The strength of usability: An attempt to measure transfer of training

Stefan Holmlid
ASLAB, Department of Computer and Information Science, Linköping University[1]

The training of end-users has become one of the most wide spread methods for companies to support the use of software in an effective and efficient manner. In order to assess whether knowledge acquired through training is transferred and utilized in the work situation, evaluations beyond knowledge tests need to be used. It is proposed that subjectively perceived usability may be a way of evaluating transfer of training. Although not equivalent to transfer of training, it is shown that subjectively perceived usability is one way of evaluating the transfer and utilization of knowledge acquired in training to a work-situation.

As a first attempt to assess the relevance of the proposal, this study applies one measure of usability, Software Usability Measurement Inventory, SUMI, to evaluate the transfer of training.

Introduction

Training end-users has earned a lot of interest during the last decade. The research has gained its perspective very much from the areas of psychology and instructional design. As a result the focus has been on 1) whether individual differences, such as age, gender, experience, education and so forth (Harrison & Rainer 1992, Szajna 1994, Martocchio 1994), influence skill levels acquired through training, or on 2) how knowledge of a system should be communicated in order for users to be able to acquire the knowledge and thus become skilful (Waern & Rabenius 1987, Martocchio 1994).

Generating skilful users is one important goal for end-user training. But what really makes a difference is if the acquired knowledge is used in the user's work situation. This is referred to as *transfer of training*. Training end-users to be very skilful will, of course, matter less to a company if the users do not use the acquired knowledge at work.

There is a need for a means to evaluate if users have transferred training to their work situation. An evaluation instrument that is capable of doing so would have the power to change the marketplace for end-user training companies. Traditional evaluation models are divided into four levels; reaction, learning, behaviour and results (Garavaglia 1993). Today most companies training end-users do only reaction-level evaluations, test if the users know what they should know through a post-test, and evaluate if the participants were pleased with the lunch. The most responsible companies also promise that users should have a certain degree of knowledge after training. Companies that will not end their efforts until users utilize their new knowledge to solve their work tasks are rare, and live a quiet life.

1. This research was made possible by the support of DataMedia, Stockholm. The author wishes to acknowledge Jurek Kirakowski for his invaluable help with SUMI, and Jonas Löwgren for invaluable support as supervisor.

A method for evaluating if transfer has occurred can be chosen on many grounds. Two common perspectives are an instructional designer's, which is based on psychological research, and a manager's, which is based on organisational theory and human resources research

We propose a third complementary perspective, from the field of human-computer interaction, reasons for which will be shortly sketched here, although the paper will not go into any depths on this. The evaluation method utilized is a usability evaluation. This makes it relevant to system developers. Usability evaluations can be done on the program itself to judge one qualitative aspect of the software. If evaluations of training are done with the same method, system developers will be able to use that information in order to make better products.

Busch (1994) states that transfer of training is an evaluation of how well knowledge acquired in training is utilized in a work situation. He describes a model of transfer of training. Transfer, in his multidisciplinary view, is the result of a complex interplay between a set of organisational and personal variables (Figure 1). Similar descriptions of transfer of training are provided by Garavaglia (1993), Stine and Wildemuth (1992) and Davis and Bostrom (1993).

Usability, can be described in many different ways (ISO-DIS 9241, Löwgren 1993, Porteous, Kirakowski & Corbett 1993). We will use the interpretation that usability is an evaluation of how well part of a work situation might be solved with knowledge on how to utilize a computer. Such an evaluation might be done, e.g., by looking at the aspects relevance, efficiency, affect and learnability (Löwgren 1993) or the aspects efficiency, affect, helpfulness, control, and learnability (Porteous et al. 1993). Both definitions might be used to construct metrics of the usability of a piece of software.

There are mainly two aspects of usability that can be measured, performance and preference (Nielsen & Levy 1994). Performance measures are of great interest if there is a need for assessments of, e.g., the possibility to make more products in less time. Preference measures on the other hand, tells us about the usability a specific user perceives. Busch (1994), among others (Harrison and Rainer 1992), argue that certain subjective measures, such as self-efficiacy, correspond well to actual performance. This indicates that a preference measure or a subjective instrument might well be useful in measuring transfer of training.

