INTERNATIONAL JOURNAL OF
HUMAN COMPUTER INTERACTION (IJHCI)

Publication Frequency: 6 Issues / Year

INTERNATIONAL JOURNAL OF HUMAN COMPUTER INTERACTION (IJHCI)

VOLUME 4, ISSUE 2, 2013

EDITED BY
DR. NABEEL TAHIR

INTERNATIONAL JOURNAL OF HUMAN COMPUTER INTERACTION (IJHCI)

CSC Publishers, 2013

EDITORIAL PREFACE

This is *Second* Issue of Volume *Four* of International Journal of Human Computer Interaction (IJHCI). IJHCI is an International refereed journal for publication of current research in Human Computer Interaction. Publications of IJHCI are beneficial for researchers, academics, scholars, advanced students, practitioners, and those seeking an update on current experience, state of the art research theories and future prospects in relation to applied science. Some important topics covers by IJHCI are affective computing, agent models co-ordination and communication, computer mediated communication, innovative interaction techniques and user interface prototyping for interactive systems etc.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Started with Volume 4, 2013, IJHCI appears with more focused issues related to human computer interaction studies. Besides normal publications, IJHCI intend to organized special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

This journal publishes new dissertations and state of the art research to target its readership that not only includes researchers, industrialists and scientist but also advanced students and practitioners. IJHCI seeks to promote and disseminate knowledge in the applied sciences, natural and social sciences industrial research materials science and technology, energy technology and society including impacts on the environment, climate, security, and economy, environmental sciences, physics of the games, creativity and new product development, professional ethics, hydrology and water resources, wind energy.

IJHCI editors understand that how much it is important for authors and researchers to have their work published with a minimum delay after submission of their papers. They also strongly believe that the direct communication between the editors and authors are important for the welfare, quality and wellbeing of the Journal and its readers. Therefore, all activities from paper submission to paper publication are controlled through electronic systems that include electronic submission, editorial panel and review system that ensures rapid decision with least delays in the publication processes.

To build its international reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc, Scribd, CiteSeerX and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJHCI. We would like to remind you that the success of our journal depends directly on the number of quality articles submitted for review. Accordingly, we would like to request your participation by submitting quality manuscripts for review and encouraging your colleagues to submit quality manuscripts for review. One of the great benefits we can provide to our prospective authors is the mentoring nature of our review process. IJHCI provides authors with high quality, helpful reviews that are shaped to assist authors in improving their manuscripts.

Editorial Board Members
International Journal of Human Computer Interaction (IJHCI)

EDITORIAL BOARD

TABLE OF CONTENTS

Pages

A Case Study on Non-Visual Pen-Based Interaction with the Numerical Data

Tatiana G. Evreinova   *Tatiana.G.Evreinova@uta.fi*
School of Information Sciences
University of Tampere
Tampere, 33014, Finland

Grigori Evreinov   *Grigori.Evreinov@uta.fi*
School of Information Sciences
University of Tampere
Tampere, 33014, Finland

Roope Raisamo   *Roope.Raisamo@uta.fi*
School of Information Sciences
University of Tampere
Tampere, 33014, Finland

Abstract

A widespread tabular format still poses great problems for screen readers because of diversity and complexity of the cells' content. How to access the numerical data presented in tabular form in a quick and intuitive way in the absence of visual feedback? We have implemented and assessed the algorithm supporting an exploration of the tabular data in the absence of visual feedback. This algorithm helps to solve the most commonly encountered problems: retrieving the position of extreme values and the target value that can also be linked to the specific content of the virtual table. The performance of 11 blindfolded subjects was evaluated when they used the StickGrip kinesthetic display and when they relied on the Wacom pen and auditory signals. The results of the comparative study are reported.

Keywords: Numerical Data, Pen-based Interaction, Kinesthetic Display.

1. INTRODUCTION

Douglas and his co-authors [1] have pointed out that according to statistical data one out of every three working blind adults have to use spreadsheets and databases on a daily basis. Visually impaired people still find it very challenging to access to composite tabular information through commercially available assistive technologies. Screen reading the table on a cell-by-cell basis can be used when the number of records (cells) is limited to a reasonable amount of data. Listening to the long sequences of the monotonous data can cause a loss of concentration on the main task, overload the person with unnecessary details and such a reading leads to frustration. Recently Burton and Lilly [2] assessed the Huntington National Bank website with different screen readers for navigating the tables and forms for bill paying and transferring funds. They found that this site was well accessible with JAWS, Window-Eyes, System Access, NVDA (NonVisual Desktop Access) and Thunder. As a result, they concluded that NVDA was the winner to support different online banking procedures.

Nevertheless, some companies (Jyske Netbank of Denmark and other international units of Danske Bank in Finland, Estonia, Latvia, Lithuania, Sweden, Norway, Ireland, UK, Germany, Russia) motivated by desire to achieve better security of the Online Banking service increase the length of identifiers and the method of logging in to their system. At that, they often ignore the abilities of the customers. In particular, bank identifiers can consist of a user ID, a personal

password and an additional security card number indicated via a key number printed in the security card (Figure 1). Moreover, if the person is not able to enter the required code correctly and within a certain time interval the security number cannot be identified then the system will require repeating the identification procedure for the new key number. Of course, the card content could be scanned (captured by camera phone), converted into the tabular format and accessed with the screen reader. However, reading of the entire security table (containing ninety 10-digit numbers) might be time consuming and not safe at all. The procedure of access to the specific number should also be optimized in terms of preventing an "observation attack" including any means of possible registration of accompanying physical signals.

Beyond the online banking service and other minor problems detected with different screen-readers [2], a slow exploration of large tabular data sets and the cells with a complex content hampers everyday activities and solving practical tasks, vocational and educational issues in the absence of visual feedback. By the large data sets, we mean tens of columns and tens of thousands of records when a combination of several digits could not be a unique as a key number in the security card.

Let us look at a few more examples of the tabular data sets. A cell of the table can display different types of data including textual description, pictures and a set of different but somehow linked together features [3]. In particular, the number 38 (Strontium) can be found 9 times in interactive periodic table (http://www.ptable.com/), an abbreviation of "Ta" (Tantalum) – 14 times if it was retrieved by the search engine and so on. The data collected in cells of the table can be the results of measurements, statistic functions, descriptions of the products or their components, vehicles, furniture, dress, apartments, etc. When the person wishes to get an overview of the empirical data or the product portfolio, what s/he needs to know in the first step?

These are the extreme values, the most expensive and the cheapest products, the products with the specific features. Nobody wishes to scan the entire table records. Instead, sighted people try to use sorting or another way of visualization, which could facilitate even visual evaluation of data: the extreme and average values, their trend and variation/volatility.

The widespread opinion that to analyze the data the person should explore perceptually all the records by sight, by hearing, or by touch is becoming obsolete. The tables were invented as a visual means facilitating data sorting and analysis. Since then, there have been developed a variety of methods and algorithms to analyze the data and present a result of the data processing in a compact and efficient way in any modality. Non-visual navigation in tables and spreadsheets based on the direct data conversion into another accessible modality and efforts spent to develop the tools to overview the specific information and to maintain the various procedures of the data processing would become useless, as such a way is often not efficient and frustrating [4], [5], [6]. To easily infer indicative features, patterns and trends, the data records have to be grouped and sorted and these procedures should not be seen as continuous efforts of the person. The tabular data can be converted into the multidimensional array and be further processed by an appropriate software application with respect to the task assigned by the user [3]. The information-processing medium is intended and must be designed to meet the needs of the users, by reducing the perceptual and cognitive load when dealing with any type of data presented in any format.

Another great problem with the numerical data manipulation stems from devices (tablet PC, smart phones and other mobile gadgets), which are exclusively optimized for visual-based touch input. At that, the touch sensitive surface cannot provide intuitive haptic feedback as mechanical keys of the ordinary keyboard. That is the question of how to access the numerical data presented in the tabular form in a quick and intuitive manner in the absence of visual feedback remains open.
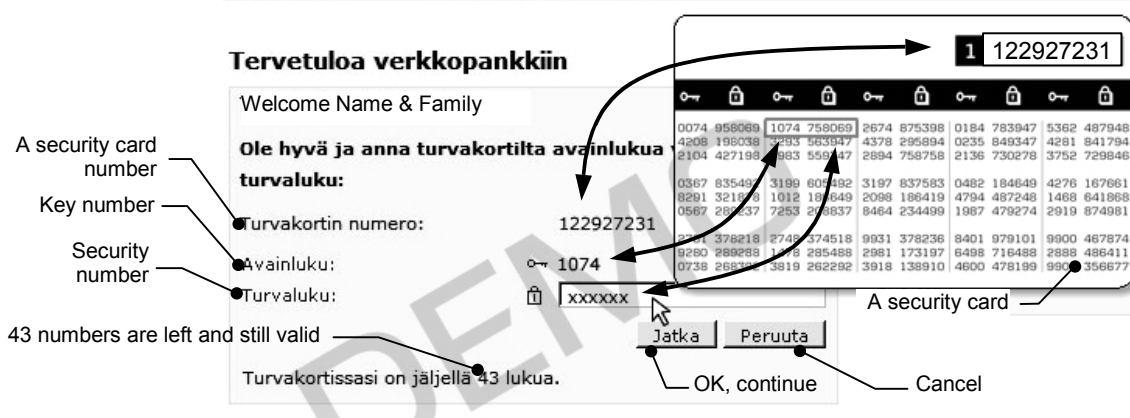
FIGURE 1: Demo shows the use of the security card when logging in to the online banking system.

2. BACKGROUND

The tabular data are usually grouped spatially and functionally. The spatial layout of the table can provide hyper and metaphorical information to the sighted person that often can not be directly extracted from the content of the individual cells (see e.g., periodic table of elements). Therefore, the navigation and selective reading the content of separate cells aggravate mental integration of the tabular data. The latter makes the relationships between the values of the cells unintelligible [7]. Spatially distributed tabular data require different overview and attentional mechanisms to explore and comprehend the content and the features of the layout. Many features have to be translated into accessible forms while preserving their relationships and generalization – spatial and functional grouping of the records of numerical, textual, graphical, audio, and other types of data values [5], [8]. At that, an alternative visualization of the hyper and metaphorical information has to demand a minimum of perceptual resources from an observer.

The vast majority of the previous research work was based on the cross-modal data transformation and focused on the use of sonification, including non-speech sounds, sound spatialization and musical techniques to convert visual information into auditory (phonological) representations [9], [10], [11], [12], [13], [14]. Various research projects have been undertaken to convert visual presentation of the entire table directly into another perceptual modality. Different generic properties of non-speech sounds such as the pitch, loudness, timbre and time (repetition rate) have been tested and combined with numerical data, their main statistical features and exploratory behavior of the user to facilitate browsing and access to large sets of data. Converting tabular data to series of pitches with different frequencies resulted in better perception among the visually impaired than when other perceptual properties were employed for data sonification [15].

Ramloll with co-authors [11] introduced a sonification approach to overview 2D tabular data arranged in 26 rows and 10 columns. The person could listen to any column or row as a continuous non-speech sound sequence by using keys on the numeric keypad for navigation (arrow keys and the center key), shortcut keys (PgUp/PgDn, Home, End), modal keys (to select 'value', 'pitch' or 'labels' mode) and overview keys ("+" and Enter). In this approach, the listener has to perform a sophisticated listening analysis to discover a tendency in the sound pattern and, thus, retrieve the relationships between the values of the cells within each row or column. Nevertheless, as cognitive load factors are not simply additive, when the number of records (i.e., the cells to be analyzed) is great enough, the complexity of perceptual parsing and the mental workload required to perform perceptual filtering and comprehension tasks in the absence of visual feedback is enormously increasing. The proposed sonification concept is only a reasonable way for presenting a limited number of data [8] or indicating general tendencies as a result of the statistical analysis without specifying exact values.

To overcome difficulties of exploration of the numerical data using the traditional speech-based techniques consuming time and working memory, Kildal and Brewster undertook several attempts [16], [17] to speed it up by sonifying data values at once in a selected row or column in very rapid succession. Duration of sound per value was 60 ms with an interval of 20 ms between sounds of neighboring cells. However, to avoid problems with indistinguishable sounds, the technique was limited for sonification of only integer values between 21 and 108, which have corresponded to 88 musical tones. Moreover, only the main characteristics of the relatively complex sound pattern could be detected and identified by the subjects in more detail. Then, in [17], the authors decreased the duration of sound per value to 7 ms by making quick overviews, like the cross-sectional slices over the entire table, distributing the complete range of 66 MIDI notes of piano instrument. They were also able to sonify the selected parts of the data in row or column with adjustable sound parameters. To provide the user with kinesthetic feedback, the researchers mapped the table to a Wacom graphic tablet. Thus, users could navigate with help of the pen relying on the sense of the movement and relative position of their limbs. The results of research have shown that while listeners managed well to perceive general overview of the data series, they were not able to assess in detail integrative picture about the specific features of the rapidly played back sound sequences. In the absence of spatio-temporal synchronization (compare the rhythmic structure of music notation), rapidly played back sound sequences were perceived as separate auditory events. The authors concluded that proposed accessibility technique did not facilitate the task of exploration of the numerical data for blind and visually impaired users.

