Developing a Usability Evaluation Method for E-learning Applications: From Functional Usability to Motivation to Learn

Panagiotis Zaharias
Athens University of Economics and Business,
pz@aueb.gr

Abstract

In this paper the development of a questionnaire-based usability evaluation method for e-learning applications is described. The method extends the current practice by focusing not only on cognitive but also affective considerations that may influence e-learning usability. The method was developed according to an established methodology in HCI research and relied upon a conceptual framework that combines web and instructional design parameters and associates them with the most prominent affective learning dimension, which is intrinsic motivation to learn. The latter is proposed as a new usability measure that is considered more appropriate to evaluate e-learning designs. Two large empirical studies were conducted in order to evaluate usability of e-learning courses offered in corporate environments. Results provide significant evidence for reliability and validity of the method; thus usability practitioners can use it with confidence when evaluating the design of e-learning applications.

1. The need to develop a usability evaluation method for e-learning applications

Organizations and educational institutions have been investing in information technologies to improve education and training at an increasing rate during the last two decades. Especially in corporate settings continuous education and training for the human resource is critical to an organization's success. Electronic learning (e-learning) has been identified as the enabler for people and organizations to keep up with changes in the global economy that now occur in Internet time. Within the context of corporate training, e-learning refers to training delivered on a computer that is designed to support individual learning or organizational performance goals (Clark and Mayer, 2003). Although e-learning is emerging as one of the fastest organizational uses of the Internet (Harun, 2002), most e-learning programs exhibit higher dropout rates when compared with traditional instructor-led courses. There are many reasons that can explain the high dropout rates such as relevancy of content, comfort level with technology, availability of technical support etc. but one major contributor is the poor usability of e-learning applications. The latter is the focal point of this study.

Evaluating the usability of e-learning applications is not a trivial task. Increase in the diversity of learners, technological advancements and radical changes in learning tasks (learner interaction with a learning/training environment is often an one-time event) present significant challenges and render the possibility of defining the context of use of e-learning applications. Identifying who are the users and what are the tasks in e-learning context impose extra difficulties. In the case of e-learning design the main task for the user is to learn, which is rather tacit and abstract in nature (Zaharias and Poulymenakou, 2006). As Notess (2001) argues "*evaluating e-learning may move usability practitioners outside their comfort zone*". Squires (1999) highlights the need for integration of usability and learning and points out the non-collaboration of workers in HCI and educational computing areas. In fact usability of e-learning designs is directly related to their pedagogical value. An e-learning application may be usable but not in the pedagogical sense and vice-versa (Quinn, 1996, Albion, 1999, Squires and Preece, 1999). Accordingly usability practitioners need to familiarize themselves with the educational testing research, learning styles and the rudiments of learning theory. Nevertheless very little has been done to critically examine the usability of e-learning applications; there is an ellipsis of research validated usability evaluation methods that address the user as a learner and consider cognitive and affective learning factors that support learners to achieve learning goals and objectives.

1.1 Beyond functional usability: the emergence of affective dimension

A major challenge of contemporary HCI research is to address user affect. It is critical that systems designers assess the range of possible affective states that users may experience while interacting with the system (Hudlicka, 2003). To this end new usability techniques and measures need to be established (Hornbaek, 2005). Traditional usability measures of effectiveness, efficiency and satisfaction are not adequate for new contexts of

use such as home technology (Monk, 2002), ubiquitous computing (Mankoff, et al., 2003) and technology supporting learning (Soloway et al., 1994). In the e-learning context, affect has recently gained attention. It has been argued that affect is the fuel that learners bring to the learning environment connecting them to the *"why"* of learning. New developments in learning theories such as constructivist approaches put an increasing emphasis on the affective domain of learning; new thinking in adult learning theory and practice stresses the need to enhance learners' internal priorities and drives that can be best described by motivation to learn. The latter, a concept intimately linked with learning (Schunk, 2000), is the most prominent affective learning factor that can greatly influence users' interaction with an e-learning application. Intrinsic motivation to learn is proposed as the anchor for the development of a new usability measure for e-learning design. Table 1 exhibits the focus of this study that responds to the need for affective considerations.

Affective HCI and implications for usability (Hudlicka, 2003)	**Affective learning and usability in this study**
Importance of affect: refers to identification of HCI contexts where affect is critical and must be addressed	**Importance of affect is critical and must be addressed in e-learning context.**
Selection of emotions: refers to which emotions should be considered in which context and for which types of users and so on.	***Motivation to learn* is selected in this study amongst other affective states and emotions;**
Measurement: focuses on how can existing usability criteria be augmented to include affective considerations.	This study combines *web usability* attributes with *instructional design* attributes and integrates them with *motivation to learn*.