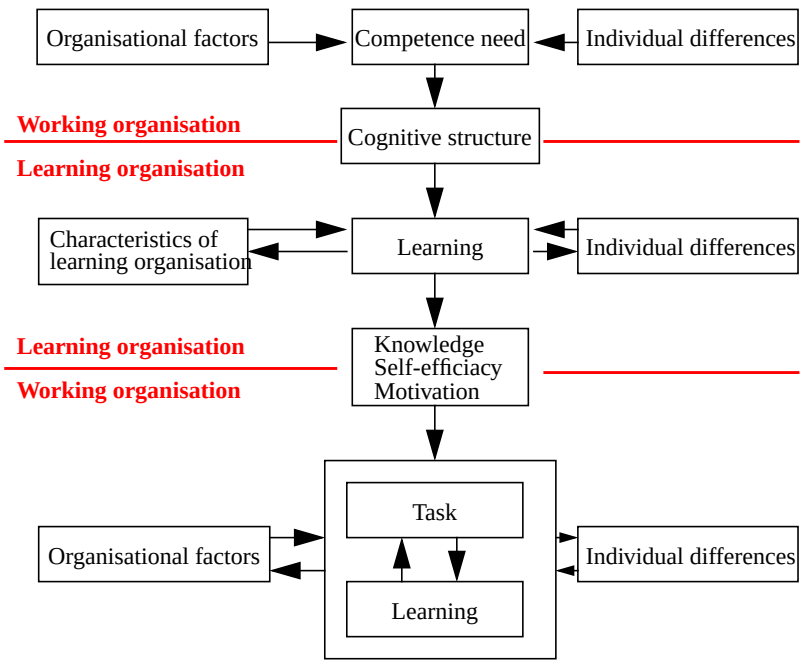

FIGURE 1. Busch (1994) multidisciplinary model of transfer of training.

Hypotheses

Usability and transfer of training not only share concepts on the level of definitions, but also more specific concepts. Motivation, e.g., is likely to be a shared concept (Harrison and Rainer 1992). If usability is to be regarded as an evaluation of the transfer of training, the measure of usability has to increase if training is transferred.

The main hypothesis is therefore that

Hypothesis 1. An increase in subjectively perceived usability indicates that transfer of training has occurred.

According to Agassi (1980) a refutation of this hypothesis has to be a ground for making useful conclusions. It is a puzzle we need to solve.

Looking at Busch model of transfer of training, transfer can fail to occur for several different reasons.

> Some participants will learn, but will not transfer knowledge to their work-situation. Other participants will not learn because they do not engage in the learning-process – thus they have nothing to transfer. The consequences are the same, but the different problems should not be mixed.
>
> Busch 1994, p 94, [author's transl.]

This means that transfer will more likely occur if the student's knowledge, self-efficiacy and motivation is not influenced. On the basis of this we formulate the first working hypothesis

Hypothesis 2. Self-efficiacy and motivation changes positively through training.

Moreover, in line with Garavaglia (1993), a student need to enter the task/learning loop of Busch's model for transfer to occur. On the basis of this we formulate the second working hypothesis

Hypothesis 3. The student's have entered the task/learning loop.

Overview

First the study undertaken will be presented. Thereafter there will be a presentation of the results. Following that will be a discussion around the hypotheses and last we look into the near future.

Empirical work

The empirical work was done in close collaboration with an instructional company, and was performed in a real training session, with the researcher acting only as an evaluator of training. The empirical material is presented more thouroughly in Holmlid (1995).

Measurement method

The instrument used in the evaluation is Software Usability Measurement Inventory, SUMI, a questionnaire especially developed to measure subjectively perceived usability, carefully constructed and validated by the Human Factors Research Group at University College Cork, Ireland (Porteous et al. 1993), within the EEC research programme MUSiC. The instrument is commercially available, which makes it easily accessible for practitioners, but hard to be too detailed about here. An evaluation using this instrument generates both the individual users' ratings of a system and a relative rating of a specific system, with a standardisation database of over 1000 evaluations as baseline.