Stockman with colleagues [18], [19] brought forward another approach for supporting interactive sonification of spreadsheets and diagrams. They provided automatic sonification of the rows and columns and activated the sonification cursor to provide the user with an overview by approximating parameters of the neighbor area of the arbitrary number of cells of the spreadsheet. Exact information was delivered through verbal feedbacks using the screen reader. The authors demonstrated that the system which is tailored to the user requirements and can display on demand overviews of the data from a range selected will decrease cognitive load and will speed up the process of exploring groups of cells for particular features [7]. The drawback of this approach was that the user had to navigate with the sonification cursor by controlling four additional keys (E-up/D-down, S-left/F-right) with the left hand while operating with the PC cursor by the right hand and arrow keys. To synchronize behavior of two logic cursors (activate/superpose/lock together), four modification keys were used and more keys were required in order to control parameters of auditory feedback. Abundance of controls and conditional signals needed for operating the application significantly aggravated the usability of the system implemented.

In order to facilitate manipulating any types of the tabular data, retaining a reasonable level of redundancy to indicate the trend of the variable in question and a possibility of navigation across the data, the application has to provide more intuitive and efficient way of processing and displaying the information relevant to the user's task, query and intention. Why the person should waste a time by listening to many sounds (e.g., 66 or 88 MIDI notes, [10]) which is not a straightforward task to detect the cells containing the highest/lowest values? This is also inconsistent with the Human Factors Guidelines on the use of non-speech acoustic signals and auditory tones [20] specifying that only tones having a frequency (pitch) between 300 Hz and 3000 Hz (that is, less than 50 notes or 4 octaves) can be employed to be clearly distinguished by the human ear.

Wall and Brewster [21] also undertook attempt to augment non-visual exploration of the numerical data with haptic force feedback. The entire tabular data were converted into a virtual 3D surface (plot) and explored with the PHANToM force-feedback device.

Other methods, techniques and tools have recently been implemented to facilitate exploration and manipulating numerical data in the absence of visual feedback. Nevertheless, when multiple records should be accessible in the absence of visual feedback the time to perform analysis of information presented is a critical factor. That is, the time and efforts (perceptual, cognitive and

behavioral) should be compatible to those that the blind person spent to comprehend the text block of about the same size accessed with a screen reader, including a minimum of additional knowledge and skills required to use the technique. For instance, like the reading of about 100 words, detection of the extreme value in the data set of 100 records should not exceed 20 s [4], [5], [6], [22].

By analyzing the results reported in the literature, we came to conclusion that rather than to spend additional efforts for understanding the encoding semantics of the tabular layout, any transformations of the data should reduce their redundancy while retaining some features of the spatial arrangement and a possibility to observe the trend of the variable in question, if necessary. At that, the important issues raised by Colwell and Sjöström should be taken into consideration [23], [24]. The researchers argued that it is necessary to provide visually impaired and blind user with auxiliary guiding feedback. When the users employ new technique, they have to get confident about the task and the workspace being explored, otherwise, they might easily become lost with any approach.

For instance, the task of the person is to detect the position of the numerical value that is a key to a specific (linked) data or another content in the table. Sighted person could get an exact value using find method in Excel and by filtering out perceptually irrelevant information or by creating an array formula. However regarding the skills and knowledge required, creating an array formula in the absence of visual feedback is not an easy task especially for elderly people (see e.g., Excel Magic Trick [25]) even with a suitable screen reader. Another approach consists in the formalization of the human-computer dialog when the application or the input technique (at the level of driver) could mediate the interaction between the user and the data. In such a way, the data analysis has to begin from a specification of the task to the application-mediator, e.g., by activating the appropriate functionality (mode). Then, instead of exploration of original tabular data the person would be able to supervise the data analysis performed by the application-mediator by making movements of the stylus or finger along (or parallel to) the physical borders of the study area (e.g., along the touchscreen bezel). At that, the size of the physical space of interaction could be optimized to better fit it to the touch-sensitive surface (touch pad, touchscreen or touch tablet). So, the user could systematically explore intermediate information presented in a suitable form.

The goal of the study presented was to examine how fast and accurately in the absence of visual feedback people are able to analyze the numerical data presented in a tabular form by following the proposed methodology. We approached this question by measuring both the mean time spent by the participants to complete the task (to search for the extremes or the given value in the array), and the number of errors committed when different modalities of auxiliary signals were used: an auditory and haptic signals accompanying the pen-based interaction.

3. METHOD
3.1 Retrieving the Position of Extreme Values in the Data Array
The algorithm has to be easily adjustable to the human perceptual and cognitive resources. Let us examine the particular task to detect the position of extremes in the 2D numerical data array. At the first step, the user has to specify the task to the application-mediator by activating the appropriate functionality and to select the desired modality of auxiliary guiding feedbacks. As shown in Figure 2, the 2D data array can be imagined as a virtual table (e.g., composed of 10 rows and 10 columns), size of which is maximized to fit the table to the size of the bordered frame (study area). Then the person should rely on the specific features of the workspace (the tablet), being explored in the particular sequence. For instance, the user can move a stylus across the rows within the left or the right exploration areas to allow the application to detect the extreme values (a local minimum or maximum) by rows and to display these values in the preferred modality (auditory or haptic) when the Y-coordinate of the pointer location corresponds to the current row. Thus, instead of listening to or relying on haptic signals to perceive many numerical values which would present the translated content of the cells in each row, the user can get

intermediate information – a local minimum/maximum value in the current row. Sliding across the rows will generate a number of local extremes. Still, these values retain information regarding their location (the order) within the array.

When a number of rows and local extremes are limited to ten or less, all of them could be converted into non-visual signals to present their value and relative position with respect to the virtual rows. At that, if the length of the array is also limited to a reasonable number of elements, it would allow to apply the linear model for cross-modal transformations, that is, preserving the same scale and relationships between non-visual values of the signals. However, such an approach is relatively inefficient when the task requires to focus on the identification and localization of the global extremum and if the number of elements in an array is large enough (typically 1000 or more).
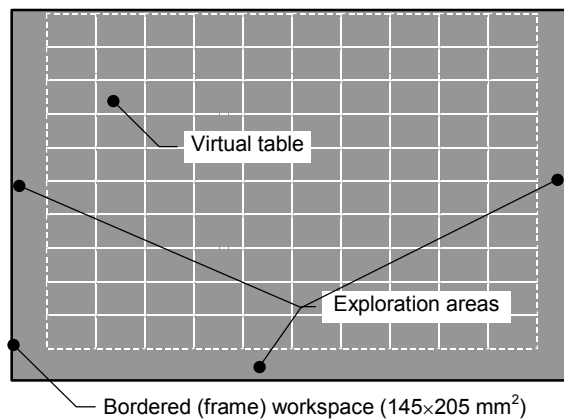
FIGURE 2: The application-mediator augmented with auxiliary guiding feedbacks projected onto the Wacom tablet. It was viewed on the monitor of the experimenter (see subsection 3.4).

Another solution consists in alternative imaging not the local extremes (their exact values) but the distance between the current extremum and the global extremum in the array. In such a way, the distance might be presented in non-visual modality using the following algorithm for imaging the distance to the target (global extremum) through a relatively limited number of non-visual signals (see also subsection 3.4 Table 1 and Table 2):

1) When the distance between a local extremum and the global extremum is less than 20 units of measurement of the values stored in the array then every value of the distance can be translated into a non-visual signal;

2) When the distance between a local extremum and the global extremum is more than 20 units and less than 50 units of measurement of the values stored in the array then every third value of the distance can be translated into a non-visual signal;

3) When the distance between a local extremum and the global extremum is more than 50 units and less than 100 units of measurement of the values stored in the array then every fifths value of the distance can be translated into a non-visual signal;

4) When the distance between a local extremum and the global extremum is more than 100 units of measurement of the values stored in the array then further increasing the distance can be accompanied with the same specific non-visual signal.

Thus, there is no need to use a wide range of non-visual signals to convert a large amount of data stored in the array. The concept of imaging the relative position of global extremum with a limited number of non-visual signals could further facilitate detection of the row where the global

extremum is located. Then, to make navigation along the row less dependent on the absence of visual feedback, by sending the command through an accessible input, the user can lock the row number and continue exploration of the row content.

By dragging the stylus along the bottom exploration area the pointer will follow within the row locked. At that again, the distance between the current column and the column where the global extremum is located can be indicated in a similar way as it was already discussed with respect to the local extremes in each row. Moreover, any local or linked information can be read aloud by the screen reader.

3.2 Retrieving the Position of Any Value in the Data Array

Processing the virtual array remains a possibility of mutually inverse operations with respect to the user task. Let us consider an array of 100 elements of integer values varying from 0 to 99. The elements of the array can be presented in the table in a random order. In the previous subsection 3.1 the algorithm of retrieving the position of extremes of the array was discussed in detail. In more general case, any element of the array can be assigned as the target having the minimum value equal to "0". Then, values of other elements of the array should be changed accordingly. That is, the value of the i-element (e.g., having the value 78) could be turned into the minimum and be equal to "0" by shifting all other numerical values accordingly. However, some elements, which had values less than the target, can get a negative value. The use of the distance of i-element from the target helps to convert the numerical values into non-visual signals regardless of their sign. This also decreases the number of non-visual signals required for translation. Thus, the task of retrieving the position of the i-element having the numerical value 78 can be reduced to the task to detect the position of the minimum. The algorithm provides quick search for numeric values in the data array of any length presented in a tabular form in the absence of visual feedback.

Moreover, non-visual imaging of intermediate data that helps to navigate within the array would require less than 50 out of 128 notes of the MIDI Synthesizer or about 40 clearly distinct states (levels of the penholder displacements) of the StickGrip kinesthetic display (see detailed description in subsection 3.4) to solve the task. Such an approach would be entirely consistent with the Human Factors Guidelines on the use of non-speech acoustic signals and auditory tones [14].

3.3 Participants

Three totally blind persons successfully tested the technique, but these data were excluded from analysis. For ethical reasons, we could not involve blind people in testing when many factors were unknown. That is, only sighted people were able to evaluate the benefits and shortcomings of the new technique introduced. Eleven volunteers from the local university participated in this study. All had normal hearing abilities (but we did not make audiometric tests) and none of the participants had any extensive training in music. They did not experience issues related to skin sensitivity or movement disorders. Five of the subjects were males and six were females. The age of the subjects ranged from 29 to 58 years with a mean of 35 years. All participants were regular computer users and reported being right hand dominant or using their right hand most frequently. None of them had participated in similar experiments before.

3.4 Apparatus and Equipment

A desktop PC with Intel Core Quad CPU Q9450 (2.66 GHz, 3.25GB of RAM) was used in the present experiment equipped with the PCI sound card Creative Sound Blaster Live! 5.1 and active stereo speaker system KOSS SA/30. The Wacom Graphire-4 USB tablet having an interactive surface of 280×260 mm2 and including a cordless stylus was employed as an input technique. In addition, Wobbrock with co-authors have proven [26] that physical edges can be efficiently used to provide greater stability and higher accuracy in the use of stylus input techniques in the absence of visual feedback. To limit extra movements of the Wacom pen out of the workspace we used the rectangular frame of 145×205×1 mm^3.

To provide the participants with auxiliary haptic feedback the motorized penholder for the Wacom pen (the StickGrip) was used [27]. The device (Figure 3) provides a range of 40 mm (±20 mm) of the vertical displacement of the penholder (the point of grasp) with an accuracy of (±0.8 mm) and with an average speed of 30 mm/s for the Wacom pen having a length of 140 mm.
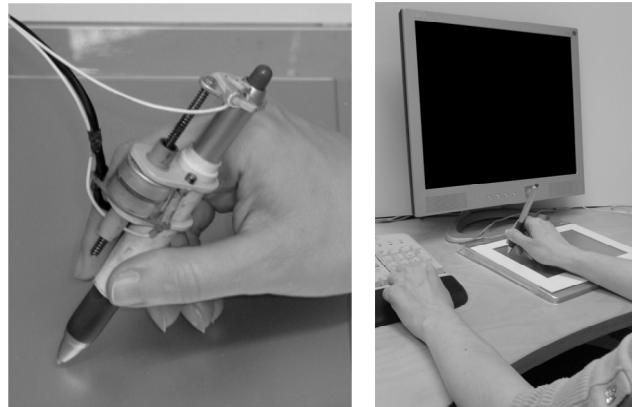

FIGURE 3: The StickGrip kinesthetic display (on the left) and experimental setup (on the right).