Table 1: Affective considerations

According to the above, the purpose of this research is to develop and empirically test a questionnaire-based usability evaluation method for e-learning applications. This method addresses the user as a learner and proposes motivation to learn as a new type of e-learning usability measurement. The development of the method was based upon a very well known methodology in HCI research and practice. Two large empirical studies were conducted examining usability of e-learning applications in authentic environments, with real learners-employees in corporate settings. The focus was on empirical assessment of the reliability and validity of the method. Results provide significant evidence for reliability and validity of the method. A synopsis of the use of questionnaires in e-learning usability is presented in the next section followed by the main body of this work, the description of method's development. The paper concludes with a summary of results and limitations and discusses future research work.

2. Related work: The use of questionnaires in e-learning usability studies

Questionnaires have been widely used for both evaluating affect and usability of interactive systems. Several studies have employed questionnaires as tools for assessing an array of affective states such as motivation (Keller and Keller, 1989), curiosity, interest, tiredness, boredom etc. (Whitelock and Scanlon, 1996) or achievement motive, creativity, sensation seeking etc. (Matsubara and Nagamachi, 1996). Regarding the use of questionnaires in usability research, the main advantage is that a usability questionnaire provides feedback from the point of view of the user. In addition usability questionnaires are usually quick and cost effective to administer and to score. A large amount of data can be collected and such data can be used as a reliable source to check if quantitative usability targets have been met. In the same vein Root and Draper (1983) argued that questionnaire methodology clearly meets the requirements of inexpensiveness and ease of application. Such requirements are of great importance in e-learning context; e-learning practitioners need a short, inexpensive and easy to deploy usability evaluation method so that e-learning economics can afford its use (Feldstein, 2002, Zaharias, 2004).

As far as concerns the development of usability questionnaires tailored to e-learning, much of the work conducted is either anecdotal or information about empirical validation is missing. Thus, it is not surprise that most e-learning researchers and practitioners usually employ some of the well-known and validated satisfaction questionnaires. Such questionnaires or variations of them have been used in several e-learning studies: Parlangeli et al. (1999) and Chang (2002) used a variation of

Questionnaire for User Interface Satisfaction (QUIS) (Shneiderman, 1987, Chin et al., 1988) and Avouris et al., (2001) used a variation of Website Analysis and MeasureMent Inventory (Kirakowski et al., 1998). In addition Avouris (1999) customized QUIS usability questionnaire for evaluation of educational material. The questionnaire contains 75 questions grouped in 10 separate categories. These categories are: 1) general system performance, 2) software installation, 3) manuals and on-line help, 4) on-line tutorials, 5) multimedia quality, 6) information presentation, 7) navigation, 8) terminology & error messages, 9) learnability and 10) overall system evaluation.

Other studies report some attempts towards the development of new questionnaires customized for e-learning; Wade and Lyng (1999) developed and used a questionnaire with the following criteria: a) naturalness, b) user support, c) consistency, d) non-redundancy and e) flexibility. Tselios et al. (2001) used a questionnaire with 10 items, which have been adapted from well-known heuristics. De Villiers (2004) developed a questionnaire by adapting "learning with software heuristics" proposed by Squires and Preece (1999). Such research efforts produce outcomes towards the direction of addressing the user as a learner but any other details about the development and psychometric properties of questionnaires such as reliability and validity are missing. Additionally most of these developments focus mainly on cognitive factors that influence the interaction with e-learning applications while they neglect the affective ones, whose importance is continuously increasing.

3. Method Development

The Questionnaire Design Methodology suggested by Kirakowski and Corbett (1990) was followed throughout the development of the proposed method. This methodology contains five stages: 1) Forming the survey, 2) Item sampling, 3) Pilot trial, 4) Production version and 5) Next version. This paper reports the work conducted along the first three stages, where reliability and validity of the proposed method is tested and established.

3.1 Forming the survey

Regarding the "Forming the survey" stage, two objectives need to be considered: a) the type, and b) the scope of the questionnaire. A psychometric-type of questionnaire was considered as the most suitable method since the major objective was to measure users' perception of e-learning usability and related affect (in this case motivation to learn).

As far as concerns the scope of the questionnaire, it has to be noted that this research focuses on usability evaluation of asynchronous e-learning applications as adult learners use them in corporate settings for training purposes. In addition this method extends the current practice by focusing on motivation to learn as an important affective learning dimension so as to address the user as a learner. This research relies mainly on Keller's (1983) theory and related model, which is perhaps the most influential model of motivation, in order to further analyze and interpret the motivation to learn construct. According to Keller (1983) motivation to learn construct is composed of four sub-constructs: attention, relevance, confidence and satisfaction. Prior to the development of the questionnaire a conceptual framework was established (Zaharias, 2005), which employs a) a combination of web design and instructional design parameters and b) motivation to learn construct (Keller, 1983). According to the conceptual framework usability parameters that were chosen from an array of studies for inclusion into the questionnaire are: Navigation, Learnability, Accessibility, Consistency, Visual Design, Interactivity, Content & Resources, Media Use, Learning Strategies Design, Instructional Feedback, Instructional Assessment and Learner Guidance & Support.