SUMI consists of five subscales (Porteous et al. 1993)

- Efficiency, which refers to the user's feeling that the software is enabling the task(s) to be performed in a quick, effective and economical manner.
- Affect, which refers to the user feeling good, warm, happy or the opposite as the result of interacting with the software.
- Helpfulness, which refers to the user's perceptions that the software communicates in a helpful way and assists in the resolution of operational problems.
- Control, which refers to the feeling the user has that the software is responding in a normal and consistent way to input and commands.

- Learnability, which refers to the feeling the user has that it is relatively straightforward to become familiar with the software and that its tutorial interface, handbooks etc., are readable and instructive.

There is also a global scale which refers to the users general perception of the usability of the software.

The questionnaire consists of 50 statements together with a three value Lickert-scale. The evaluator is asked to check a box if she Agrees, Disagrees or Don't know whether she agrees or disagrees, with the statement.

The maximum score a factor might receive is 100 and due to the scoring procedure which relates a specific evaluation to the standardisation database, with a mean of 50 and a standard deviation of 10.

Subjects and procedure

A total of 10 subjects, equal proportions of male and female subjects, voluntarily participated in the study. They were all working at the same medium sized Swedish marketing firm, and were given training, adapted to their individual experience, on a well-known word processor. The trainer was an employee from the instructional company who also had planned the training

The study undertaken consisted of three evaluations. One at the beginning of the training period, one at the end of the training period, and the last evaluation two weeks after the training period ended. The pre-study was performed during the first week of the training period. The subjects received the SUMI questionnaire during their first training session and were asked to bring it to the next training session. The post-study was performed during the last training session of the training period, eight weeks after the pre-study. Every subject was asked to fill out the SUMI questionnaire during the last training session. The delayed study was performed two weeks after the post-study, and every subject used approximately half an hour for the evaluation. The subjects were asked at all three times of evaluation to fill out the questionnaire according to the instructions.

Results

Table 1 presents the median score and the standard deviation of the evaluations. The figures in the table show that the subjectively perceived usability in the pre study is poor, and around normal in the delayed study. The standard deviations shows that the users for most of the scales agree less on the perceived usability in the delayed study than they do in the pre-study.

An analysis of variance, ANOVA, was performed in order to assess if the differences in distribution were significant. The result of the ANOVA yielded that there were some significant difference in SUMI-evaluations. A post-hoc test after analysis of variance gave the results in Table 2. In Table 2 the F-ratio of difference between means is presented.

TABLE 1. Median and standard deviation over the three evaluations

	Pre study		**Post study**		**Delayed study**	
	Median	**SD**	**Median**	**SD**	**Median**	**SD**
Global	43	11,48	49	11,15	53	15,56
Efficiency	47	13,27	41	12,67	44	11,43
Affect	62	11,43	62	8,07	65	9,10
Helpfulness	51	9,62	47	12,40	51	13,14
Control	40	9,04	49	11,24	53	13,85
Learnability	46	12,48	49	13,99	52	15,88

The figures in that table show if the difference between evaluations for every single factor is significant.

TABLE 2. The F-ratio of difference between means (n=10, * p<0.05, ** p<0.01, *** p<0.001). ns=non-significant

	Post-Pre	Delayed-Post	Delayed-Pre
Global	0,71 ns	0,07 ns	1,23 ns
Efficiency	0,80 ns	0,00 ns	0,90 ns
Affect	5,14 *	1,88 ns	13,23 ***
Helpfulness	1,74 ns	1,23 ns	0,04 ns
Control	7,51 **	0,47 ns	11,74 ***
Learnability	0,23 ns	1,34 ns	0,47 ns

This data shows that there is no significant difference between the evaluations of subjectively perceived usability, i.e., the *global* factor does not change significantly over the three studies.

Discussion

Looking at the score on overall usability there is no support for the hypothesis that subjectively perceived usability increases when transfer of training occurs. This means that either subjectively perceived usability can not indicate whether transfer occurs or that no transfer has occurred in this case.

This rejection will be a guide into making useful conclusions.

We now turn to the working hypothesis, that are concerned with the occurrence of transfer.