Thus, when the participants explored the data distributed within the workspace, they could feel as a point of grasp is sliding up and down the shaft of the Wacom pen and their hand being displaced towards and away from the physical surface of the pen tablet. Distance and direction of the penholder displacements were coordinated with a distance to target – the results of the data processing by the application-mediator (see subsections 3.1, 3.2). The software was developed in the Microsoft Visual Basic environment under Windows XP. An embodied Microsoft Speech API was used to announce short verbal messages or any variable on demand (by clicking the arrow key down of the keyboard).

The easily distinguishable MIDI sounds of the patch 21 (Accordion instrument) were generated via the JWCS MIDI Synthesizer Control [8], [28]. Table 1 and Table 2 demonstrate how the distances to the target were encoded with a minimum number of MIDI notes to detect the position of the global extremum values. We used an original sound without any additional effects and modification by equalizer.

DISTANCE (D)	MIDI NOTE (D)	Note number // name // Frequency in Hz
0 to 19	60 + d	60, 61, 62…79 // C5, C#5…G6 // 262, 277…784
20 to 49	75 + Integer(d/3)	81, 82, 83…91 // A6, A#6…G7 // 880, 932…1568
50 to 99	83 + Integer(d/5)	93, 94, 95…102 // A7, A#7…F#8 // 1760, 1865…2960

TABLE 1: MIDI Notes (0-127) Used to Indicate the Distance to a Minimum Value in the Virtual Array.

Distance (d)	MIDI Note (d)	Note number // name // Frequency in Hz
99 to 80	82 + (d-80)	102, 101…82 // F#8, F8…A#6 // 2960, 2793…932
79 to 50	54 + Integer(d/3)	80, 79, 78…70 // G#6, G6…A#5 // 830, 784…466
49 to 0	60 + Integer(d/5)	69, 68, 67…60 // A5, G#5… C5 // 440, 415…262

TABLE 2: MIDI Notes (0-127) Used to Indicate the Distance to a Maximum Value in the Virtual Array.

3.5 Experimental Procedure
The experiment was conducted in the usability laboratory at the local university. The subjects were tested individually. Prior to data collection, the experimenter briefly introduced the

techniques and the goal. The detailed instructions were given for the use of kinesthetic display, the task and overall testing procedure. The formal consent of the participants to collect and use the data obtained were gathered. No private or confidential information was collected and stored.

Still the participants were given an opportunity to refuse the continuation of the experiment at any point without any explanation of the reason and they were told that such a refusal would not have any further negative effect to them or to the experiment.

The subjects were advised to anchor the hand on the desk where the Wacom pen tablet was located. They were asked to leave the hand relaxed while grasping the pen or the StickGrip. The participants could perceive MIDI sounds, speech and haptic feedback cues when they explored the numerical data relying on hearing, kinesthetic and proprioceptive senses. No visual feedback was provided during the test. The monitor was switched off during the whole experiment, however the participants were asked to keep their eyes closed in order to better focus on the feedback signals.

In the beginning of each session, the subjects were allowed to practice shortly with the sequence of actions in the absence of visual feedback. The activity of the subjects was monitored on the computer screen of the experimenter. The participants could ask questions if they felt that the testing procedure required more clarification.

3.5.1 The First Experimental Task

The first task consisted in detection the position of the global extremum value in the virtual table. The experimenter did set the appropriate mode of the application according to the task. To accomplish this task, the participants used Wacom tablet bordered with a rectangular frame, as shown in Figure 4. The preferred positions of the keyboard and the tablet were adjusted beforehand.

The subjects started the trials by clicking the spacebar on the keyboard. To find out the location of the virtual cell containing the extremum (minimum or maximum value) the participants had to move the Wacom pen (Figure 3) or the StickGrip up and down along the borders of the rectangular frame, within the left or the right exploration areas (Figure 3). When the Y-coordinate of the pointer location was within the row of the virtual table, the application made analysis of the row data and generated the signal, which designated the "relative distance" between the local extremum value and the target value (the global extremum).

The participant had to perform an auditory or haptic perceptual analysis of only 10 non-visual signals and make a decision in which row the global extremum is located. The minimum value of the distance was translated into the lowest sound pitch or position of the penholder. The highest sound pitch and displacement of the penholder designated the maximum distance to the target value. After detection the row in question where the exact position of the global extremum value might be located, the subject had to press the "0" (zero) key on the numeric keypad to lock this row. After the first step, the subjects had to continue an exploration of the global extremum position within the row. The participants had to move the Wacom pen or the StickGrip along the bottom exploration area (Figure 3) of the frame. They should press the 0-key again when the target value was located. When the virtual cell (extremum value) was incorrectly pointed out, the person heard the error sound to stimulate him/her to be more attentive toward the analysis of the non-visual signals. Manipulating with the spacebar and 0-key of numeric keypad did not cause any problems in the absence of visual feedback. Nevertheless, there was a possibility to assign any other two keys on demand.

The participants had to complete two blocks of 20 trials each (10 tasks to detect the minimum and 10 tasks – the maximum value), under two conditions of feedback with no time limit. In one block, a pen input was accompanied with auditory signals, in another block the StickGrip was used as a kinesthetic display of the relative distance to target.

The numerical data were placed in the virtual table in a random order but the conditions, type and location of extremum were presented in a counterbalanced order. The type of extremum in question was announced as a verbal message ("minimum" or "maximum"). The end of the test was accompanied with the applause sound. Between trials, the participants were allowed to have a short break to ask questions and give comments.

A session of two blocks lasted for an hour and the test was performed 3 times per week, during 2 weeks, according to the ethical standards. Thus, each of eleven subjects performed 40 trials per day and 240 trials in a total. The data recorded were stored in a log file for further analysis. With a break of one week, the same group of participants performed the second task.

3.5.2 The Second Experimental Task

The second experimental task was to locate any value in the virtual table (data array). Described in subsection 3.2, the algorithm was used to support non-visual interaction with numerical data. The elements of the array for each task were transposed automatically when the value in question was selected and announced verbally in synthesized speech. The value, which has to be detected, was designated as a global minimum. The value of i-element (Val_i) in the array after transposition was converted and presented as follows:

$$Val_i = Abs\ (Val_i - Val_{target}) + 1 \tag{1}$$

where Val_{target} was the announced value in question.

After listening to the task, the participants followed the procedure similar to the first experimental task (see subsection 3.5.1). The protocol of the testing procedure was also similar and comprised of 5 sessions completed during one week with 40 trials per day (2 blocks). All cells of the virtual table (100 values of the array) were presented in a counterbalanced order.

4. RESULTS

Within a bordered area of the tablet, the subjects demonstrated a similar navigation activity required by the task and the techniques used. At the beginning of the test, the subjects needed some practice to comprehend the way of interaction with tabular data in the absence of visual feedback to learn how to manage with the Wacom pen or the StickGrip device and how to interpret non-visual signals indicating the relative distances to the target value instead of the true values. They moved the input device very carefully and slowly back and forth along the left or right edges of the frame in order to inspect the sequences of sounds or the penholder displacements. The subjects started the test from upper row. After a series of repeated observations and a thorough perceptual analysis, they confirmed their choice according to the test protocol. With the experience gained during the first session, the subjects felt more confident in the absence of visual feedback relying on auxiliary information and they performed the task faster already during the second session. At that point, the heterogeneity in decision-making has increased and the subjects have committed a greater number of errors.

At the beginning of the test, 4 of 11 subjects tried to detect the extremum in the first (upper) or the last (bottom) two rows. They were expecting to directly find out the extremum value encoded by sound pitch, while they were told that non-visual signals indicate the relative distance to the target cell and not the true value. Still, they immediately confirmed their choice without making thorough inspection of all the rows to be sure that the decision was made correctly. On the contrary, (based on the subject's reports) the penholder displacements seemed "more obvious", "natural", "intuitive" and "realistic" non-visual indicator of the physical distance.

The mean times spent by the subjects during the first week to detect the position of the global extremum in the virtual table using the Wacom pen and sound signals (Condition 1) are presented in Figure 4. The results revealed that 3 of 11 participants had difficulties to identify the minimum value than the maximum value. This would be possible due to the fact, that we

intentionally did not apply any adjustment of the audio system. The participants were the computer users and non-musicians and could have different hearing sensitivity at different frequencies (pitch and volume). However, by retrieving the position of the minimum values the subjects made a thorough inspection with fewer errors.
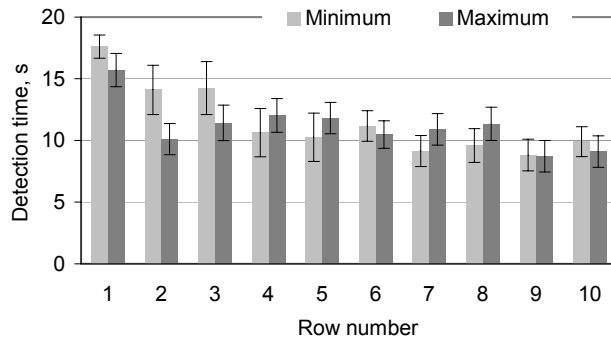

FIGURE 4: The mean time spent by the subjects to detect the global extremum (target) value in the virtual table using the Wacom pen and auditory signals. The data of the 1st week of testing were averaged over 33 trials in each row.

The mean task completion time of detection the position of global minimum was about 11.5 s with a standard deviation (SD) of about 2.9 s, varying from 9 to 17 s. To find the position of global maximum they spent in average 11.2 s (SD=1.9 s). This time varied from 8 to 15 s. The paired samples t-test revealed no statistically significant difference in this task under Condition 1: $t(9) = 0.592$ ($p>0.5$), the correlation index of these data was 0.694 ($p>0.01$).

The subjects had to analyze all the local extrema and consequently all non-visual signals indicating their distances to the target (global extremum) over 10 rows. Therefore, it took more time to identify the cell position when it was located in the upper row than in the bottom row due to differences in the travel distance towards the bottom row and backwards to the target. If the global extremum value was located on the way towards the bottom row, it took less time to find the position of extremum. The similar situation was observed independently of the signals modality.

FIGURE 5: The mean time spent by the subjects to detect the global extremum value in the virtual table using the Wacom pen and auditory signals. The data of the 2nd week of testing were averaged over 33 trials in each row.

During the 2nd week of testing, the mean task completion time of detection the position of global minimum was about 9.3 s (SD=0.65 s), varying from 8 to 11 s (see Figure 5), while to find the position of global maximum the subjects spent in average 10.4 s (SD=1.9 s). This time varied from 8 to 14 s. The paired samples t-test revealed a statistically significant difference in this task under Condition 1: $t(9) = 4.4$ ($p<0.005$), the correlation index was high and significant 0.85 ($p<0.005$).

The mean times spent during the 1st and 2nd weeks to detect the position of global extremum using the StickGrip (Condition 2) are presented in Figure 6 and Figure 7.

FIGURE 6: The mean time spent by the subjects to detect the global extremum value in the virtual table using the StickGrip device. The data of the 1st week of testing were averaged over 33 trials in each row.

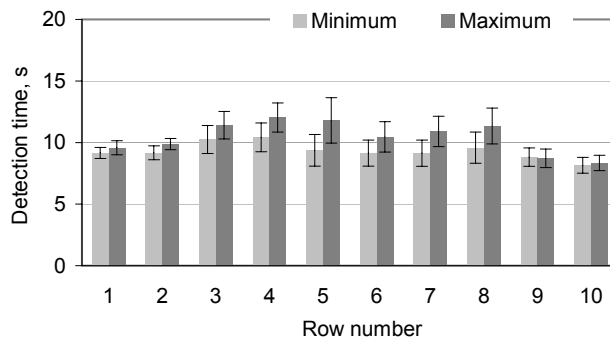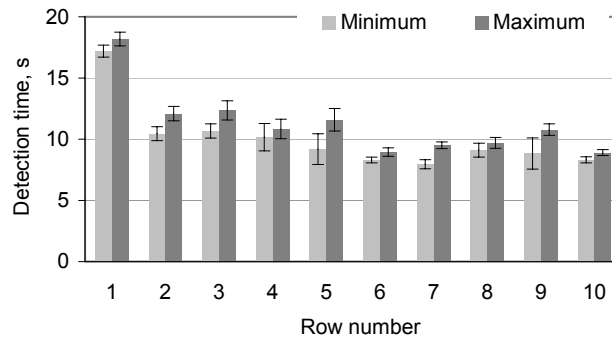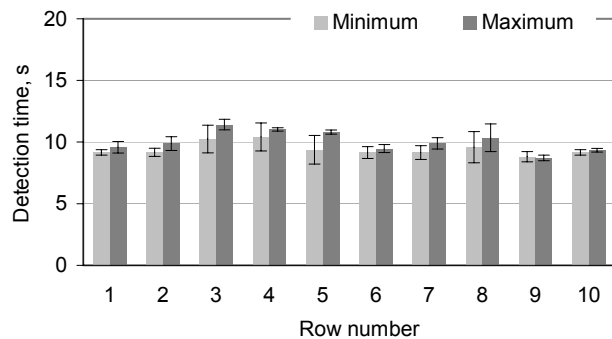**FIGURE 7:** The mean time spent by the subjects to detect the global extremum value in the virtual table using the StickGrip device. The data of the 2nd week of testing were averaged over 33 trials in each row.