Thus a new set of usability parameters (derived from web usability and instructional design literature) for e-learning is proposed along with a new type of usability measurement: motivation to learn.

3.2 Item sampling

The design parameters included in the conceptual framework were the main constructs included in the questionnaire. These constructs were measured with items adapted from prior research. To identify items for possible inclusion in the questionnaire, an extensive review of prior studies referring to e-learning and web design was conducted. More specifically a number of web and instructional design guidelines (Lynch and Horton, 1999, Weston et al., 1999, Nielsen, 2000, Johnson and Aragon, 2002) have been reviewed, as well as a number of usability evaluation heuristics, checklists and questionnaires (Quinn, 1996, Horton, 2000, Reeves et al., 2002). Items were carefully selected so that to cover all parameters included in the conceptual framework. The items in the questionnaire were presented in groups relating to each parameter; the aim of this questionnaire was to capture usability parameters that seem to have an effect on motivation to learn when measuring the

usability of e-learning courses rather than to develop an equal scale of each parameter (i.e. parameters represented by an equal number of items). The questionnaire development started with an initial item pool of over 90 items. The items were examined for consistency of perceived meaning by getting 3 subject matter experts to allocate each item to content areas. Some items were eliminated when they produced inconsistent allocations. Prior to completion of the first version of questionnaire, a mini pilot test was undertaken to ensure that items were adapted and included appropriately in the questionnaire. An online version of the questionnaire was administered to 15 users in academic settings (2 of them were usability experts) that had some prior experience with e-leaning courses. Data obtained was analyzed mainly for response completeness; some adjustments were made and subsequently some items were reworded. The whole procedure led to the first version of questionnaire, which consisted of 64 items: 54 items measuring web and instructional design usability and 10 items measuring motivation to learn. Criteria corresponding to each usability parameter were assessed on a 5 point scale, where the anchors were 1 for strongly disagree and 5 for strongly agree (table 5). Additionally an option "Not Applicable" was also included outside the scale for each item included in the questionnaire. Such an option has been used in several usability questionnaires (Lewis, 1995). There was also space for free-form comments. The first version of questionnaire was tested in pilot trial 1.

3.3 Pilot trial 1

The first version of the questionnaire was used and empirically tested during an international research project. This project aimed at enhancing Information and Communication Technologies' skills (ICT) in South-Eastern (SE) European countries. The strategic goal of this project was to set up an e-learning service that provides e-learning courses on ICT skills and competences. Four user-organizations representing four different countries in the region participated in the project. These user organizations used the e-learning service in order to train a subset of their employees (mostly IT professionals) in the following topics: "IT Business Consultancy" and "Software and Applications Development". The main pillars of the e-learning service that was developed during the project were two asynchronous e-learning courses covering the two aforementioned topics respectively.

3.3.1 Subjects and method for testing

A summative type of evaluation was conducted after the implementation of the e-learning service. Usability evaluation of the e-learning courses was a main component of the summative evaluation. The trainees (employees of user organizations) that participated in the project interacted with the e-learning service and the e-learning courses during a period of four months and then they were asked to participate in the summative evaluation study. In this study a web-based version of the questionnaire was developed, which is particularly useful in web usability evaluation when the users are geographically dispersed (Tselios et al., 2001). The online questionnaire targeted the whole trainee population who used the e-learning service and it was released right after the end of the project's formal training period. The respondents were asked to evaluate the e-learning courses that had already used and interacted with. They self-administered the questionnaire and for each question, were asked to circle the response which best described their level of agreement with the statements. The total number of trainees who participated in usability evaluation was one 113 (63 male and 50 female).

3.3.2 Analysis and results

First a factor analysis was conducted in order to identify the underlying dimensions of usability parameters of e-learning courses as perceived by the trainees. The Kaiser-Mayer-Olkin (KMO) Measure of Sampling Adequacy was 0.846, which is comfortably higher than the recommended level of 0.6 (Hair at al., 1998). A principal components extraction with Varimax rotation was used. Using a criterion of eigenvalues greater than one an 8-factor solution was extracted explaining 65,865% of the variance (table 2). In order to assess the internal consistency of the factors scales Cronbach's Alpha was utilized.