Users' self-conception influences usability

In order to be able to evaluate the possibilities for transfer to occur, a learning organisation can choose to evaluate the students knowledge, their motivation and their self-efficiacy (i.e., the extent to which a person believes she is capable of performing an action) at the end of a learning period.

Self-efficiacy can be interpreted as self confidence. The factor *control* describes the user's feeling of control over the system. As control is evaluated subjectively a user's self-efficiacy is likely to affect the evaluation of control. A user with low confidence in her ability to perform actions will probably rate her control over the system lower than a person with a high degree of confidence. A user feeling in control over the system probably will show a high degree of confidence, and vice versa.

Motivation can be both inner, e.g. how satisfying a task is, and outer, such as merit wages. The factor *affect* describes the user's perception of liking of the system. It is likely that a person who feels affect to a system also has inner motivation for using the system. A user who has inner motivation to use a specific system before another in solving her tasks is likely to be more fond of that system's ability to support her in solving her tasks.

Without claiming that self-efficiacy and motivation are equivalent to the factors *control* and *affect*, those factors are used in this study to assess the level of self-efficiacy and motivation.

In the study performed here both *control* and *affect* changed significantly by training, see Table 2. This demonstrates that self-efficiacy and motivation, in the light of the conceptual linkage between control and affect and self-efficiacy and motivation respectively, is affected by training. It also indicates the close relationship between subjectively perceived usability and users' self conception. The first working hypothesis is corroborated, and thus provide for a useful conclusion, in spite of the rejection of the main hypothesis.

One pre-requisite for transfer of training is confirmed and we have found a way to measure that through the use of subjectively perceived usability.

Shift of concern in usability evaluations

indicate transfer of training

A simple goal, then, would be to strive for maximization of these factors for every single user, and believe that one has done what could be expected from a learning organisation. This would be a valid approach taking a more traditional perspective. Organisational factors are to be regarded only in addition to, e.g., mental models and motivation (Sein, Bostrom & Olfman 1986).

It is also the responsibility of a learning organisation to provide for awareness that organisational factors and individual differences will affect the transfer of training. Sometimes even specific actions need to be taken, e.g., at the organisational level (Busch 1994).

In order to be able to evaluate whether transfer in fact has occurred a learning organisation need to assess whether users have entered the task/learning loop. One sign that users have entered the task/learning loop, would be that they, in the delayed study, do not view learning the system as a primary goal but that they still feel that it is stimulating to work with the system. It would mean that their focus is on the task, and that they have motivation and confidence in their ability to use, and learn about, the system for solving their tasks. Another sign that users have entered the task/learning loop would be that their concern about the system changes throughout the study from a general fear and need for safety contrasted against their work, over concern about the system as such, to a concern about how to learn new things to apply in their work.

The interpret evaluations at this level of detail, and qualitatively, the focus needs be specific items in the questionnaire.

Furthermore the procedure of the usability evaluations used in this study provide a possibility to compensate for the different situations under which the evaluations are done. First, the evaluation form was filled out two weeks after the training period had ended, i.e. after the users had returned to the working organisation, which gives us a picture of the difference between when the user was part of the learning organisation and part of the working organisation. Second, the questionnaire was filled out in the context of their current situation, either in the learning organisation or the working organisation and in close connection to a period of computer use, which ensures that we will get perspectives from both environments. Third, a measure of subjectively perceived usability has the possibility to include personal characteristics, such as if the user is interested in learning more or interested in using the system to solve more tasks, into an evaluation.

The evaluations support the first sign. In the delayed study four users answered that they never would learn everything offered by the system, and three users answered that they did not know if they would. Seven users answered that they felt that it is mentally stimulating to use the system, while three did not know. In contrast, in the pre study, six users thought that they would learn to use everything offered by the system, while four did not know. Moreover, in the delayed study, every user would like to use the system every day and every user would recommend it to their colleagues.

The second sign is also supported by the evaluations. In the pre study users showed fear of the system, and they express a need for safety. They contrast the system with their work. Their concern is how the system could support them in their work, on a general level. In the post study their concern is no longer coupled with their work tasks. The anxiety is about the system itself. They are concerned, e.g., with how to be able to judge on information presented in order to be able to continue working with the system. It is the internals of the system that concerns the users. In the post-study the users are concerned with trying out new things to support them in their work-tasks.