At the beginning of the test (Figure 6), the subjects also spent more time to detect a position of the global maximum with an average value about 11 s (SD=2.7 s) varying from 8 to 18 s. Detection the global minimum required of about 10 s (SD=2.7s), varying from 7.5 to 17 s. As regards the task completion time the results of the paired samples t-test revealed a statistically significant difference in this task under Condition 2: $t(9) = 6.178$ ($p<0.001$), the correlation index was high and significant 0.97 ($p<0.001$).

Still a practice with the StickGrip took a short time to learn how to distinguish and interpret haptic signals arising from vertical displacements of the penholder. To the end of the test, the subjects mastered their skills and it was required less time to find a location of extremum. To detect the global minimum the subjects spent on average about 9.4 s (SD=0.5 s) and the global maximum of about 10 s (SD=0.8 s). The paired samples t-test also revealed a statistically significant difference in this task under Condition 2: $t(9) = 4.4$ ($p<0.005$), the correlation index was high and significant 0.885 ($p<0.005$).

The ANOVA analysis strengthened these findings and has shown that the novice-to-expert transition was significantly above chance when the participants was asked to detect the position of the global extremum in the virtual table with two different techniques. The data from the 1st and 6th test sessions were statistically assessed and compared (Table 3).

Extremum	Condition 1	Condition 2
Maximum	F = 2.36; *p < 0.001*	F = 4.15; *p < 0.001*
Minimum	F = 1.63; *p < 0.001*	F = 3.75; *p < 0.001*

TABLE 3: ANOVA results of detection the position of extremum values with two techniques in the 1st and 6th test sessions (10 trials, 11 participants).

As it can be seen from Figure 8, the subjects demonstrated a tendency to commit a greater number of errors when they were asked to detect the position of the global maximum under Condition 1. In particular, the average number of errors recorded from 11 subjects per session (110 trials) was about 7.4 (SD=4.3) for maximum and only 1.5 (SD=0.8) for the global minimum. The paired samples t-test revealed a statistically significant difference in this task under Condition 1: t(5) = 4.344 (p<0.005), still the correlation index was low and not significant -0.4 (p>0.1). The greater number of errors made in the beginning of the test was also influenced by misinterpretation of sounds indicating the distance to the target.

Relying on haptic signals (Condition 2), the subjects experienced similar difficulties to detect the position of the global extremum (Figure 9). They needed time to learn how to interpret the displacements of penholder with the distance to the target. However, when they were asked to detect the global minimum, the subjects again committed fewer mistakes.

The mean number of errors recorded from 11 subjects per session (110 trials) was about 4.7 (SD=3.6) when the subjects were asked to detect the position of the global minimum and about 9.1 (SD=4.6) when global maximum was asked to locate. The paired samples t-test revealed a statistically significant difference in this task under Condition 2: t(5) = 6.2 (p<0.001), a correlation was 0.855 and significant (p<0.001).

FIGURE 8: The average number of errors committed by 11 participants (110 trials per session) during detection of the global extremum value in the virtual table relying on the Wacom pen and auditory signals throughout six test sessions.

Figure 10 demonstrates the variations in the mean times spent by participants to detect the position of the numerical values (the second task) under two conditions. The data in each row were averaged over 110 trials.

In general, the participants spent a little less time to detect the position of numerical value being converted into the global minimum when relied on the StickGrip device: 14.3 s (SD=3.9s) vs. 15.6 s (SD=3.2 s). The paired samples t-test revealed no statistically significant difference in this task: t(9)=1.296 (p>0.1), the correlation index was 0.646 (p>0.01). This result was confirmed by the ANOVA p<0.001, F = 0.97.
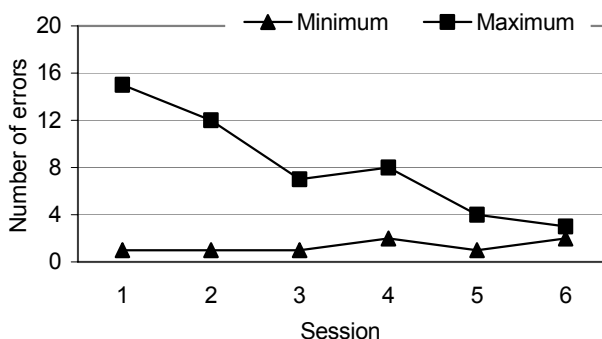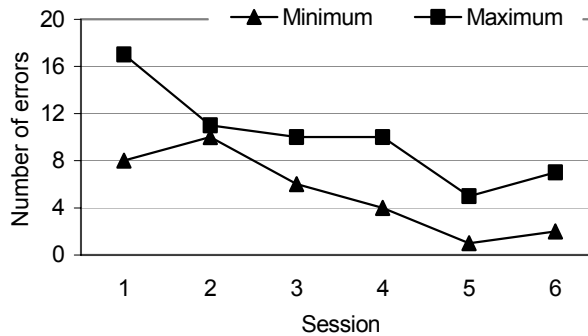
FIGURE 9: The average number of errors committed by 11 participants (110 trials per session) during detection of the global extremum value in the virtual table relying on the StickGrip kinesthetic display throughout six test sessions.
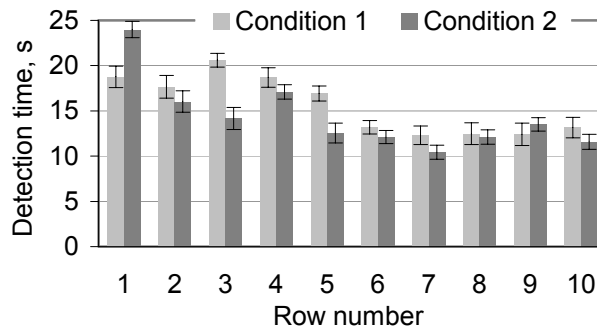

FIGURE 10: The mean time spent to detect a position of any value of 100 in the virtual table using the Wacom pen and auditory signals (Condition 1), and the StickGrip kinesthetic display (Condition 2). The data in each row were averaged over 110 trials.

Even with getting a practice in non-visual inspection of the virtual table during the first task the subjects spent greater time to detect the position of the cell when it was situated in upper part (rows 1-5) rather than in bottom rows (6-10). The mean of detection time in rows 1-5 was about 18.5 s (SD=2.4 s) vs. 17s (SD=4.5 s) varying from 13 to 23 s and from 9.6 to 26 s under Condition 1 and 2 accordingly. When the value in question was located in bottom rows 6-10 an average time was about 12.7 s (SD=2.1s) vs. 11.9s (SD=1.7s) varying from 8 to 17s and from 7.5 to 16 s under Condition 1 and 2 accordingly. As it was already discussed, it took more time due to differences in the travel distance towards the bottom row and back to the target position.

5. CONCLUSION

Different authors have tried to find a solution to improve accessibility to the data presented in a tabular form. However, both the data used, techniques, experimental conditions and methods of testing and analysis of the results reported in the literature vary widely. The available experimental results, which can still be compared to our approach, are presented in the Table 4.

We have implemented, described and assessed the technique and algorithm of retrieving the position of any value in the data records which are appropriate for non-visual navigation with numerical data when the number of elements in the array is large enough (typically 1000 or more). Moreover, the navigation within the virtual table was performed using the touch sensitive surface and pen-based input techniques. The number of non-visual cues required to support navigation is almost independent on the amount of data and was less than 50 signals, including the short verbal messages.

Still the goal of this study was to examine human performance and an accuracy of solving the most common tasks dealt with the virtual table composed of 100 cells. The tasks were retrieving the position of extreme values and the target value that can be linked to the specific content, as it was discussed in the scenario of access to the security number in the card (table) of online banking system.

Data format Techniques [Ref.]	Elements in array	Tasks assigned included: identification of	Auxiliary modality Auditory		Auxiliary modality Haptic	
			Time, s (SD)	Accuracy %	Time, s (SD)	Accuracy %
Tabular MIDI vs. StickGrip [new]	100	- the maximum value	10.8(1.9)	92.6	9.7(1.8)	90.9
	100	- the minimum value	10.4(1.6)	98.5	10.5(1.6)	95.3
	100	- the specific value of 100 integers	15.6(3.2)	100	14.3(3.9)	100
Tabular Speech vs. MIDI [11]	260	- the highest number of 100 integers (speech)	133(55)	4.6	--	--
	260	- the highest number of 100 integers (127 notes)	44(14)	5.3	--	--
Tabular MIDI [17]	28, 168, 744	- the highest/lowest value (66 notes)	*22(10) *26(10) *32(10)	*87(10) *90(10) *80(10)	--	--
Tabular Speech vs. Braille [6]	49	- the maximum value (browsing with speech)	42	88	--	--
	70	- the minimum value (browsing with a formula)	30	75	--	--
	21	- the specific cell (browsing with Braille)	*15	--	18	20
Graph Tactile paper vs. Novint Falcon [5]	11 (points)	- the max / min point (tactile paper)	--	--	23	42
		- the max / min point (Novint Falcon)	--	--	15	100

* The data were not presented exactly in the paper; these values were extracted from the figures.

TABLE 4: Comparison of the new approach with existing methods.

The assessment of human performance revealed that the subjects committed the greater number of errors in the beginning of the test and relying on haptic signals rather than when they used the Wacom pen and auditory signals. Nevertheless, an experience with the StickGrip device took a short time to learn how to distinguish and interpret haptic signals arising from vertical displacements of the penholder.

The time required to detect the position of the virtual cell containing the extremum announced verbally was about 10.6 s (SD=1.75 s) when they used the Wacom pen and auditory signals vs. 10.1 s (SD=1.7 s) when the subjects relied on the StickGrip device. During the second task of detection any value announced verbally, the task completion time was about 15.6 s (SD=3.2 s) when the subjects relied on the Wacom pen and auditory signals vs. 14.3 s (SD=3.9 s) when they used the StickGrip device. However, the variation of the task completion time might have resulted from differences in the travel distances to the target position and the limited speed of penholder displacements of the kinesthetic display. Based on the subjective evaluation reports, we can

conclude that the StickGrip device has a potential as intuitive kinesthetic display to substitute auditory feedback in non-visual interaction with numerical data. The penholder displacements seemed to the subjects as "more obvious", "natural", "intuitive" and "realistic" non-visual indicator of the physical distance.

The algorithm allowing non-visual navigation in the virtual table on a touch sensitive surface and access to any value without extensive learning the numerical data converted into tactile and auditory patterns constitute a novel approach to make access and manipulating the tabular data in the absence of visual feedback.

6. FUTURE RESEARCH

Exploration of the large data set cannot be efficient when the cell content (e.g., numerical values) being converted into non-visual signals of any accessible modality would become redundant. Even detail overview of pieces of data does not allow to retrieve the data with a minimum number of steps. Rather than to spend additional efforts in understanding the encoding semantics of the tabular layout, the data translation methods should reduce their redundancy to a reasonable level while retaining some features of the spatial arrangement and a possibility to access to key variables indicating the trend of the parameter and descriptive statistics. The alternative approach might consist in the formalization of the human-computer dialog when the special application and the input technique could mediate the human-data interaction with respect to the user task. Then, only a limited number of navigational signals might be enough to find out a specific position in the virtual table and to display intermediate information if necessary during the data processing.

For visually impaired persons and in extreme situations, haptic channel is the most natural and perhaps the most efficient way of presenting a large amount of data (both discrete and continuous) in a limited amount of space [27], [29]. The StickGrip kinesthetic display was used in this study to provide haptic signals to support an intuitive interaction within the numerical data array: by indicating the relative distances (closer and further) between the position of the pointer and the target numerical value, providing an easy access to intermediate information and navigation in the data set. There are also many other tasks related to the interactive data visualizations.

