-TLX measurement questionnaire. In the future we hope to develop the subjective analysis further.

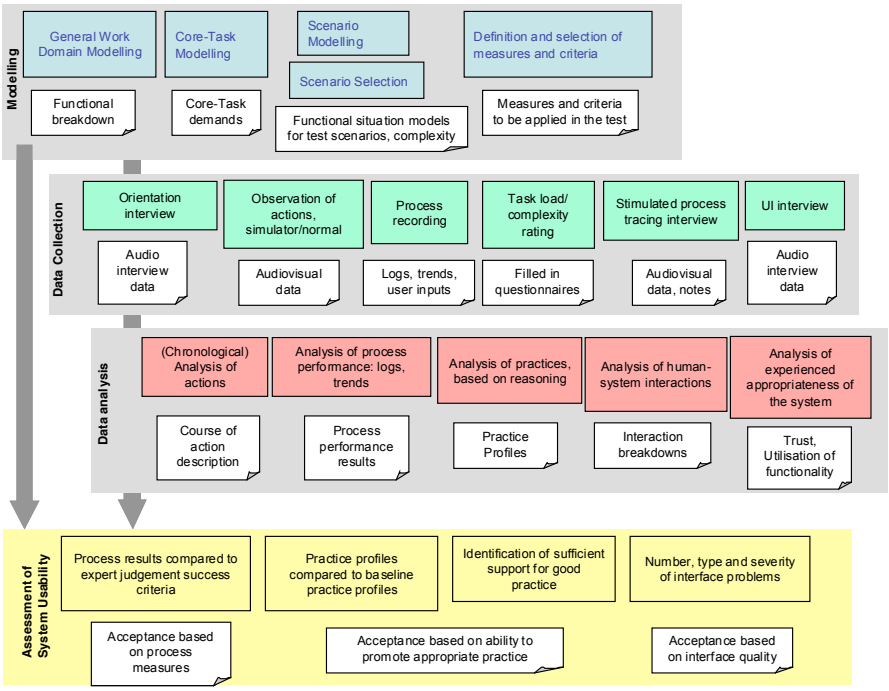

Figure 2: Phases and the deliverables of evaluating human system interaction of complex systems

3- In the third analysis phase, the work practice analysis, the objective is to analyse operators' habits of action (working practices). We clarify how operators use the available tools, e.g. information, operational methods and procedures when they interact with the process, with each other or in the control of their own personal resources. These different types of interactions serve definite situated goals and tasks:

— Operator-process: monitoring and situation assessment, planning and execution of control, stabilisation or testing actions, diagnosis and fault detection, definition of operational goals or executing planned changes in operational stage (so called primary tasks);

— Operator-operator: management, co-ordination, collaboration, communication;

— Operator-Self: control of the use of tools (so called secondary tasks), focusing and attention, feedback, checking, learning.

The analysis is based on indicators that relate to the above listed tasks and portray how the core-task demands of the operator work are taken into account in accomplishing the tasks.

3.4.-The assessment of systems usability

The assessment phase is the part of the human factors evaluation in which the acquired results are compared to the acceptance criteria.

As Figure 2 indicates the assessment is to be accomplished from three different perspectives. First it is evaluated whether the human-system interaction is "good" in the sense that the process may be held in acceptable boundaries. The second aspect of evaluation refers to the "good" of the work performance and the practices of the operators. This evaluation focuses especially on the adequacy,

adaptability and core-task orientedness of performance. These attributes of the human-system interaction are considered relevant for strengthening the development of the activity. Finally, it is also necessary to evaluate whether the human-system interaction has acceptable interface qualities.

4. Conclusions

The motive in the development of the Core-Task Analysis –based CASU method is to improve present conceptions of what is "good" design and how to "measure" it. Smart objects, environments and infrastructures of the knowledge society should meet the new requirements of *Systems Usability* that we proposed as a central concept of the method.

The notion of usability mediates well the basic requirement for human-centred design, i.e. products or tools should promote human possibilities to achieve goals in a defined context of use. We propose, however, that the notion should be extended in the following five aspects:

— *Integrative*: Integration is needed in evaluation between the two purposes of evaluation: innovation and acceptance of design solutions, between the different design phases, and between the different levels of detail of the design.

— *Comprehensive*: There is a need that artefacts should be considered as part of a meaningful activity. The cultural-historical theory of activity provides a holistic approach and a context to evaluate how the system promotes the different actions or tasks accomplished in the organisation and how it facilitates the objectives and activity of the organisation.

— *Context-dependent.* A widely accepted requirement for human-centred design and usability studies is that evaluations need to be made in the context of use. Our proposal is to use a new Core-Task Analysis methodology to accomplish this requirement.