As table 2 exhibits all factors show high internal consistency as indicated by high Alpha coefficients (ranges from 0.707 to 0.901), which exceed the recommended level of .70 (Lewis, 1995, Hair et al., 1998). In addition the composite variable *Motivation to Learn* shows a very high internal consistency as Alpha coefficient indicates (a =0.926). After conducting factor and reliability analysis a new version of the questionnaire (version 2) was derived. Data analyses led to the refinement of the questionnaire and a more parsimonious solution has been reached with 8 factors representing usability parameters of e-learning courses: Interactive content, Instructional Feedback & Assessment, Navigation, Visual Design, Learner

Guidance & Support, Learning Strategies Design, Accessibility and Learnability.

It must be noted that the first version of the questionnaire contained 54 items representing 12 usability parameters and 10 items representing the *Motivation to Learn* construct, a total of 64 items.

Factors	**Reliability Cronbach Alpha**	**Eigenvalue**	**Percentage of variance explained**
Interactive content	$\propto$ = 0.901	15,229	36,259
Instructional Feedback & Assessment	$\propto$ = 0.852	3,090	7,357
Navigation	$\propto$ = 0.801	2,341	5,573
Visual Design	$\propto$ = 0.793	1,798	4,280
Learner Guidance & Support	$\propto$ =0.805	1,549	3,688
Learning Strategies Design	$\propto$ =0.821	1,354	3,223
Accessibility	$\propto$ = 0.718	1,169	2,784
Learnability	$\propto$ = 0.707	1,134	2,700
Percentage of total variance explained			65,865

Table 2. Factor and reliability analysis in pilot trial 1.

The second version of the questionnaire contains 41 items representing 8 usability parameters (the 8 factors extracted as already presented) and 10 items representing the *Motivation to Learn* construct, a total of 51 items. This new version of the questionnaire, which shows a very high reliability as measured by Cronbach Alpha coefficient (0.961) is used and empirically tested in the second large-scale study (pilot trial 2) to be described in the next section.

3.4 Pilot trial 2

The second large-scale study was conducted in collaboration with a large private Greek organization that operates in banking industry. The organization has an extensive network of branches all over Greece and trainees were geographically distributed. The organization had implemented e-learning services and numerous e-learning courses during the last two years. The total number of e-learning courses was 23. Some of them were offered only through the e-learning mode (asynchronous e-learning courses) while some others were offered as supplementary to traditional, classroom-based courses.

3.4.1 Subjects and Method for testing

The second version of the questionnaire was paper-based and it was delivered in Greek language. The target population of the study was the whole set of employees that were trained via e-learning courses. Thus, each participant in the study had experience in using e-learning courses. The questionnaire version 2 was sent to the respective branches and the respondents (employees of these branches) were asked to evaluate the e-learning courses that had already used and interacted with. They self-administered the questionnaire and for each question, were asked to circle the response which best described their level of agreement with the statements. Out of the 500 questionnaires that were distributed, 260 complete responses were returned and thus the response rate was 52%. Four responses were considered as invalid while cleaning the data; therefore data from two 256 respondents were actually used in the analysis. Among them 110 were male and 146 were female.

3.4.2 Analysis and results

Forty-one items representing eight usability parameters were factor analyzed using principal components method with a Varimax rotation. The Kaiser-Mayer-Olkin (KMO) Measure of Sampling Adequacy was 0.914, which is comfortably higher than the recommended level of 0.6 (Hair et al., 1998). The following criteria were used in extracting the factors: a factor with an eigenvalue greater than one and factors that account for a total variance at least 50% would be selected. Further, only items with a factor loading greater than 0.32 would be included in each factor grouping. After performing two iterations of factor analysis, the pattern of factor loadings suggested that a seven-factor solution should be extracted. The seven-factor solution explained 54% of the variance. These factors were: Content, Learning & Support, Visual Design, Navigation, Accessibility, Interactivity and Self Assessment & Learnability. In order to assess the internal consistency of the factors' scales Cronbach's Alpha was utilized.

According to the above analysis, it is evident that all factors show high internal consistency (table 3) as indicated by high Alpha coefficients, which exceed the recommended level of 0.70 (Lewis, 1995, Hair et al., 1998). In addition the composite variable *Motivation to Learn* shows a very high internal consistency as Alpha coefficient indicates (α=0.900). After conducting factor and reliability

analysis a new more parsimonious solution emerged and a new version (version 3) of the questionnaire was derived. The overall alpha for the new version of questionnaire was very high, a=0.934. This version of questionnaire contains 39 items measuring e-learning usability parameters plus 10 items measuring *Motivation to Learn*, a total of 49 items (table 4 presents an excerpt of the questionnaire version 3).