In further research we also will focus on the issues allowing the person in the absence of visual information to interact with the data by perceiving, exploring, and understanding the data values and their relationships. The development of intuitive data visualization and exploration tools making human-data interaction comfortable and effective in the total absence of visual support remains one of the main challenges in the field of data visualization and related areas of research in human-computer interaction.

7. ACKNOWLEDGMENT

8. REFERENCES

[1] G. Douglas, C. Corcoran, S.A. Pavey "The role of the WHO ICF as a framework to interpret barriers and to inclusion: visually impaired people's views and experiences of personal computers," Br J of Vis Impair, 25(1) 32–50, Jan. 2007, DOI: 10.1177/0264619607071773.

[2] D. Burton, J. Lilly "NonVisual Desktop Access and Thunder: A Comparison of Two Free Screen Readers," Access World 12(8) Aug. 2011.

[3] Y. Wang, I.T. Phillips, R.M. Haralick "Table structure understanding and its performance evaluation," Pattern Recognition, Vol. 37, No 7 Jul. 2004, pp. 1479-1497.

[4] M. Chiousemoglou, H. Jürgensen "Setting the table for the blind," in Proc. of PETRA, Int. Conf. on Pervasive Technologies Related to Assistive Environments, Crete, Greece, ACM Press, New York, 2011, doi>10.1145/2141622.2141624.

[5] I.A. Doush, E. Pontelli "Non-visual navigation of spreadsheets," UAIS, Mar. 2012, pp. 1-17, http://dx.doi.org/10.1007/s10209-012-0272-1.

[6] I.A. Doush, E. Pontelli "Non-visual navigation of tables in spreadsheets," in Proc. of ICCHP 2010, Vienna, Austria, Part I, LNCS 6179, Springer, Berlin, 2010, pp. 108–115.

[7] C. Jay, R. Stevens, R. Hubbold, M. Glencross "Using haptic cues to aid nonvisual structure recognition," ACM Trans Appl Percept. 5(2), Article 8, May 2008, 14 p.

[8] T.V. Evreinova, G. Evreinov, R. Raisamo "Non-visual Gameplay: Making Board Games Easy and Fun," in ICCHP 2008, Linz, Austria, Springer, Berlin Heidelberg, LNCS 5105, 2008, pp. 561-568.

[9] L.M. Brown, S.A. Brewster, R. Ramloll, W. Yu, B. Riedel "Browsing Modes For Exploring Sonified Line Graphs," in British HCI Conf, London, UK, vol. 2, 2002, pp. 2-5.

[10] R. Ramloll, S.A. Brewster "Generic approach for augmenting tactile diagrams with spatial non-speech sounds," in CHI Ext. Abstracts on Human factors in computing systems, ACM Press, NY, USA, 2002, pp. 770–771.

[11] R. Ramloll, S. Brewster, W. Yu, B. Riedel "Using Non-speech Sounds to Improve Access to 2D Tabular Numerical Information for Visually Impaired Users," in BCS IHM-HCI, Lille, France, Springer, Berlin, 2001, pp. 515–530.

[12] S. Smith, R.D. Bergeron, G.G. Grinstein "Stereophonic and Surface Sound Generation for Exploratory Data Analysis," in CHI SIGCHI Conf on Human factors in comp systems, ACM Press, NY, USA, 1990, pp. 125-132.

[13] B.N. Walker, M.A. Nees "Theory of sonification," in Hermann T, Hunt A, Neuhoff J.G., (eds), The Sonification Handbook, Logos Publishing House, Berlin, Germany, ch. 2, 2011, pp. 9-39.

[14] B.N. Walker, J.T. Cothran "Sonification Sandbox: a graphical toolkit for auditory graphs," in Proc. of ICAD Int. Conf. on Auditory Displays, 2003, pp. 161-163.

[15] L.M. Brown, S.A. Brewster, R. Ramloll, M. Burton, B. Riedel "Design Guidelines For Audio Presentation of Graphs and Tables," in Proc. of ICAD Int. Conf on Auditory Displays, Boston, MA, USA, 2003, pp. 284-297.

[16] J. Kildal, S.A. Brewster "Explore the matrix: Browsing numerical data tables using sound," in ICAD Proc. of Int. Conf. on Auditory Displays, Limerick, Ireland, 2005, pp 300-303.

[17] J. Kildal, S.A. Brewster "Providing a size-independent overview of non-visual tables," in Proc. of ICAD Int. Conf. on Auditory Displays, London, 2006, pp. 8-15.

[18] T. Stockman "The Design and Evaluation of Auditory Access to Spreadsheets," in Proc. of ICAD Int. Conf. on Auditory Displays, Sydney, Australia, 2004, at: http://icad.org/Proceedings/2004/Stockman2004.pdf [accessed on Mar. 2013].

[19] T. Stockman, C. Frauenberger, G. Hind "Interactive Sonification of Spreadsheets," in Proc. of ICAD Int. Conf. on Auditory Displays, Limerick, 2005, pp. 134–139.

[20] "Human Factors Guidelines for ICT products and services," Design for All, ETSI EG 202116, V1.2.1 (2002-09), p. 146, available at: http://www.cettico.fi.upm.es/aenor/etsi202116v010201p.pdf [accessed on Mar. 2013].

[21] S. Wall, S. Brewster "Providing External Memory Aids in Haptic Visualisations for Blind Computer Users," Int. J. of Disability and Human Dev, Vol. 4, No 4, May 2005, pp. 331-338.

[22] I.A. Doush, E. Pontelli, D. Simon, T.C. Son, O. Ma "Making Microsoft ExcelTM: multimodal presentation of charts," in Proc. of ACM SIGACCESS Conf. on Computers and Accessibility, ACM Press, NY, USA, 2009, pp. 147-154.

[23] C. Colwell, H. Petrie, D. Kornbrot, D. Kornbrot "Haptic virtual reality for blind computer users," in Proc. of ASSETS Int. ACM Conf on Assistive technologies, ACM Press, NY, USA, pp. 92–99, 1998.

[24] C. Sjöström "Designing haptic computer interfaces for blind people," in Proc. of ISSPA Int. Symp. on Signal Processing and its Applications, Workshop on Website Evolution. Kuala Lumpur, Malaysia, IEEE CERTEC, Lund, Vol 1, pp. 68-71, 2001.

[25] Excel Magic Trick 731: "Retrieve Cell Address of Minimum Value in Column," at: http://www.youtube.com/watch?v=cpRlYaaztx4 [accessed on Mar. 2013].

[26] J.O. Wobbrock, B.A. Myers, J.A. Kembel "EdgeWrite: A stylus-based text entry method designed for high accuracy and stability of motion," in Proc. of UIST the ACM Symp. on User Interface Soft and Tech, Vancouver, British Columbia, Canada, ACM Press, NY, USA, 2003, pp. 61-70.

[27] T.V. Evreinova, G. Evreinov, R. Raisamo "An Evaluation of the Virtual Curvature with the StickGrip Haptic Device: a Case Study," UAIS 12(2) June 2013, pp. 161-173. DOI:10.1007/s10209-012-0273-0.

[28] T.G. Evreinova, G. Evreinov, R. Raisamo "Camera-Based Target Acquisition Augmented with Phosphene Sensations," in ICCHP 2010, Vienna, Austria, Springer, Berlin Heidelberg, LNCS 6180, 2010, pp. 282-289.

[29] A. Israr, P.H. Meckl, Ch.M. Reed, H.Z. Tan "Controller design and consonantal contrast coding using a multi-finger tactual display," J. Acoust. Soc. Am. 125(6), June 2009, pp. 3925-3935.

Generating a Domain Specific Inspection Evaluation Method through an Adaptive Framework: A Comparative Study on Educational Websites

Roobaea S. AlRoobaea   *r.alrobaea@uea.ac.uk*
Faculty of Computer Science and
Information Technology Taif
University, Saudi Arabia and
School of Computing Sciences,
UEA, Norwich, UK

Ali H. Al-Badi   *aalbadi@squ.edu.om*
Department of Information
Systems, Sultan Qaboos
University, Oman

Pam J. Mayhew   *p.mayhew@uea.ac.uk*
School of Computing Sciences,
UEA, Norwich, UK

Abstract

The growth of the Internet and related technologies has enabled the development of a new breed of dynamic websites and applications that are growing rapidly in use and that have had a great impact on many businesses. These websites need to be continuously evaluated and monitored to measure their efficiency and effectiveness, to assess user satisfaction, and ultimately to improve their quality. Nearly all the studies have used Heuristic Evaluation (HE) and User Testing (UT) methodologies, which have become the accepted methods for the usability evaluation of User Interface Design (UID); however, the former is general, and unlikely to encompass all usability attributes for all website domains. The latter is expensive, time consuming and misses consistency problems. To address this need, new evaluation method is developed using traditional evaluations (HE and UT) in novel ways.

The lack of a methodological framework that can be used to generate a domain-specific evaluation method, which can then be used to improve the usability assessment process for a product in any chosen domain, represents a missing area in usability testing. This paper proposes an adapting framework and evaluates it by generating an evaluation method for assessing and improving the usability of a product, called Domain Specific Inspection (DSI), and then analysing it empirically by applying it on the educational domain. Our experiments show that the adaptive framework is able to build a formative and summative evaluation method that provides optimal results with regard to the identification of comprehensive usability problem areas and relevant usability evaluation method (UEM) metrics, with minimum input in terms of the cost and time usually spent on employing UEMs.

Keywords: Heuristic Evaluation (HE), User Testing (UT), Domain Specific Inspection (DSI), Adaptive Framework, Educational Domain.

1. INTRODUCTION

The growth of the Internet and new technologies has created new dynamic websites that are growing rapidly in use and are having a significant impact on many businesses. These dynamic websites are becoming increasingly developed in the midst of the Internet revolution and ever-improving information technologies. For example, e-learning websites are now essential for all universities that have a physical workplace. They have websites, which have become an integrated part of their business, particular in their e-learning systems. Nowadays, the Internet revolution has led to a large number of universities being solely online, without needing a physical workplace. To keep pace with this development, some companies and organizations seek to build free online learning websites that are oriented to world-class education for all educational levels, such Intel® Education and the BBC. This development in lifelong learning has made learners' intention to continue using e-learning an increasingly critical issue. Consequently, quality is considered crucial to education in general, and to e-learning in particular. Web design is a key factor in determining the success of e-learning websites, and users should be the priority in the designers' eyes because usability problems in educational websites can have serious ramifications, over and above the users failing to meet their needs.

It is clear that Heuristic Evaluation (HE) and User Testing (UT) are the most important traditional usability evaluation methods for ensuring system quality and usability [10;13]. Currently, complex computer systems, mobile devices and their applications have made usability evaluation methods even more important; however, usability differs from one product to another depending on product characteristics. It is clear that users have become the most important factor impacting on the success of a product; if a product is produced and is then deemed not useful by the end-users, it is a failed product (nobody can use it and the company cannot make money) [42]. Nayebi et al., (2012) asserted, "companies are endeavouring to understand both user and product, by investigating the interactions between them" [44].

Traditional usability measures of effectiveness, efficiency and satisfaction are not adequate for the new contexts of use [52]. HE has been claimed to be too general and too vague for evaluating new products and domains with different goals; HE can produce a large number of false positives, and it is unlikely to encompass all the usability attributes of user experience and design in modern interactive systems [21; 11]. UT has been claimed to be costly, time consuming, prone to missing consistency problems and subject to environmental factors [44]. Several studies have also emphasised the importance of developing UEMs as a matter of priority, in order to increase their effectiveness. To address these challenges, many frameworks and models have been published to update usability evaluation methods (UEMs) [3; 20]; however, these frameworks and models are not applicable to all domains because they were developed to deal with certain aspects of usability in certain areas [16].

The main objective of this paper is to address these challenges and to construct a methodological framework and then to test its validity by applying it on the educational domain, through three case studies. Furthermore, it is to conduct a comprehensive comparison between UT, HE and our domain specific investigation (DSI) method, which is from the adaptive framework, in terms of discovering the number of real usability problems and their severity in each of the usability problem areas, UEM metrics, and other measurements. The paper is organized in the following way. Section 2 starts with a brief literature to this study and includes a definition of usability problems, a severity rating and related work. Section 3 describes the construction of the adaptive framework. Section 4 is details the research methodology. Section 5 details the set of measurements and analysis metrics. Section 6 validates the adaptive framework by applying the new method (DSI), HE and UT in practice to three cases and provides an analysis and discussion of the results. Section 7 presents a discussion of the findings. Section 8 presents the conclusion and future work.

