Tikka, P. 2005. Design and Evaluation Conventions in Information Design, Usability Engineering and Readability Research: An Overview. In Kuure, L., Kärkkäinen, E. & Saarenkunnas, M. (eds), *Kieli ja sosiaalinen toiminta – Language and Social Action*. AFinLA Yearbook. Publications de l'association finlandaise de linguistique appliquée 63, 335-346.

▼▼▼▼

# DESIGN AND EVALUATION CONVENTIONS IN INFORMATION DESIGN, USABILITY ENGINEERING AND READABILITY RESEARCH: AN OVERVIEW

**Piiastiina Tikka**
University of Oulu

The present paper reviews some of the central design and evaluation conventions in information design, usability engineering and readability research. The main aim is to compare key features in these conventions and methods used in these disciplines so that eventually it may be possible to find ways of improving and adding to available evaluation methods for assessing text-based communication in various settings. Comparison of methods shows that all reviewed disciplines have common practices in place that can be used as a starting point for developing assessment of text-based communication. One such common practice is the expert evaluation method, which in usability engineering is considered a fast and cost-effective method with an acceptable reliability level, and the conclusions in the present paper suggest developing the heuristic expert evaluation method for information accessibility assessments.
**Keywords:** information accessibility, readability, design evaluation

## 1 INTRODUCTION

The present paper reviews some of the central design and evaluation concepts and conventions in information design[1], usability engineering[2], and readability research. These disciplines are all concerned with making it easier for a user (or reader) to perform some function or achieve some set goal, whether to use a computer program, read a children's book or employ some device with the help of a technical manual. In other words, they all have the user in common. The broader context for the present paper is efficient text-based communication[3] on small screen devices. By

[1] A high-level term describing various aspects of technical communication such as clear, understandable communication with concern over structure, context and presentation of data.
[2] A systematic approach to making systems and interfaces easier to use.
[3] Any communication in text format, distinguishing text from other possible electronic communication methods of voice, video, images or animation.

comparing key features in development conventions and evaluation methods used in these user-centred disciplines it may be possible to find ways of improving and adding to the available evaluation methods for assessing text-based communication in various settings – including small screen mobile devices.

The need to bring usability engineering methods to bear in information design stems from the increasing amounts of information communicated through such extremely fast distribution channels as the Internet and mobile communications. Vast amounts of data are being written and distributed for example as websites, and the complexity of many modern technical devices requires help systems and other documentation. Organisations and individuals pour out more information onto all available communications channels in all possible formats faster than can ever be read, but this does not mean that organisations could afford to decrease the level of quality for their information products. To have an economical yet efficient method of checking how comfortable and efficient a given piece of written communication is before it is released to the public will potentially save both time (and thus also money) and an organisation's image – not to mention end users' efforts in wading through all this information.

The present paper is a part of a broader research on information accessibility on mobile devices with small screens. However, within the scope of this paper it is more feasible to have a broad view of existing conventions, and only later to analyse and apply the findings to a specific field.

## 2 DESIGN CONVENTIONS SURVEY

The present section examines some established design conventions in information design, user interface design and readability research. Final usability of a product is to a great extent a result of the design process, and it is next to impossible to assess the usability of an object without understanding for what (and whom) it was designed. Therefore the present chapter is concerned with the above-mentioned design conventions, and the conventions of evaluating designs and products have been allotted a chapter of their own (see chapter 3).

### 2.1 INFORMATION DESIGN CONVENTIONS

Documented methods and processes in information design emphasise planning and the necessity of user and task analysis. A basic information design project (for example, creating a user guide for some device) is built up of three basic project phases: planning – implementation – evaluation. Iteration as such has not been discussed so much as a process in information design as it has been in the context of user interface

design, but clearly the conventions of reviews and prototyping show that iterative design process is used widely.

An information design project rarely has genuine substance to be evaluated at the planning phase; only the plans themselves can be reviewed by other information design professionals and the product development team. Draft documentation sets can be produced later on during the implementation phase and these can be assessed for example in tests on the actual products (for example, when testing the installation process of a software product the test person can go through it with the help of the draft documentation).

User and task analysis, often presented in a more comprehensive information plan (Hackos 1994: 81), is one important source of data for information designers to use in planning the structure, look and content of a given product's documentation. In the information design context a user and task analysis aims at gathering data on how a given product's end users benefit from the accompanying documentation, and how they use it.