Factors	**Reliability Cronbach Alpha**	**Eigenvalue**	**Percentage of variance explained**
Content	a = 0.824	12.277	30.691
Learning & Support	a = 0.868	2.530	6.326
Visual Design	a = 0.775	1.684	4.211
Navigation	a = 0.762	1.510	3.774
Accessibility	a = 0.734	1.313	3.282
Interactivity	a = 0.782	1.194	2.986
Self-Assessment & Learnability	a = 0.724	1.087	2.716
Percentage of total variance explained			53.986

Table 3. Factor and reliability analysis in pilot trial 2.

4. Validity analysis

The validity of a questionnaire concerns what the questionnaire measures and how well it does so. There are three types of validity: content validity, criterion-based validity and construct validity (Anastasi and Urbina, 1997). Unlike criterion validity, evidence for construct validity cannot be obtained from a single study, but from a number of inter-related studies (Saw and Ng, 2001). In the following, efforts made to assess content and criterion validity are presented.

As far as concerns content validity, the usability attributes were thoroughly examined and chosen based on the HCI and more specifically on web usability literature, and instructional design literature. An extensive literature review was conducted in order to select the appropriate usability attributes; items for inclusion within the questionnaire were selected from a wide range of web course design guidelines, checklists and questionnaires. The factor analyses that were performed on data collected during trial 1 and trial 2 support the content validity, since meaningful unitary constructs emerged.

Furthermore criterion validity can take two forms: concurrent and predictive validity. Concurrent validity is a statistical measure in conception and describes the correlation of a new instrument (in this case the questionnaire) with existing instruments, which purport to measure the same construct (Rust and Golombok, 1989). No attempt to establish concurrent validity was done. The main reason is the lack of other research-validated usability questionnaires that measure *Motivation to Learn* for cross-validating purposes.

Regarding predictive validity a multiple regression analysis was performed during pilot trial 1. The main objective was to assess the efficacy and effectiveness of the proposed usability parameters in explaining and predicting *Motivation to Learn*. In this research, the proposed usability parameters (i.e. the factors identified in factor analysis) are the independent variables (IVs) and the composite variable *Motivation to Learn* is the dependent variable (DV). The composite dependent variable was consisted of the ten items used to measure *Motivation to Learn*. All independent variables were entered into the analysis simultaneously in order to assess the predictive strength of the proposed model. When all independent variables entered into the multiple regression model results showed an R square of 0.716 and adjusted R-square of 0.691 (table 5). An analysis of variance revealed an F (8,93) of 29.264, which is statistically significant ($p < .001$). These findings reveal that the eight usability parameters (factors extracted from factor analysis in pilot trial 1) when entered together in the regression model, accounted for 71.6% (Adjusted R Square 69.1%) of the variance in *Motivation to Learn*. Such findings delineate good results for the questionnaire and can be considered as preliminary evidence of the validity of the proposed method.

Model Summary				
Model	R	R Square	Adjusted R Square	Sig. F Change
1	,846	,716	,691	,000
Predictors: (Constant), Learnability, Instructional Feedback & Assessment, Navigation, Accessibility, Learning Strategies Design, Visual Design, Interactive content, Learner Guidance & Support				

Table 5. Usability parameters predicting motivation to learn

Construct validity has two components (Anastasi and Urbina, 1997, Rust and Golombok, 1989): *convergent validity*, which demonstrates association with measures that are or should be related, and *divergent or discriminant validity*, which demonstrates a lack of association with measures

Criteria	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree	NA
Content						
Vocabulary and terminology used are appropriate for the learners.						
Abstract concepts (principles, formulas, rules, etc.) are illustrated with concrete, specific examples.						
Learning & Support						
The courses offer tools (taking notes, job-aids, recourses, glossary etc.) that support learning						
The courses include activities that are both individual-based and group-based.						
Visual Design						
Fonts (style, color, saturation) are easy to read in both on-screen and in printed versions						
Navigation						
Learners always know where they are in the course.						
The courses allow the learner to leave whenever desired, but easily return to the closest logical point in the course.						
Accessibility						
The course is free from technical problems (hyperlink errors, programming errors etc.)						
Interactivity						
The courses use games, simulations, role-playing activities, and case studies to gain the attention, and maintain motivation of learners.						
Self-Assessment & Learnability						
Learners can start the course (locate it, install plug-ins, register, access starting page) using only online assistance						
Motivation to learn						
The course incorporates novel characteristics						
The course stimulates further inquiry						
The course is enjoyable and interesting						
The course provides learner with frequent and varied learning activities that increase learning success						

Table 4. An excerpt of the questionnaire version 3.

that should not be related. Unlike criterion validity, evidence for construct validity cannot be obtained from a single study, but from a number of inter-related studies (Saw and Ng, 2001). Nevertheless a first attempt to assess construct validity was made within the context of this research. A variation of "*Multitrait Multimethod Matrix*" technique as suggested by Campbell and Fiske (1959) was used. Practically according to this technique, convergent validity tests whether the correlations between measures of the same factor are different than zero and large enough to warrant further investigation of discriminant validity. In order to assess convergent validity the smallest within-factor correlations are estimated. The correlation analysis revealed the smallest within-factor correlations: Content =0.230; Learning & Support=0.346; Visual Design=0.411; Navigation=0.289; Accessibility=0.486; Interactivity=0.264; and Self-Assessment& Learnability=0.326. These correlations are significantly different than zero ($p < 0.01$) and large enough to proceed with discriminant validity analysis.