1.1 Research Hypotheses
This research hypothesizes that:

1. There are significant differences between the results of HE and DSI, where the latter method outperforms the former in terms of achieving higher ratings from evaluators on the issues relating to the number of usability problems, the usability problem areas, the UEM performance metrics, and the evaluators' confidence, concluding that it is not essential to conduct HE in conjunction with DSI.

2. There are significant differences between results of UT and DSI, where the latter method outperforms the former in terms of achieving higher ratings on the issues relating to the number of usability problems, the usability problems areas, the UEM performance metrics, concluding that it is not essential to conduct UT in conjunction with DSI.

2. LITERATURE REVIEW AND RELATED RESEARCH

A website is a product, and the quality of a product takes a significant amount of time and effort to develop. A high-quality product is one that provides all the main functions in a clear format, and that offers good accessibility and a simple layout to avoid users spending more time learning how to use it; these are the fundamentals of the 'usability' of a product. Poor product usability may have a negative impact on various aspects of the organization, and may not allow users to achieve their goals efficiently, effectively and with a sufficient degree of satisfaction [26]. The growth of the Internet has led to an explosion of educational website content, rising in accordance with demand. E-learning occurs when students in any place and time access the Internet to proceed through the sequence of teaching, completing the learning activities and achieving learning results and objectives. This could be part of a winning strategy for particular needs, such as decongestion of overcrowded education facilities, and support for students or teachers and adult education [1; 7]. However, some of these websites are difficult to use due to the inexperience of many of the designers and the lack of effective, efficient and accurate appropriate guidelines for performing this task. Consequently, users spend more time learning how to use the website than learning the educational content, causing frustration leading to abandonment of the site. Alkhattabi et al. 2010 state, "quality is considered a crucial issue for education in general, and for e-learning in particular" [4]. Thus there is a need for e-learning websites to be of sufficiently high quality. In this, it is extremely important to classify suitable criteria for addressing and assessing their quality [48].

The reviewed literature shows that the techniques for measuring the quality of user experience have been classified under the heading of ergonomics and ease-of-use, but more lately under the heading of usability [44]. This aims to ensure that the user interface is of sufficiently high quality. Usability is one of the most significant aspects affecting the quality of a website and its user experience. Nielsen (1994b) stated, "usability is associated with learnability, efficiency, memorability, errors and satisfaction" [38]. Muir et al. (2003) defined pedagogic usability as a branch of usability that "affects educational website design and development, particularly in the context of supported open and distance learning" [35]. Usability is not a single 'one-dimensional' property of a user interface. There are many usability attributes that should be taken into account and measured. Shackel and Richardson (1991) proposed four-dimensional attributes that influence the acceptance of a product, which are effectiveness, learnability, flexibility and attitude [46]. Nielsen (1994b) introduced five major attributes of usability based on a System Acceptability model [38], and they are as follows; 1) Easy to learn: a system should be easy to learn for the first time; 2) Efficient to use: the relationship between accuracy and time spent to perform a task; 3) Easy to remember: a user should be able to use the system after a period without spending time learning it again; 4) Few errors: the system should prevent users from making errors (this also addresses how easy it is to recover from errors); and 5) Subjectively pleasing: this addresses the user's feeling towards the system.

Usability evaluation methods (UEMs) are a set of techniques that are used to measure usability attributes. They can be divided into three categories: inspection, testing and inquiry. Heuristic Evaluation (HE) is one category of the inspection methods. It was developed by [34], and is guided by a set of general usability principles or 'heuristics' as shown Table 1. It can be defined as a process that requires a specific number of experts to use the heuristics in order to find usability problems in an interface in a short time and with little effort [33]. It can be used early in the development process, and may be used throughout the development process [40]. However, it is a subjective assessment and depends on the evaluator's experience, and can produce a large number of false positives that are not usability problems at all or can miss some real problems [23; 39; 11; 21].

Heuristic Evaluation
Visibility of system status
Match between system and the real world
User control and freedom
Consistency and standards
Error prevention
Recognition rather than recall
Flexibility and efficiency of use
Aesthetic and minimalist design
Helps users recognize, diagnose, and recover from errors
Help and documentation

TABLE 1: Heuristic Evaluation.

There are two kinds of expert evaluators. One is a 'single' evaluator, who can be defined as a person with general usability experience. The second is a 'double' evaluator who can be defined as a person with a usability background in a specific application area. Molich and Nielsen (1990) recommended from previous work on heuristic evaluation that between three and five single expert evaluators are necessary to find a reasonably high proportion of the usability problems (between 74% and 87%) [34]. For the double expert evaluators, it is sufficient to use between two and three evaluators to find most problems (between 81% and 90%). There is no specific procedure for performing HE. However, Nielsen [37] suggested a model procedure with four steps. Firstly, conducting a pre-evaluation coordination session (a.k.a training session) is very important. Before the expert evaluators evaluate the targeted website, they should take few minutes browsing the site to familiarize themselves with it. Also, they should take note of the actual time taken for familiarisation. If the domain is not familiar to the evaluators, the training session provides a good opportunity to present the domain. Also, it is recommended that in the training session, the evaluators evaluate a website using the heuristics in order to make sure that the principles are appropriate [14]. Secondly, conducting the actual evaluation, in which each evaluator is expected to take around 1 to 1.5 hours listing all the usability problems. However, the actual time taken for evaluation should always be noted. Next, there should be a debriefing session, which would be conducted primarily in a brainstorming mode and would focus on discussion of possible redesigns to address the major usability problems and general problematic aspects of the design. A debriefing is also a good opportunity for discussing the positive aspects of the design, since heuristic evaluation does not otherwise address this important issue. Finally, the results of the evaluations are collected into actual evaluation tables, and then combined into a single table after removing any redundant data. After the problems are combined, the evaluators should agree on the severity of each individual problem [37].

In the present context and in relation to HE, usability testing (also known as user testing), is another important evaluation method for ensuring system quality, in particular for websites. It needs participants to perform a set of tasks, usually in a laboratory. These tasks are performed without information or clues as to how to complete them, and with no help provided to the user during the test session. Also, the completion of these tasks is monitored and assessed by an

observer who records the usability problems encountered by the users. All the observed data, such as error numbers, time spent, success rate and user satisfaction, need to be recorded for analysis [38]. Dumas and Redish (1991) stressed that a fruitful usability testing session needs careful planning and attention to detail [17]. Accordingly, there is a general procedure for conducting user testing, thus: 1) Planning a usability test; 2) Selecting a representative sample and recruiting participants; 3) Preparing the test materials and actual test environment; 4) Conducting the usability test; 5) Debriefing the participants; 6) Analysing the data of the usability test; and 7) Reporting the results and making recommendations to improve the design and effectiveness of the system or product. The Think-Aloud technique (TA) is used with UT. There are three TA types, which are concurrent, retrospective and constructive interaction. The concurrent TA type is the most common; this involves participants verbalising their thoughts whilst performing tasks in order to evaluate an artefact. Retrospective TA is less frequently used; in this method, participants perform their tasks silently, and afterwards comment on their work on the basis of a recording of their performance. Constructive interaction is more commonly known as Co-Discovery Learning, where two participants work together in performing their tasks, verbalising their thoughts through interacting [51].

One important factor in usability testing is setting the tasks. Many researchers are aware that task design is an important factor in the design of adequate usability tests. The tasks designed for Web usability testing should be focused on the main functions of the system. The tasks should cover the following aspects: 1) Product page; 2) Category page; 3) Display of records; 4) Searching features; 5) Interactivity and participation features; and 6) Sorting and refining features [50]. Dumas and Redish (1999) suggested that the tasks could be selected from four different perspectives [17]. These are: 1) Tasks that are expected to detect usability problems; 2) Tasks that are based on the developer's experience; 3) Tasks that are designed for specific criteria; and 4) Tasks that are normally performed on the system. They also recommended that the tasks be short and clear, in the users' language, and based on the system's goals [17]. Alshamari and Mayhew (2008) found that task design can play a vital role in usability testing results, where it was shown that changing the design of the task can cause differences in the results [6].

The result of applying HE and UT is a list of usability problems [37]. These problems are classified into different groups to which a numeric scale is used to measure the severity of each problem. Firstly, this issue is not a usability problem at all. Secondly, this is a cosmetic problem that does not need to be fixed unless extra time is available on the project. Next, this issue is a minor usability problem; fixing this should be given low priority. Then, this is a major usability problem; it is important to fix this, so it should be given high priority. Finally, this issue is a usability catastrophe; it is imperative to fix this before the product can be released.

In the early years of computing, HE was widely applied in measuring the usability of Web interfaces and systems because it was the only such tool available. These heuristics have been revised for universal and commercial websites as HOMERUN heuristics [41]. Furthermore, [11;12] have proposed UEMs called HE-Plus and HE++, which are extensions to HE by adding what is called a "usability problem profile". However, some researchers have found that their tested websites failed in certain respects according to these extended or modified heuristics [49; 5]. On the other hand, many researchers then sought to compare and contrast the efficiency of HE with other methods such as UT. They found that HE discovered approximately three times more problems than UT. However, they reported that more severe problems were discovered through UT, compared with HE [30; 18; 27]. Lately, researchers' findings have been almost unanimous in one respect: HE is not readily applicable to many new domains with different goals and are too vague for evaluating new products such as web products because they were designed originally to evaluate screen-based products; they were also developed several years before the web was involved in user interface design [24; 25; 31]. Thus, each method seems to overcome the other method's limitations, and researchers now recommend conducting UT together with HE because each one is complementary to the other, and then combining the two methods to offer a better picture of a targeted website's level of usability [36; 32]. To address these challenges, many frameworks and models have been published to update usability

evaluation methods (UEMs) [3]; [20]; however, these frameworks and models are not applicable to all domains because they were developed to deal with certain aspects of usability in certain areas [16].

It can be seen from the above that there is need to an effective and appropriate methodology for evaluating the emerging domains/technology to measure their levels of efficiency, effectiveness and satisfaction, and ultimately to improve their quality. Also, there is need for a method that is context of use and that considers expert and user perspectives. This finding and the criticality of website usability has encouraged researchers to formulate such a framework. This framework should be applicable across numerous domains. In other words, it should be readily capable of adapting in any domain and for any technology. This paper constructs this framework for generating a context- specific method for the chosen domain that can be applied without needing to conduct user testing. However, developing and testing a method is not quick and it should involve some key stages. The next section describes the steps employed in the adaptive framework, also, it describes the process used to test it.

3. CONSTRUCTION OF THE ADAPTIVE FRAMEWORK

The adaptive framework was developed according to an established methodology in HCI research. It consists of two distinct phases: 1) Development phases that consist of four main steps for gathering together suitable ingredients to develop a context-specific method (DSI) for website evaluation; and 2) Validation phases for testing the developed DSI method practically (these are outlined in Figures 1 and 2). Below is an explanation of the four development steps:

Development Step One (D1: Familiarization): This stage starts from the desire to develop a method that is context of use, productive, useful, usable, reliable and valid, and that can be used to evaluate an interface design in the chosen domain. It entails reviewing all the published material in the area of UEMs but with a specific focus on knowledge of the chosen domain. Also, it seeks to identify an approach that would support developers and designers in thinking about their design from the intended end-users' perspective.

Development Step Two (D2: User Input): This stage consists of mini-user testing (task scenarios, think aloud protocol and questionnaire). Users are asked to perform a set of tasks on a typical domain website and then asked to fill out a questionnaire. The broad aim of this is to elicit feedback on a typical system from real users in order to appreciate the user perspective, to identify requirements and expectations and to learn from their errors. Understanding user needs has long been a key part of user design, and so this step directly benefits from including the advantages of user testing.

Development Step Three (D3: Expert Input): This stage aims to consider what resources are available for addressing the need. These resources, such as issues arising from the mini-user testing results and the literature review, require a discussion amongst experts (in the domain and/or usability) in order to obtain a broader understanding of the specifics of the prospective domain. Also, it entails garnering more information through conversations with expert evaluators to identify the areas/classification schemes of the usability problems related to the selected domain from the overall results. These areas provide designers and developers with insight into how interfaces can be designed to be effective, efficient and satisfying; they also support more uniform problem description and they can guide expert evaluators in finding real usability problems, thereby facilitating the evaluation process by judging each area and page in the target system.