Another important tool for an information design project is content specification, which is a detailed guide to what a given documentation set is to contain – how the needs discovered in the user and task analysis are satisfied. (Hackos 1994: 34) Content specification provides a means for other project members to visualise the documentation set before any of the content is actually produced: a good content specification will show basic structure (information order) and list illustration needs and planned contents, thereby making it possible to catch any deficiencies or shortcomings (or unnecessary extras) during reviews before the documentation set is actually produced. In effect, content specifications are early prototypes (Hackos 1994: 428).

In the planning phase an information design project will list the usability requirements set for the documentation product. For example, a usability goal in electronically distributed help systems can be that the content in the majority of help topics is fully visible without having to scroll the help window at all. Other basic requirements, typically, relate to the basic functionality of a documentation set: hyperlinks work and lead to the correct reference, index and glossaries are accurate, illustrations are clear in quality, and so on. Testing the usability of a documentation set is really only possible when there are documented usability goals to test the set against (Hackos 1994: 118). Aspects of documentation (that is, information design) usability are considered in more detail in section 3.2.

## 2.2 USER INTERFACE DESIGN CONVENTIONS

User interface (UI) design is an integral part of any product development that involves human-computer interaction. The industry conventions are by now fairly standard, if one is to believe the multitude of textbooks

produced on the topic. Although the present paper is concerned with those conventions and patterns documented in various industry publications, we should always remember that user interfaces are typically designed for a new product with some novelty value and that the industry cannot afford to remain static but must always be on the lookout for new methods, processes and opportunities.

Some principles collected and presented in Shneiderman and Plaisant (2004) include a thorough user and task analysis and selecting an interaction method (for example, direct manipulation, menu selection or command language). These principles include what they call the "eight golden rules of interface design" and include concepts such as consistency, feedback, sensible dialog design and reversal of user actions. There is so much agreement over these principles in the industry that Shneiderman and Plaisant are merely one example of the many UI design books and authors listing them in one form or another. User and task analysis is, as mentioned, also a central tool in information design (see section 2.1) and in the case of user interfaces Hackos and Redish have described the method in detail in *User and Task Analysis for Interface Design* (Hackos & Redish, 1998).

# 3 EVALUATION CONVENTIONS SURVEY

Usability engineering has evolved around the need to assess usability of all types of devices and tools used by people every day. The conventions and usability parameters have become so established they have been standardised to some extent. Readability research approaches contents from a different direction by assessing texts after they have been completed. Information design evaluation methods combine some elements of usability engineering, but the statistical methods more familiar from readability research are not alien to it either. The present chapter deals with each of these disciplines in individual sections, and conclusions regarding the disciplines are presented in chapter 4.

## 3.1 USER INTERFACE DESIGN USABILITY EVALUATION

'Usability' here is used as a generic term to denote all such activities and methods that are used in improving or studying user experience. Usability in general falls under various titles in literature and research: human-computer interaction (HCI), user-centred design (UCD), man-machine interface (MMI), human-machine interface (HMI), operator-machine interface (OMI), user interface design (UID), human factors (HF) and ergonomics (Nielsen 1993: 23).

Usability in itself is an abstract, and requires closer definition through more concrete concepts in order to gain a more tangible shape. These more concrete concepts are: learnability, efficiency, memorability, errors

and satisfaction. In addition, usability is always relative to a set task and set users and the optimal usability defined for one such set may not be directly applicable to another. (Nielsen 1993: 26-27) Furthermore, it may not be possible to end up with the optimal result in all the listed categories (for example, a highly learnable system will probably not reach optimal efficiency levels in the beginning), and usability trade-offs are necessary. This means balancing the importance of the usability categories according to usability goals and criteria set in a development project. (Nielsen 1993: 42)

Usability engineering and information design use similar methods for defining their ultimate goals. At the core of these methods is a close analysis of users and their tasks. Nielsen (1993: 43) recognises three areas in which users (even within a basically homogenous group) will differ from each other on an individual level: their experience with the system in question, with computers in general and with the task domain in question. However, typically user experience is discussed on a simplified axis of novice – expert (or somewhere in between) (ibid.). The difference between system expertise and domain expertise can be considerable in cases where an established domain expert is adopting new software. Expertise in the domain field allows for specialised terminology in the interface, but lack of experience in the system requires usability parameters appropriate for a novice.