Discriminant validity for each item is tested by counting the number of times that the item correlates higher with items of other factors than with items of its own theoretical factor. For discriminant validity, Campbell and Fiske (1959) suggest that the count should be less than one-half the potential comparisons. According to the correlation matrix and the factor structure there are 690 potential comparisons. A careful examination of correlation matrix revealed 356 violations of the discriminant validity condition, which is slightly higher than the one-half the potential comparisons (345). This cannot be considered as a major problem since –as already mentioned- construct validity cannot be obtained from a single study but from a number of inter-related studies. Therefore further examination of construct and more specifically divergent validity is an issue of further research.

5. Discussion and further work

5.1 Summary of the results & limitations

Motivated by the need to address the specificities of e-learning design a usability evaluation method was developed. The proposed questionnaire-based usability evaluation method extends conventional web usability criteria and integrates them with criteria derived from instructional design so that to address specificities of e-learning design and address the users as learners. The proposed method also extends the current practice in usability evaluation by measuring users' affective engagement and proposing *motivation to learn* as a new type of usability measurement. This new type of measurement has been tested for reliability and validity: overall internal consistency of the questionnaire is very high while content and criterion validity have been adequately achieved.

Besides the accomplishments of this study there are still certain limitations that practitioners should be aware of. The first limitation has to do with the use of the questionnaire as a method to assess affective states. Questionnaires have been accused of being static methods that cannot easily detect more transient respondents' characteristics; further social-emotional expectations and awareness of the respondent can greatly influence what is reported. In addition the questionnaire was focused only on asynchronous e-learning applications since this kind of delivery mode is the dominant approach nowadays. Finally further studies are needed to provide more empirical evidence regarding construct validity of the method. It is widely recognized that validation is a process (Hornbaek, 2005) and adequate evidence for construct validity cannot be obtained from a single study but from a number of inter-related studies (Anastasi and Urbina, 1997, Saw and Ng, 2001).

5.2 Future studies

Firstly future studies can be designed in order to address the above limitations. As already mentioned using a questionnaire as a method to assess an affective state has advantages and some weaknesses as well. Future research efforts can employ a combination with other methods so to gather other information about more transient characteristics and more qualitative usability data. A combination of methods can give stronger results. Such methods could include the use expert systems and sentic modulation, which is about detecting affective states through sensors such as cameras, microphones, wearable devices etc. (Picard and Daily, 2005).

Moreover, further consideration is needed to explore usability attributes and role of affect in synchronous e-learning courses and environments. Synchronous e-learning is characterized by different types of interactions through chats, real time audio, application sharing, whiteboards, videoconferencing etc., which imposes additional considerations concerning usability and its evaluation.

Regarding the additional evidence for reliability and validity, a modification of the third version of the questionnaire can be conducted by replacing

and re-wording some of the few items in the questionnaire which did not discriminate well and further test the method with different kinds of e-learning courses, different types of learners, different corporate environments. Alternative statistical analyses can also be used to shed light to reliability and validity issue; for example Confirmatory Factor Analysis can be used to determine convergent and divergent (or discriminant) validity (Wang, 2003). The advantages of applying CFA as compared to classical approaches to determine convergent and divergent validity are widely recognized (Anderson and Gerbing, 1997). Another idea to further assess reliability is to use "test – retest" reliability (Wang, 2003), which examines the stability of an instrument over time.

Besides the confrontation of the limitations future research can focus on the following:

- Use of the proposed questionnaire as a formative evaluation method: This paper described the use of the questionnaire-based usability evaluation method in two large-scale empirical studies as a summative evaluation method. The proposed usability evaluation method can also provide useful design guidelines during the iterative design process as a formative evaluation method. Currently, the proposed usability evaluation method can point towards specific usability problems with an e-learning application. A more systematic exploitation of using such method for formative evaluation can be realized through the development of a database where the results of a number of different usability studies can be stored so that the knowledge obtained can be re-used.
- Benchmarking: The proposed questionnaire-based usability evaluation method can also provide benchmark information like other research-validated questionnaires (for example WAMMI) do so. This means practically that the usability of an e-learning application could be tested against others. A standardized database can be developed that contains the usability profiles of existing e-learning applications and, thus, can facilitate designers compare the usability of one application with a series of other e-learning applications.
- Next Version of questionnaire: focusing on other affective/ emotional states. Future research should seek a deeper understanding of the design issues that influence learners' affect and emotions. Emotions such as fear, anxiety, apprehension, enthusiasm and excitement as well as pride and embarrassment, (Ingleton and O'Regan, 1998, O'Regan, 2003) along with the flow experience (Csikszentmihalyi, 1975, 1990, Konradt and Sulz, 2001) can provide significant input to e-learning design and shed light in explaining learners' behaviour; such emotions and their assessment can also be taken into consideration in the next version of the questionnaire.