Both signs are supported by our study. *This result* suggests that subjectively perceived usability has the potential to assess the whether users have entered the task/learning loop. It also indicates a relationship between subjectively perceived usability and the situation of use. The second working hypothesis is corroborated, and thus provide for a useful conclusion, in spite of the rejection of the main hypothesis.

The task/learning loop has been entered, and another pre-requisite for transfer of training has been confirmed.

As far as subjectively perceived usability can tell, training has been transferred.

Main conclusions

The main conclusion of this study is that usability, in the shape used here, has a potential of being used as one instrument to indicate if transfer of training has occurred.

The secondary conclusions is that subjectively perceived usability is affected by users' self-conception and that subjectively perceived usability can be used to assess whether the task/learning loop has been entered.

Limitations

The research performed here is not without limitations.

First of all we fail to find support for our main hypothesis. One reason for that can be found in the way the global factor of SUMI is calculated from the other factors. We have a shown a larger interest in what is specific about users, instead of what is a normalisation. It can also be found in the fact that SUMI's factors are not developed in order to measure transfer of training explicitly, but the usability of a software product. It would therefore be of great interest to develop a usability construct which could be used for both purposes.

A minor observation is that usability, in the form used here, do not provide for an easy differentiation on any level between espoused theories and theories in use (Busch 1994). It is a tedious, and incomplete, process to evaluate if users have only a mental model of the system which is not reflected in their use of the system (espoused) or if they also use the system according to a model acquiring a high degree of usability (in use).

One obvious limitation of this study is the narrow perspective both on usability, transfer and training. Similar investigations with different usability constructs, different views on transfer and different ways of training would complement this research in a fruitful way.

This study thus can act as a ground for a continuing research on whether training can provide for usability of a software product, and in what ways.

References

Agassi, J. (1980). Between science and technology. *Philosophy of Science* 47:82-99.

Busch, T. (1994). Overföring av läring. *Nordisk Pedagogik*, 2:87-100.

Davis, S., A., Bostrom, R., P. (1993). Training end users: An experimental investigation of the roles of the computer interface and training methods. *MIS Quarterly*, 17(1):61-85.

Garavaglia, P., L. (1993). How to ensure transfer of training. *Training & Development*, oct:63-68.

Harrison, A. W., Rainer, R. K. Jr (1992). The influence of individual differences on skill in end-user computing. *Journal of Management Information Systems*, 9(1), 93-111.

Holmlid, S. (1995). *The effect of end-user instruction on usability: Usability as a key quality of instruction*. Final report, Dept. Computer and Information Science, Linköping University, S-581 83 Linköping, Sweden.

ISO-DIS 9241 (1994). *Ergonomics requirements for office work with visual display terminals (VDTs):- Part 11: Guidance on usability.* Available FTP: ftp.nlp.co.uk Directory: pub/hci File: ISO-9241-11.

Löwgren, J. (1992). *Human-computer interaction: What every system developer should know.* Lund, Studentlitteratur.

Martocchio, J. J. (1994). Effects of conceptions of ability on anxiety, self-efficiacy, and learning in training. *Journal of Applied Psychology*, 79(6):819-825.

Nielsen, J., Levy, J. (1994). Measuring usability: Preference vs. performance. *Communication of the ACM*, 37(4). 67-75.

Porteous, M., Kirakowski, J., Corbett, M. (1993). *SUMI user handbook*. Human Factors Research Group, University College Cork, Ireland.

Sein, M. K., Bostrom, R. P., Olfman L. (1987). Training end users to compute: Cognitive, motivational and social issues. *INFOR*, 25(3):236-255.

Stine, W., D., Wildemuth, B., M. (1992). The training of microcomputer users: Insights from two disciplines. *Journal of Education for Library and Information Science*, 33(2):100-109.

Szajna, B. (1994). An investigation of the predictive validity of computer anxiety and computer aptitude. *Educational and Psychological Measurement*, 54(4), 926-934.

Waern, Y., Rabenius, L. (1987). On the role of models in the instruction of novice users of a word processing system. *Zeitschrift fuer Psychologie*, Suppl 9:59-73.