Other aspects in which users can differ from each other are (for example): age and gender (which are relatively easy to observe), spatial memory, reasoning ability and preferred learning style (which are more difficult to observe). In addition, there will be differences in the general capacity and ability for carrying out given tasks: some people are simply more productive and efficient than others. (Nielsen 1993: 47) Because of these differences it is important in usability engineering to know as much about the actual users as possible. Through an understanding of the user demographics it is possible to set goals for the complexity levels of a given system under development. In addition to users themselves, it will help to know the context of use (office environment, noise levels, etc.). (Nielsen 1993: 73-75).

User analysis is complemented with detailed task analysis (Nielsen 1993: 75), which is also a set method in information design. Through task analysis system developers gain essential input to their design, and usability engineering is able to set usability goals for the system. Some aspects of user tasks that task analysis should address are: overall goals of a given task, users' current approach to that task, users' information needs and how the users deal with exceptional circumstances and emergencies. (Ibid.)

Usability engineering is a process that goes hand in hand with a product development process. This means, in practice, that usability is taken into consideration in the early stages of product development (user and task analysis provide data not just for system designers but also for usability engineering), early designs are tested with suitable usability

evaluation methods, designs are iterated accordingly, new versions produced and again tested, and so forth. According to Nielsen (1993: 16), what makes a good user interface cannot be defined in a vacuum but can only be achieved by applying the usability process in product development. Because of this close proximity to system and user interface development, usability activities and goals are often interwoven with development methods. For example, parallel design is one such method where several preliminary designs are produced in the beginning, evaluated for usability, with one selected for further development (Nielsen 1993: 85).

One major aspect of user interfaces and systems that is scrutinised in usability engineering is consistency. The usability activities for coordinating the interface do not only cover what is on the computer screen, but the entire package: application screens, documentation, online help and training material (for example tutorials). Consistency also applies to further releases of the same system. (Nielsen 1993: 90).

## 3.2 READABILITY EVALUATION

A specific field devoted to evaluation of the functionality, complexity and communicativeness of written language is the study of readability. At its broadest, readability study takes into account all those aspects that influence reading and comprehension.

There are number of tools devised in readability studies to measure and evaluate readability of texts. Gilliland (1972: 83) lists formulae, tables and charts, sentence completion and cloze procedure, objective question and answer techniques and subjective assessment. Readability assessment of the three main elements of readability (as stated in the definitions) involve

1) Ease of reading
2) Interest
3) Ease of understanding.

(Gilliland 1972: 83).

The first item on the list, ease of reading, concerns basic reading skills (word recognition, error rates, eye fixations, etc.) and physical aspects of visibility and legibility. The second is – understandably – studied from a human interest point of view. Studies of the third group have focused on words and sentences (their length, frequency, etc.). (Gilliland 1972: 83-84) The results derived from studying these three areas of definition separately are not, according to Gilliland (ibid.), comparable.

Readability formulas can be useful after a user and task analysis phase has been completed and there is a clear understanding of what is being measured and with whom. Such reader-oriented aspects as interest and motivation cannot be left out of the equation, and user and task

analysis may provide some data that helps to negotiate the interest and motivation variance among readers. For example, instructional text is used with the aim of completing some task. The readers therefore share a common motivation. A history book for schools, however, will have a readership where some readers already have an interest in the subject matter, while other readers have no personal motivation to read and study regardless of possibly having to pass exams on the subject later. First impressions, then, are that a readability formula may be of some use in the first example, provided that the readership is otherwise homogenic enough, but that there may be difficulties with the second example group.

The idea that exact mathematical (and statistical) methods could be imposed on a sample of written text is, probably, quite appealing if someone is looking for an engineering solution for their language evaluation needs. However, as the adverse comments by Bruce & Rubin and Anderson & Davison (for example) in Davison, A. & Green, G.M. (eds.) 1988 show, these formulas at best give very little insight into the true level of language complexity and into how readable or unreadable a piece of writing is. According to these articles, readability formulas ignore some of the very central human factors in how people read, comprehend and perceive written text. For example, replacing the influence of a person's prior knowledge of a particular topic he or she is reading about with a vocabulary difficulty rating (Davison & Green 1988: 44) does not cover the full range of that reader's life experience, genre literacy or any personal competence in reading. There are no means of grading a reader's level of interest, prior knowledge, motivation (whether natural/personal or conscious direction of focus), which means that in effect formulas will overlook all individual aspects of comprehension and reading ability. The result, therefore, is that we are left with formulas that are practically incapable of producing usable data for evaluation of human communication.