Development Step Four (D4: Draw Up DSI Method: data analysis): The aim of this step is to analyse all the data gathered from the previous three. Then, the DSI method will be established (as guidelines or principles) in order to address each area of the selected domain.
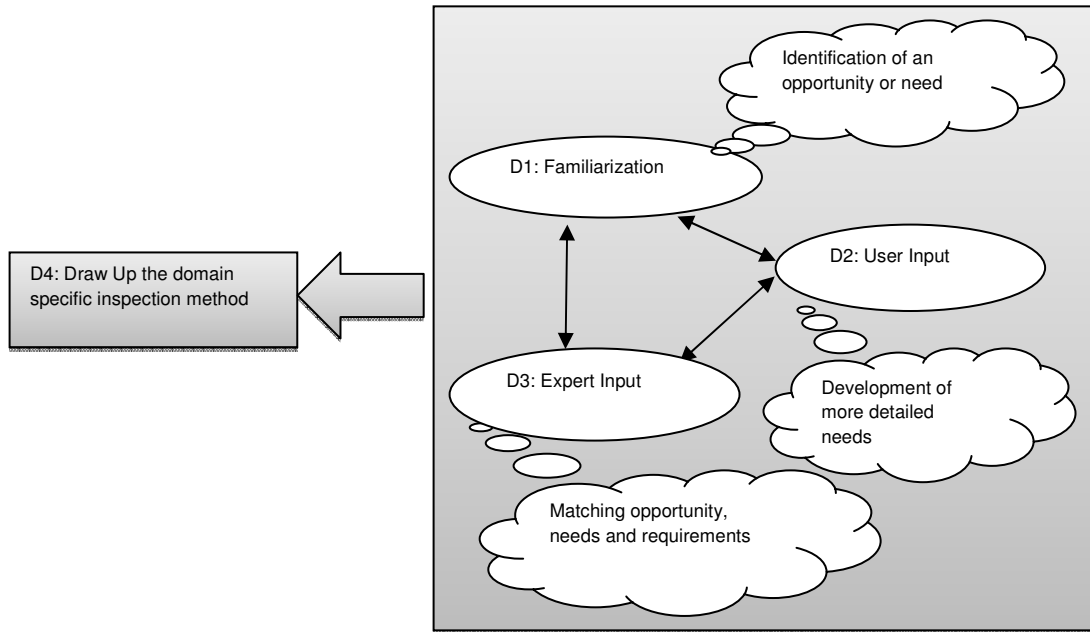
FIGURE 1: Development Stages of the Adaptive Framework.

After constructing the DSI framework, the researchers test it intensively through rigorous validation methods to verify the extent to which it achieves the identified goals, needs and requirements that the method was originally developed to address (this validation is outlined in Figure 2).
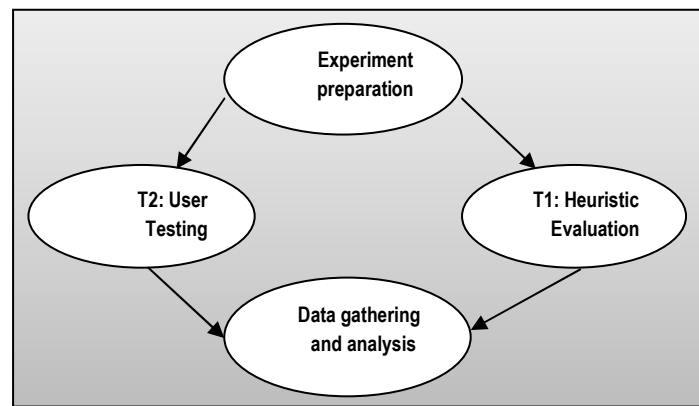

FIGURE 2: Testing stages of the adaptive framework

The validation phase of the adaptive framework again consists of four separate steps, as explained below:

1. Experiment Preparation (for DSI, HE and UT): Before the actual evaluation formally starts, the following initial preparations are needed: 1) Select a number of systems/websites that are typical of the chosen domain; 2) Recruit expert evaluators and users; 3) Plan the sequence of conducting the evaluations by each group in such a way as to avoid any bias; and 4) Prepare the experiment documents. The initial experiment preparation phase is concluded with a pilot experiment to make sure that everything is in place and ready for the actual evaluation.

2. Heuristic Validation (Expert Evaluation (HE)): The aim of this step is the validation of the newly developed method by conducting a heuristic evaluation (HE). Expert evaluators need a familiarization session before the actual evaluation. The actual expert evaluation is then conducted using the newly developed DSI method alongside HE. The aim of this process is to collect data ready for analysis (analytically and statistically), as explained in step 4.

3. Testing Validation: (User Evaluation (UT)): The aim of this step is to complement the results obtained from the expert evaluation, by carrying out usability lab testing on the same websites. [36] recommends conducting usability testing (UT) with HE because each one is complementary to the other. Then, the performance of HE is compared with the lab testing to identify which problems have been identified by UT and not identified by HE and/or DSI, and vice versa. The aim of this process is to collect data ready for analysis (empirically and statistically) in step 4.

4. Data Analysis: This step aims to analyse all the results and to answer all the questions raised from the above steps in a statistical manner. It is conducted in two parts; one focused on HE and the other on UT. The researchers extract the problems discovered by the experts from the checklists of both DSI and HE. Then, they conduct a debriefing session with the same expert evaluators to agree on the discovered problems and their severity, and to remove any duplicate problems, false positives or subjective problems. Then, the problems approved upon are merged into a master problem list, and any problems upon which the evaluators disagree are removed. Ultimately, the researchers conduct a comparison on the results of both methods (DSI and HE) in terms of the number of problems discovered (unique and overlapping), their severity ratings, which problems are discovered by HE and not discovered by DSI and vice versa, the areas of the discovered problems, the UEM performance metrics, evaluator reliability and performance, and the relative costs entailed in employing the two methods.

In the second part, the researchers conduct a debriefing session with independent evaluators to rank the severity of the problems derived from the user testing and to remove any duplicate problems. Following this, they establish the list of usability problems for UT. Subsequently, a single unique master list of usability problems is consolidated from the three methods. A comparison of the results of the three methods is then conducted in terms of the number of problems discovered (unique and overlapping), their severity ratings, and the areas of the discovered problems; this is to identify which problems were discovered by HE and DSI and not discovered by UT, and vice versa. Also, the UEM performance metrics of each method are measured, together with other measures, such as their relative costs and reliability. Moreover, this final step seeks to prove or refute the efficacy of conducting UT and HE with DSI.

Having proposed the framework above, it was decided to evaluate its practicality by applying it to a real-life experiment. From the literature review, it was found that the evaluation of the free educational websites domain is a subject area that has not yet been fully explored, nor have any context-specific methods been generated for this domain (to overcome the shortcomings of HE and UT); this is an important area of research because these websites are now essential to many users and companies. A well-designed educational website (i.e. one that is aesthetically attractive and is easy to use) can positively affect the number of people who become members.

4. RESEARCH METHODOLOGY
The experimental approach was selected to address the research hypotheses outlined above. Essentially, this section describes the methodology employed in this comparative study. Before conducting this experiment, a set of procedures were followed by the researchers, as follows:

4.1 Design
This experiment employs the between-subject and within-subject designs. The independent variables are the three methods (HE, DSI and UT). The dependent variables are the UEM

performance metrics, which are calculated from the usability problems reported by the evaluators/users, and from the reliability and efficiency measurements.

4.2 Evaluation of the Practicality of the Framework

In the first stage, the researchers conducted a literature review on the materials relating to usability and UEMs as well as on the requirements of the educational websites domain. In stage two, a mini-user testing session was conducted through a brief questionnaire that entailed four tasks, which were sent to ten users who are regular educational website users. In stage three, a focus group discussion session was conducted with eight experts in usability and the educational domain (i.e. single and double experts). Cohen's kappa coefficient was used on the same group twice to enable a calculation of the reliability quotient for identifying usability problem areas. In stage four, the researchers analysed the results of the three stages and incorporated the findings. The intra-observer test-retest using Cohen's kappa yielded a reliability value of 0.8, representing satisfactory agreement between the two rounds. After that, the usability problems areas were identified to facilitate the process of evaluation and analysis, and to help designers and programmers to identify the areas in their website that need improvement. Then, the DSI method was established, closely focused on educational websites, taking into an account what is called "learner-centred design". The method was created and classified according to the usability problem areas detailed in Table 2 below.

Usability problem area/attributes	Domain Specific Inspection (DSI)
User usability	Supports modification and progress of evaluation
	Supports user tasks and avoids difficult concepts
	Feedback and support services
	Easy to remember
Motivational factors	Supports leaner curiosity
	Learning content design and Attractive screen design
	Motivation to learn
Content information and process orientation	Relevant, correct and adequate information
	Reliability and Validity
	Privacy and Security
Learning process	Assessment
	Interactivity
	Evokes mental images for the learner
	Resources
	Learning management
	Learnability
Design and media usability	Multimedia representations
	Accessibility and compatibility of hardware devices
	Functionality
	Navigation and Visual clarity

TABLE 2: Final Version of Domain Specific Inspection (DSI).

4.3 Selection of the Targeted Websites

The first step in an initial preparation phase is selecting the websites. The researchers sought to ensure that the selected websites would support the research goals and objectives. The selection process was criteria-based; six aspects were determined and verified for each website, and these are: 1) Good interface design, 2) Rich functionality, 3) Good representatives of the free educational websites, 4) Not familiar to the users, 5) No change will occur before and during the actual evaluation, and 6) Completely free educational websites. In order to achieve a high level of quality in this research, the researchers chose three well-known websites in this domain. The first website was 'skoool'. It is an Intel driven initiative that delivers highly innovative and interactive learning resources via cutting-edge technologies and devices [47]. The second website was

AcademicEarth. It is an organization founded with the aim of giving everyone on earth access to a world-class education [2]. The third website was BBC KS3bitesize. This website helps school students from 11 to 14 with their coursework, homework and test preparation [8]. All of these have all the aspects mentioned above.

4.4 Recruitment of Experts and Users
The selection of usability experts and users is the second important step in the initial preparation phase in this experiment. The researchers decided to recruit eight expert evaluators, divided into two groups of four, who were carefully balanced in terms of experience. In each group, there are two double expert evaluators (usability specialists in educational websites) and two single expert evaluators (usability specialists in general). Each evaluator was to conduct his/her evaluation separately in order to ensure independent and unbiased evaluations [37]. Also, each group employed two methods, namely DSI and HE, to evaluate the three different websites. The evaluation was carried out in a prescribed sequence, i.e. one group used HE, DSI and HE on Skoool, AcademicEarth and BBC KS3bitesize, while the second group used DSI, HE and DSI as shown in Figure 3 below. The researchers adopted this technique to avoid any bias in the results and also to avoid the risk of any expert reproducing his/her results in the second session through over-familiarity with one method, i.e. each evaluation was conducted with a fresh frame of mind.
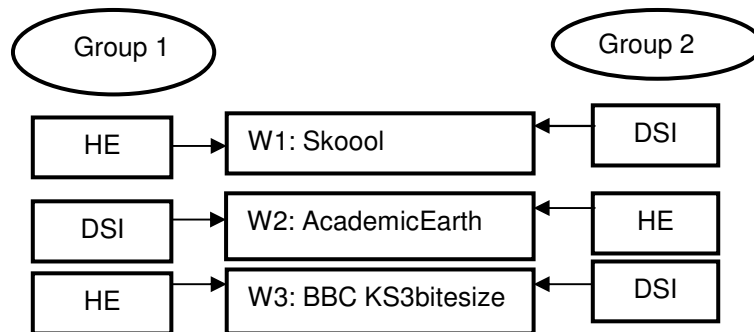

FIGURE 3: Usage of Methods by Group.

Selecting and recruiting users must be done carefully; the participants must reflect the real users of the targeted website because inappropriate users will lead to incorrect results, thereby invalidating the test. Appropriate users will deliver results that are more reliable; they will also be motivated to conduct the experiment [17]. There is no agreement on how many users should be involved in usability testing. Dumas and Redish (1999) suggested that 6 to 12 users are sufficient for testing, whereas other studies have recommended that 7, 15 and 20 users are the optimal numbers for evaluating small or large websites; particularly 20 users if benchmarking is needed [39]. At this point, 60 users were engaged; they were chosen carefully to reflect the real users of the targeted websites and were divided into three groups for each website, i.e. a total of 20 users for each website. The majority of the users are students, and they were mixed across the three users groups in terms of gender, age, education level and computer skills.