Acknowledgments

The author wishes to thank all the trainees - participants in the empirical studies, as well as all the partners in the international project during pilot trial 1 and the Greek private bank (pilot trial 2) for their collaboration.

References

Albion, P. R. (1999). Heuristic evaluation of educational multimedia: from theory to practice. In *Proc. of 16th Annual conference of the Australasian Society for Computers in Learning in Tertiary Education*, ASCILITE, 1999.

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. 7th Edition, Prentice Hall.

Anderson, J. C., & Gerbing, D. W. (1997). Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin* 103, 3 (1997) pp.411–423.

Avouris, N., Tselios, N. Fidas, C. & Papachristos, E. (2001). Website evaluation: A usability-based perspective. In *Proc. of 8th Panhellenic Conference on Informatics*, vol II, pp. 108-117, November 2001.

Avouris, N. (1999). *Educational Software Evaluation Guidelines*, Technical Report, Working Group on Educational Software Evaluation Procedure, Pedagogical Institute, Athens, 1999.

Campbell, D.R., & Fiske, D.W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin*. 56, (2), pp. 81–105.

Chang, Y. (2002). Assessing the Usability of MOA Interface Designs. In *Proc. of eLearn 2002*, Vol. 2002, Issue. 1, pp.189-195.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the Human-Computer Interface. In *Proc. of CHI'88: Human Factors and Computing*

Systems. ACM Press/Addison Wesley, (1988), pp.213-218.

Clark, R. C., & Mayer, R. E. (2003). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. Jossey-Bass/Pfeiffer. San Francisco, California.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety: The experience of play in work and games*. San Francisco: Jossey-Bass.

Csikszentmihaly, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

De Villiers, R. (2004). Usability Evaluation of an E-Learning Tutorial: Criteria, Questions and Case Study. In *Proc. of the 2004 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries SAICSIT '04.*, 2004.

Feldstein, M. (2002). What is "usable" e-learning?. *eLearn*. Vol. 2002, (9).

Hair, J. F., Anderson, R. E., Tatham, R. L. T., & Black, W. C. (1998). *Multivariate data analysis*, Prentice-Hall Inc., Upper Saddle river, New Jersey.

Harun, M. H. (2002). Integrating e-Learning into the workplace. *Internet and Higher Education* 4, pp.301–310.

Hornbaek, K. (2005). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies* (in press).

Horton, W. (2000). *Designing Web-Based Training*. John Wiley & Sons.

Hudlicka, E. (2003). To feel or not to feel: the role of affect in human-computer interaction *International Journal of Human-Computer Studies*. 59, pp. 1-32.

Ingleton, C., & O'Regan, K. (1998). Recounting mathematical experiences: Exploring the development of confidence in mathematics. In *Proc. of the AARE (Australasian Association for Research in Education) Conference*.

Johnson, S., & Aragon, S. (2002). An Instructional Strategy Framework for Online Learning Environments. In *Proc. of World Conference on E-Learning in Corp., Govt., Health., & Higher Ed*. Vol. 2002, (1), pp.529-536.

Keller, J.M. (1983). Motivational design of instruction. In Reigeluth, C. M. (ed.), *Instructional Design Theories and Models: An overview of their current status.* Hillsdale, NJ: Erlbaum.

Keller, J.M., & Keller, B.H. (1989). *Motivational delivery checklist*. Florida State University.

Kirakowski, J., Claridge, N., & Whitehead, R. (1998). Human Centered Measures of Success in Web Site Design. In *Proc. of Our Global Community*. http://www.research.att.com/conf/hfweb/proceedings/

Kirakowski, J. & Corbett, M. (1990). *Effective Methodology for the Study of HCI.* Elsevier.

Konradt, U., & Sulz, K. (2001) The Experience of Flow in Interacting With a Hypermedia Learning Environment. *Journal of Educational Multimedia and Hypermedia*. 10, 1, pp.69-84.

Lewis, J. R. (1995). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*. 7, 1, pp.57-78.

Lynch, P.J., & Horton, S. (1999). Interface design for WWW Web Style Guide. Yale Style Manual. http://info.med.yale.edu/caim/manual/interface.html .