Readability formulas tend to focus on items such as word difficulty levels and syntax complexity as charted from a small sample of text. Owing to the limitations on what can be deduced from a text sample for statistical analysis, results of this kind of analysis should not be used for revising texts (Davison & Green 1988: 25). That being the case, an immediate question arises: if a method that attempts to reveal unnecessary complexity and hidden difficulty in texts cannot be used in trying to improve those texts, precisely what purpose do the methods serve? As Bruce & Rubin (in Davison & Green 1988: 16-18) found, simplification of a given text reduced its cohesion and overall communicativeness. At times intricate issues may require intricate communication, which, in turn, may make a given text look complex when evaluated with such crude tools as readability formulas. Bruce and Rubin also point out (ibid.) that it is possible to devise texts which seem to be very readable when tested with the measures used, but which are in fact impossible to comprehend owing to lack of genuine desire to communicate.

Cloze procedure has been used in readability assessment as well, but according to Gilliland (1972: 103) it has come in for some criticism, due to the extent of variance in test setting. It has also been suggested that it is possible for test subjects to fill in the gaps from their experience and knowledge of familiar speech and language patterns. Despite shortcomings, however, with proper test setting cloze procedures should have some use for readability studies. A major benefit of the method is that it efficiently takes the individual reader into account: the test reflects the full reading ability and linguistic capability of the reader and in this manner satisfies all aspects of the earlier definitions of readability. Also, the test itself matches a reader and a text intended for that reader. (Gilliland 1972: 103 –105)

Question and answer techniques have, in practice, not only been used to measure the difficulty of a passage but to provide a base line against which other measures can be compared. These question and answer techniques are primarily intended for measuring the comprehension of a passage, and do not produce data for assessing overall readability (such as fluency). There are also various technical shortcomings that limit the applicability of the question and answer technique. For example, although it is possible to assess whether a given text is easy or difficult to understand, it is impossible to determine which are the factors that lead to such a conclusion (whether the subject matter is too difficult for the reader or the phrasing too complex). Also the testing situation (for example time limits) may affect the result. (Gilliland 1972: 87-88) Thus it seems that although question and answer techniques do have a place in readability studies, their use should be considered carefully and the resulting data used critically to indicate no more than what the tests are capable of measuring: comprehension.

Because it is so difficult to find accurate quantitative methods for measuring overall readability, the qualitative paradigm may have something to offer researchers and practitioners. Gilliland refers to a study by Moyle (1971, in Gilliland 1972: 85) on consistency of grading carried out by a group of people rather than by individuals. The study showed that grading of books was more consistent when this was done by a committee (in this case a group of experienced teachers) than when undertaken by similarly experienced teachers individually (ibid.). Gilliland points out results by Chall (1958, in Gilliland 1972: 85) that indicate inconsistency and unreliability in such collective assessment. Gilliland does not, however, pursue the conflict further, and more information on the details of each study is required to see whether they are at all comparable. Expert evaluation techniques used in usability engineering rely on setting clear targets and acceptability guidelines regarding the experts' assessment. In these evaluations the evaluators are expected to review a given user interface within certain parameters (see section 3.3).

There are a great number of further studies into assessing readability. Michael Pressley (in Farstrup and Samuels, 2002: 291-299) refers to a

number of thinking-aloud methods that have been used in studying the active comprehension strategies of good readers. Thinking-aloud is also a useful technique in usability engineering. In such a test, the test subjects are asked to voice their thoughts on what they are doing, thinking and/or feeling. The consistency of results derived from the tests referred to by Pressley (over 40 published studies), indicating certain mainstream strategies common to good readers, suggests that thinking-aloud may also be an effective test method in text evaluation as defined in the context of the present study. The test method itself is highly subjective and the interpretation of results requires a considerable input from the test moderator, but it would seem highly feasible to use the technique as a complementary test method to any other tests made on text and information accessibility.