— *Performance-based.* The quality of the artefact is typically checked against available guidelines or standards. We see that there is a need to improve the performance-based evaluation of technologies and see draw on the three generic functions of a tool or medium in action, as an instrument, as a psychological tool and as medium of communication and consciousness.

— *Semiotic Perspective.* Our attempt is to extend usability evaluation to consider how the new forms of representation that the digital media makes available may best be connected to the content that the process necessitates for appropriate human process control activity.

Acknowledgements

The work described in the has been financed by the Finnish research project on nuclear safety (SAFIR) and by VTT. We express our thanks to M.Sc (Tech) Leena Salo and Dr. (Psych) Jari Laarni who have collaborated with us in the research.

Referencing

Norros, L.L., & Savioja, P.J. (2007). Towards a theory and method for usability evaluation of complex human-technology systems. *@ctivités*, 4(2) pp. 143-150, http://www.activites.org/v4n2/v4n2.pdf

References

Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engeström, R. Miettinen, & R.-L.Punamäki [Eds.], *Perspectives in Activity Theory* (pp. 19-38). Cambridge: Cambridge University Press.

Klemola, U.-M., & Norros, L. (2001). Practice-based criteria for assessment the anaesthetists' habits of action.Outline for a reflexive turn in practice. *Medical Education*, 35, 455-464, 2001.

Leont'ev A.N. (1978). *Activity, Consciousness, and Personality*. Englewood Cliffs: Prentice Hall.

Norros, L. (2004). *Acting under Uncertainty. The Core-Task Analysis in Ecological Study of Work*. Espoo: VTT, Available also URL: http//www.vtt.fi/inf/pdf/

Norros, L., & Nuutinen, M. (2005). Performance-based usability evaluation of a safety information and alarm system. *International Journal of Human-Computer Studies*, 63 (3), 328-361.

Nuutinen, M., & Norros, L. (2007). Core-Task Analysis in accident investigation – analysis of maritime accidents in piloting situations. *Cognition Technology and Work* (in press).

O'Hara, J., Higgins, J., Persensky, J., Lewis, P., & Bongarra, J. (2002). *Human factors engineering program review model*. Washington, DC: United States Regulatory Comission.

Peirce, C.S.(1958). Letters to Lady Welby. In P. Wiener (Ed.), *Selected writings of C.C. Peirce* (pp. 380-432). New York: Dover Publications.

Peirce, C.S. (1998). The Harvard Lectures on Pragmatism. In Project TPe (Ed.), *The Essential Peirce. Selected Philosophical Writings*, Vol. 2 (pp. 133-241). Bloomington and Indianapolis: Indiana University Press.

Savioja, P., & Norros, L. (2004). *Developing the concept of system usability.* IFAC/IFIP/IFORS/IEA Symposium Analysis, design, and evaluation of human-machine systems. Atlanta, GA

Vygotsky, L.S. (1978). *Mind in Society. The Development of Higher Psychological Processes*. Cambridge, Mass.: Harvard University Press.

Résumé

Un des rôles centraux de la recherche en ergonomie est de fournir une base normative pour évaluer l'adaptation des objets destinés à être utilisés dans les activités humaines. Dans cet article on propose une approche basée sur la théorie de l'activité, appelée «*analyse de la core task»*. Inspirée par cette dernière, une méthode d'évaluation, appelée « *l'évaluation contextuelle de l'utilisabilité des systèmes* » a été développée pour analyser l'appropriabilité des systèmes complexes d'information et de conduite de processus. Les principes de base de cette méthode, et les indicateurs qui permettent l'évaluation sont décrits succinctement.

Mots clés

théorie de l'activité, analyse de la tâche focale, utilisabilité des systèmes, Tabitus, modélisation fonctionnelle, systèmes complexes.

Resumen

Uno de los roles centrales de la investigación en ergonomía es de aportar una base normativa que permita evaluar la adaptación de los artefactos utilizados en diversas actividades humanas. En este artículo presentamos un nuevo enfoque basado en la teoría de la actividad, denominado Core-Task Analysis (análisis de las tareas centrales). Inspirados en este último, se ha desarrollado un método llamado «evaluación contextual de la usabilidad de los sistemas » que permite evaluar cuán apropiados son los sistemas complejos de información y de control de procesos. Se describen sucintamente, los principios de base de este método, y los indicadores que permiten la evaluación.

Palabras clave

Teoría de la actividad, análisis focal de la tarea, usabilidad de los sistemas, habitus, modelización funcional, sistemas complejos.