Mankoff, J., Dey, A.K., Hsieh, G., Kientz, J., Ames, M., and Ledered, S. (2003). Heuristic evaluation of ambient displays. In *Proc. of ACM Conference on Human Factors in Computing Systems*. ACM Press, New York, NY, (2003), pp.169-176.

Matsubara, Y. & Nagamachi, M. (1996). Motivation systems and motivation models for intelligent tutoring, in Claude Frasson, et. al., (Eds.), in *Proc. of the Third International Conference in Intelligent Tutoring Systems*, (1996), pp.139-147.

Monk, A. (2002). Noddy's guide to usability. *Interfaces* 50, pp.31-33.

Nielsen, J. (2000). *Designing Web Usability: The Practice of Simplicity*. New Riders: New York.

Notess, M., 2001, Usability, User Experience, and Learner Experience. http://www.elearnmag.org/.

O'Regan, K. (2003). Emotion and E-Learning. *Journal of Asynchronous Learning Networks*. 7, 3, pp.78-92.

Parlangeli, O., Marchigianni, E., & Bagnara, S. (1999). Multimedia System in Distance Education: Effects on Usability. *Interacting with Computers*. 12, pp.37-49.

Picard, R. W. & Daily, S. B. (2005). Evaluating affective interactions: Alternatives to asking what users feel. *CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches*, April 2005.

Quinn, C.N. (1996). Pragmatic evaluation: lessons from usability. In *Proc. of 13th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*, (1996).

Reeves, T. C., Benson, L., Elliott, D., Grant, M., Holschuh, D., Kim, B., Kim, H., Lauber, E., and Loh, S. (2002). Usability and Instructional Design Heuristics for E Learning Evaluation. In *Proc. of World Conference on Educational Multimedia, Hypermedia and Telecommunications*. Vol. 2002, 1, pp.1615-1621.

Root, R. W., & Draper, S. (1983). Questionnaires as a software evaluation tool. In *Proc. of CHI 83,* New York: NY: ACM, pp.83-87.

Rust, J., & Golombok, S. (1989). *Modern Psychometrics: The Science of Psychological Assessment*. Routledge, London and New York.

Saw, S.M., & Ng, T.P. (2001). The Design and Assessment of Questionnaires in Clinical Research. *Singapore Med Journal*. 42, 3, pp.131-135.

Schunk, D. (2000). *Learning theories: An educational perspective* (3rd Ed). Upper Saddle river, NJ: Prentice-Hall.

Shneiderman, B. (1987). Designing the user interface: Strategies for effective human-computer interaction. Reading, MA: Addison-Wesley.

Soloway, E., Guzdial, M., & Hay, K. (1994). Learner-Centered Design. The Challenge For HCI In The 21st Century. *Interactions*. 1, 2, pp.36-48.

Squires, D. (1999). Usability and Educational Software Design: Special Issue of Interacting with Computers. *Interacting with Computers*. 11, 5, pp.463-466.

Squires, D., & Preece, J. (1999). Predicting quality in educational software: Evaluating for learning, usability and the synergy between them, *Interacting with Computers*. 11, 5, pp.467-483.

Tselios N., Avouris N., Dimitracopoulou A, & Daskalaki S. (2001). Evaluation of Distance-learning Environments: Impact of Usability on Student Performance, *International Journal of Educational Telecommunications*, vol 7, 4, pp.355-378.

Wade, V. P., & Lyng, M. (1999). Experience Improving WWW based Courseware through Usability Studies. In *Proc. of World Conference on Educational Multimedia,Hypermedia and Telecommunications*. Vol. 1999, 1, pp.951-956.

Wang, Y.S. (2003). Assessment of learner satisfaction with asynchronous electronic learning systems. *Information & Management*. 41, pp.75–86.

Weston, C., Gandell, T., McApline, L., & Filkenstein, A. (1999). Designing Instruction for the Context of Online Learning, *The Internet and Higher Education*. 2, 1, pp.35-44.

Whitelock, D. & Scanlon, E. (1996). Motivation, media and motion: Reviewing a computer supported collaborative learning experience. In Brna, P., et. al. (Eds.), *Proc. of the European Conference on Artificial Intelligence in Education*.

Zaharias, P. & Poulymenakou, A. (2006). Implementing Learner-Centered Design: The interplay between usability and instructional design practices. *Journal of Interactive Technology and Smart Education* (in press).

Zaharias, P. (2004). Usability and e-Learning: The road towards integration. *ACM eLearn Magazine*, Vol.2004 (6).

Zaharias, P. (2005). E-Learning Design Quality: A Holistic conceptual framework. In *Encyclopedia of Distance Learning*. Edited by Howard et al., Idea Group Inc., New York.