Baker, Atwood, and Duffy (in Davison and Green 1988: 55-61) mention, along with prediction formulas, "judgment of expert writers" as an alternative to actual testing. They find this judgment to be inadequate as a stand-alone, but when the evaluators in their example test were given user test data, such as transcripts of test subjects' thinking aloud, the evaluators were able to improve the readability of the tested documents significantly.

Baker, Atwood, and Duffy go on and to suggest a number of strategies of testing users (readers) in real scenarios, for example by a thinking-aloud test. One test setting designed for a technical manual audience is to have the test subjects perform a given task following the directions in the manual and think out loud while performing it. Also, the test must be calibrated so that it reflects the practical information goals of the readers. In other words, if the text is used as a reference point only, it is not reasonable to test how well the readers have learned it. (Ibid.)

## 3.3 INFORMATION DESIGN USABILITY ASSESSMENT

The present overview approaches documentation usability to a large degree as a question of user interfaces. For example, a user manual for some software product can be distributed as an electronic documentation library that is accessed through certain operations on a computer and is browsed and read on the screen. Accessing the content of that documentation library requires that the user is able to open the library, see the content, navigate the contents, perhaps change some settings, find commands for, say, printing items off the set, and eventually close the set. The ability to perform all of these functions is really a question of user interface design. Reading-specific ergonomic issues, such as the physical size of a documentation set, guiding the eye, selected typeface and font size, colour combinations and icon and symbol design, are information design considerations, although they do overlap with UI design and can be tested alongside the rest of a product's UI at a convenient stage.

Hackos (1994) reminds information designers not to overdo testing: if there are some general matters that have already been tested, such as optimal font size or colour combinations that are simply impossible, there really is no need to repeat these tests (Hackos 1994: 438). Usability data, if collected and documented properly, is reusable. Hackos has divided documentation usability testing into three phases: 1) planning, 2) implementation and prototyping, and 3) production and evaluation. The planning phase contains the basic assessment that sets style standards and helps in making design decisions. After planning, more detailed assessment is possible: technical reviews and performance testing are possible on draft content, and the documentation drafts (that is, prototypes) can be tested together with the product in one observed performance test. (Hackos 1994: 438). Hackos (ibid.) also lists cognitive walkthroughs as one assessment method in which experts simulate users progressing through an interface (or, in this case, documentation set) carrying out typical tasks (Shneiderman & Plaisant 2004: 142). In this respect the assessment method is related to the heuristic usability assessment method. These prototyped assessments are finally repeated in the process close to or soon after actual product and documentation release to get a final assessment that will help towards revising standards and planning for the next release (Hackos 1994: 439).

## 4 CONCLUSIONS

The overview of evaluation methods presented above suggests that neither quantitative nor qualitative evaluation methods alone are enough in themselves to produce good, generally applicable evaluation results. Usability engineering, according to the popular industry guidelines as presented by Nielsen and Shneiderman, for example, promotes qualitative methods in the pursuit of evaluating user experience. Although some work has been done in readability research on more subjective, qualitative methods, the search for quantitative (statistical) methods still weighs heavily over the field. Information design evaluation methods often overlap with those in usability engineering, but naturally focus the methods on the information products rather than the actual product or system itself. Within that focused area, information design has the opportunity to use readability research evaluation methods as well.

To summarise the main findings in the present overview, both UI and information design (judging by the number of publications dedicated to them and the level of agreement among these publications) have well documented support, methods and guidelines available for their processes. Furthermore, both UI and information design have correlating methods in these design process models, most interestingly the use of iteration as an in-built evaluation stage. Readability research approaches contents from a different direction, evaluating text after it is set and

released rather than during the creation process. Some researchers in this field have pointed out the usefulness of both think-aloud and expert evaluation methods, which are well-established methods in usability engineering.

Expert evaluations in information design tend to focus on measurable items such as use of certain terminology, sentence length, form and visual design of a documentation set, to name a few. Evaluations are performed iteratively (as in usability engineering in a UI context) on early prototypes (for example, a context specification), and typically the experts in these evaluations represent the knowledge domains of technical writing and product design. The quantitative approach should be able to answer the most important question a technical writer will probably want to ask: can the user perform a given task with the help of this documentation, and will he be able to perform it efficiently? Typically technical documentation is intended as guidance for performing set tasks with a specific device or tool, and the more specialised the device or tool, the more reasonable it is to expect the readership to have common characteristics in their technical and professional capacity. The homogenous nature of technical documentation readership and the limited area of application of the documentation (practically all readers can be expected to have the same motivation for reading) mean that this one question is enough. However, such a homogenous readership and limited application area are the reasons why a quantitative approach – regardless of expert reviews – is not enough when the object of evaluation is more general in nature (that is, not a guide for a specific product or task) or its application area (context of use) is more flexible than that of a technical manual.

One of the goals for the present paper was to see if the overview of established design and evaluation methodology would allow adding to and enhancement so as to find a method that enabled faster and more cost-effective evaluation of written content. The overlapping areas listed above show that the basic tools are already available, though they will have to be calibrated for information accessibility. Comprehensive laboratory testing and quantitative measurables may provide a great deal of the data required for improving a product, but laboratory testing is time consuming and expensive, and quantitative, statistical methods for evaluating language are not likely to be reliable enough alone. One method that answers the need for speed with reasonable reliability in usability engineering is the heuristic expert evaluation. Also, the usefulness of thinking-aloud tests on actual test subjects has been acknowledged in the reviewed literature, and Baker, Atwood and Duffy (see section 3.1) point out the benefits of providing expert evaluators with other data (namely, thinking-aloud transcripts) to support their evaluations.

Based on these observations it seems reasonable to assume that it should be possible to derive a set of heuristics that would apply to the use and design of language on screen. The heuristics, used either alone or

together with additional data, would add a more formal and reproducible method for language evaluation in the iterative development process, for example in an information design setting.

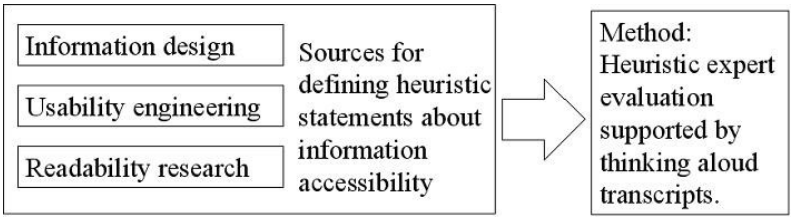


**Figure 1.** Simplified process description for creating a heuristic expert evaluation method based on the three disciplines discussed in the present overview.

Using existing literature on readability, ergonomics of reading, documentation usability and usability engineering together and collecting new observation data on users, those aspects of information accessibility that can be anticipated in a design process could then be isolated and expressed as heuristic statements. These statements would then have to be calibrated and tested against more formal usability data (that is, laboratory test data) in order to ensure their accuracy and to see if the reliability level was acceptable (an estimated target should be close to the established reliability of heuristic evaluations in usability engineering).

## REFERENCES

Anderson, R. & Davison, A. 1988. Conceptual and Empirical Bases of Readability Formulas. In Davison, A. & Green, G.M. (eds.)

Baker, E., Atwood, N. & Duffy, T. 1988. Cognitive Approaches to Assessing the Readability of Text. . In Davison, A. & Green, G.M. (eds.), 55-83.

Bruce, B. C. & Rubin, A. D. 1988. Readability Formulas: Matching Tool and Task. In Davison, A. & Green, G.M. (eds.), 5-22.

Davison, A. and Green, G.M. (eds.), *Linguistic Complexity and Text Comprehension: Readability Issues Reconsidered*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.

Farstrup, A.E. & Samuels, J.S. (eds.). 2003 (2002). *What Research Has to Say About Reading Instruction*. Delaware: Internal Reading Association

Gilliland, J. 1972. *Readability*. London: University of London Press Ltd

Hackos, J.T. 1994. *Managing Your Documentation Projects.* New York: John Wiley & Sons, Inc.

Hackos, J.T. & Redish, J.C. 1998. *User and Task Analysis for Interface Design.* New York: Wiley Computer Publishing, John Wiley & Sons, Inc.

Nielsen, J. 1993. *Usability Engineering*. San Diego: Academic Press.

Pressley, M. 2002. Metacognition and Self-Regulated Comprehension. In Farstrup, A. & Samuels, J.S. (eds.), 291-309.

Shneiderman, B., and Plaisant, C. 2004 Designing the User Interface: Strategies for Effective Human-Computer interaction, 4th edition.