4.5 Building the Experiment Documents
In experiment preparation phase, the third step was building the set of preparation documents for HE, DSI and UT, such as an introduction sheet, heuristics checklists, tasks sheet, ranking problems sheet, observer sheet, demographics, satisfaction and Likert questionnaires, problems sheet and a master problem list. The introduction sheet contains the goals and objectives of the evaluation and the roles of users and experts. Before starting actual evaluation, users and experts completed a demographics questionnaire to obtain more information about them. Expert evaluators used checklists that had been developed by the researchers to facilitate the evaluation process for DSI and HE. Users used the task sheet that was designed according to the main functions that users would normally expect to perform on the three educational websites. A combination of task designs and TA approaches (as mentioned in the literature review) were used in this experiment. There were four sub-tasks in each three task group, they were kept to a

reasonable time limit and they were interesting and engaging enough to hold the users' attention. As briefly mentioned above, usability testing requires an observer, and the researcher adopted this role in all the sessions, noting all the comments made by the users. The researcher used a stopwatch to record the time spent by each user on each task, and an observation sheet to write down the behaviour of each user and the number of problems encountered. The ranking sheet aims to help the expert evaluators and independent evaluators (for user testing) to rank the severity of usability problems by using Nielsen's scale as mentioned above. After the evaluators had finished their evaluations and had ranked HE and DSI problems, they were asked to complete a satisfaction questionnaire using the System Usability Scale (SUS) on both methods. It is made up of ten items in the form of scale questions ranging from 0 to 100 to measure the satisfaction of expert evaluators [9]. Also, when the users finished their tasks, they were asked to rate their level of satisfaction in a questionnaire on a scale of one to seven, where one refers to 'highly unsatisfactory' and seven indicates 'highly satisfactory'. This scale has been suggested to truthfully measure the levels of satisfaction that are felt by users on a website interface following a test [39]. Evaluators and users were asked to fill in an open-end questionnaire by writing down their comments and feedback on the methods used and explaining any reaction that was observed during the test. Subsequently, the Likert scale was used by the evaluators for measuring either positive or negative responses to a statement in both the DSI and HE methods. Moreover, the researchers extracted the problems of three methods from the problems sheet and removed all false positive ('not real') problems, evaluators' 'subjective' problems and duplicated problems during the debriefing session. The problems agreed upon were merged into a unique master problem list and any problems upon which the evaluators disagreed were removed.

4.6 Piloting the Experiment
The forth step in an initial preparation phase is a pilot experiment. It was conducted by two independent evaluators and fifteen users. All materials were checked to make sure that there were no spelling or grammatical errors and no ambiguous words or phrases, and that all of sentences in the instruments (heuristics, check-lists, task scenarios, questionnaires and procedures) were sufficiently clear to be used by the evaluators. A fewer minor improvements were made, such as before establishing the final version of DSI, the independent evaluators suggested removing game heuristics from Design and Media Usability area because approximately 95% of games websites are not educational websites as confirmed by [22].

Furthermore, to assess the time needed for testing, the fifteen users were divided into three groups (five users in each). Each group performed its tasks. The users' behaviour was monitored, and all the usability measures were assessed as they would be in real testing. All of these steps resulted in useful corrections and adjustments for the real test. Also, the test environment was a quiet room. We attempted to identify what equipment the users regularly use and set it up for them before the test, for example, using the same type of machine and browser.

4.7 Actual Evaluation
The Heuristics Validation phase started with a training (familiarization) session for the eight expert evaluators. They were given a UEM training pack that contained exactly the same information for both groups. The researchers emphasized to each evaluator group that they should apply a lower threshold before reporting a problem in order to avoid misses in identifying real problems in the system. Then, the actual expert evaluation was conducted and the evaluators evaluated all websites consecutively, rating all the problems they found in a limited time (which was 90 minutes). After that, they were asked to submit their evaluation report and to complete the five-point scale in the SUS questionnaire (1 for strongly disagree and 5 for strongly agree) to rate their satisfaction on the evaluation method they had used (DSI or HE), and to give feedback on their own evaluation results.

The Testing Validation phase started with a training (familiarization) session for the 60 users; it involved a quick introduction on the task designs, the TA approach and the purpose of the study. The next step entailed explaining the environment and equipment, followed by a quick demonstration on how to 'think aloud' while performing the given tasks. Prior to the tests, the

users were asked to read and sign the consent letter, and to fill out a demographic data form that included details such as level of computer skill. All the above steps took approximately ten minutes for each test session. The actual test started from this point, i.e. when the user was given the task scenario sheet and asked to read and then perform one task at a time. Once they had finished the session, they were asked to rate their satisfaction score relating to the tested website, to write down their comments and thoughts, and to explain any reaction that had been observed during the test, all in a feedback questionnaire. This was followed by a brief discussion session.

5. DATA ANALYSIS AND MEASUREMENTS

To determine whether our adaptive framework has generated an evaluation method of sufficiently high quality, the results of the comparison process between the three methods needed a meta-analysis to be performed, as follows:

1. Compare the average time spent by each group when using each method during the evaluation sessions.

2. Compare the results of the usability problems and their severity in order to assess the performance of each method in terms of identifying unique and overlapping problems and of identifying real usability problems in the usability problem areas.

3. Comparing the satisfaction scores of HE and DSI by using System Usability Scale (SUS).

4. Reliability of HE and DSI: This can be measured from employing the 'evaluators' effect formula' (Any-Two-Agreement). It is used on each single evaluators in order to measures the performance of the evaluators individually [21].

   Any-Two-Agreement = Average of $|Pi \cap Pj|$ / $|Pi \cup Pj|$ over all $\frac{1}{2}$ n (n-1) pairs of evaluators, where Pi is the set of problem discovered by evaluator i and the other evaluator j, and n refers to the number of evaluators.

5. Evaluators' Performance: This can be measured by the performance of single and double expert evaluators in discovering usability problems by using HE and DSI in each group and website.

To make further comparisons between the performance of HE, DSI and UT in identifying usability problems, a set of UEM and other metrics were used for examining their performance; none of these metrics on their own addresses errors arising from false positive, subjective and missed problems. They are efficiency, thoroughness, validity, effectiveness, reliability and cost. Efficiency in UEMs is the "ratio between the numbers of usability problems detected to the total time spent on the inspection process" [19]. Thoroughness is perhaps the most attractive measure; it is defined as a measure indicating the proportion of real problems found when using a UEM to the total number of known real problems [29]. Validity is the extent to which a UEM accurately identifies usability problems [45]. Effectiveness is defined as the ability of a UEM to identify usability problems related to the user interface [28]. The reliability of user testing can be measured by the mean number of evaluators to the number of real problems identified [11]. The cost can be calculated by identifying the cost estimates. It can be done fairly simply by following Nielsen's equation who estimated the hourly loaded cost for professional staff at $100 [37]. All of them are computed as follows:

1. Efficiency = (No. of problems)/(Average time spent)

2. Thoroughness = (No. of real usability problems found)/(Total no. of real usability problems present)

3. Validity = (No. of real usability problems found)/(No. of issues identified as a usability problem)

4. Effectiveness = Thoroughness × Validity

5. Reliability of UT = (Mean no. of evaluators)/(No. of real problems identified)

6. Cost = (No. of evaluation hours) × (Estimate of the loaded hourly cost of participants)

To test the research hypotheses and choose the correct statistical test in SPSS, the normality of the data should be examined. The results of both Skewness and Kurtosis are equal to 0, and the Sig-value of the Kolmogorov-Smirnov test are greater than 0.05. As a result, the data are normally distributed, and t-Test, One-way ANOVA and Pearson correlation were chosen at 5% significance level as our methods for statistical analysis, as the dependent variables in our data are independent of each other, improving the validity of using analysis of variance. The Mann-Whitney test was used to analyse the Likert score (considered as an ordinal scale).

6. ANALYSIS AND DISCUSSIONS
This section describes the results obtained from using the three method adopted in this study. It starts by detailing the result of the HE and DSI methods separately, including quantitative and qualitative analyses. This is followed by detailing the result of the UT method alone, including quantitative and qualitative analyses. Ultimately, all the results derived from the three methods were compared in terms of the numbers of problems and types, as well as the other usability metrics as mentioned above.

6.1 Analysis for HE and DSI Results
6.1.1 Time spent: It can be seen from Tables 3 and 4 that the average time taken for doing the three experiments using DSI was 24.25 minutes with a standard deviation of 6.7, whereas the DSI average was 42.58 minutes with a standard deviation of 7.1. This difference in time spent is not significant (F = 0.199, p = 0.660) using the t-test. The group who used DSI managed to evaluate the website more quickly than the other group but discovered fewer usability problems. The group that used DSI spent almost double the time evaluating the website but discovered almost three times as many real usability problems. There was a statistically significant positive relationship between time spent and problems discovered by using the Pearson correlation test, where the Sig value is 0.020 at the 0.05 level. This result reveals that the users who spent more time were able to discover more usability problems.

Website	Skoool	AcademicEarth	BBCKS3bitesize
Evaluator	**Time**	**Time**	**Time**
1	25	45	24
2	30	50	40
3	25	55	22
4	20	29	23
Heuristics	HE	DSI	HE
# of problems	10	29	2
Mean time taken	25	45	27

TABLE 3: Average Time Taken and Number of Problems for Group 1.

Website	Skoool	AcademicEarth	BBCKS3bitesize
Evaluator	**Time**	**Time**	**Time**
1	42	30	50
2	40	17	38
3	38	15	35
4	45	20	44
Heuristics	DSI	HE	DSI
# of problems	33	13	12
Mean time taken	41	21	42

TABLE 4: Average Time Taken and Number of Problems for Group 2.

An explanation for the differences in time spent and number of problems located is gleaned from the evaluators' feedback. They said that HE was not particularly helpful, understandable or memorable for them. However, DSI helped them to develop their skills in discovering usability problems in this application area; also, this set was more understandable and memorable during their evaluation. To further analyse these factors of time spent and number of problems discovered, efficiency metrics were applied. DSI proved to be more efficient than HE in discovering usability problems (DSI = 1 vs. HE = 0.7), as Table 5 shows.

Method	Skoool	Academic Earth	BBC KS3bitesize	**Mean**
	Efficiency	**Efficiency**	**Efficiency**	
HE	0.77	1.1	0.2	0.7
DSI	1.55	1.2	0.6	1

TABLE 5: Mean Score of Efficiency for Two Methods.

6.1.2 Number of Usability Problems

Table 6 shows that HE was able to uncover 25% of the total number of real usability problems. However, DSI was able to uncover 75% of the total number of real usability problems in the websites (no false and subjective problems).

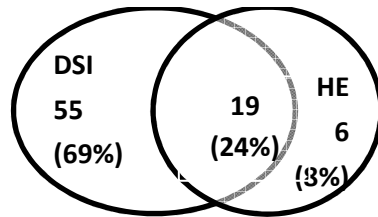| Website	Group	Expert and type	Method	# of problems found by each evaluator	# of problems with repetition	# of problems without repetition	# of problems with repetition between groups	% of problems found by each evaluator	% # of problems found by each group
Skoool	G 1	Ev. 1[+]	HE	8	19	10	43	19%	23%
		Ev. 2[^]	HE	5				12%	
		Ev. 3[^]	HE	1				2%	
		Ev. 4[+]	HE	5				12%	
	G 2	Ev. 1[+]	DSI	21	64	33		49%	77%
		Ev. 2[^]	DSI	15				35%	
		Ev. 3[+]	DSI	13				30%	
		Ev. 4[^]	DSI	15				35%	
Academic Earth	G 1	Ev. 1[+]	DSI	10	42	29	42	24%	69%
		Ev. 2[^]	DSI	11				26%	
		Ev. 3[^]	DSI	9				21%	
		Ev. 4[+]	DSI	12				29%	
	G 2	Ev. 1[+]	HE	6	23	13		14%	31%
		Ev. 2[^]	HE	6				14%	

FIGURE 5: Overlapping Problems between Both Methods.

Table 8 shows the severity ratings for the real problems discovered (cosmetic, minor, major and catastrophic). Overall, a great many real usability problems were discovered. The most significant results were obtained from using DSI, while HE found fewer (or no) usability problems. The t-test reveals that there is significant difference between the two methods in terms of severity problems ($t = -1.877$, $df = 4$, $p = 0.026$).

Website	Severity of problems	Type of method			
		HE		DSI	
Skoool	Cosmetic	Group 1	3	Group 2	12
	Minor		3		11
	Major		2		6
	Catastrophic		2		4
	Severity (average)	2.3		2.1	
AcademicEarth	Cosmetic	Group 2	2	Group 1	11
	Minor		7		11
	Major		3		4
	Catastrophic		1		3
	Severity (average)	2.2		2	
BBC KS3bitesize	Cosmetic	Group 1	2	Group 2	2
	Minor		0		6
	Major		0		3
	Catastrophic		0		1
	Severity (average)	1		2.25	
Overall severity (average)		1.8		2.1	
No. of discovered real problems		25		74	

TABLE 8: Total Number of Usability Problems with Severity Ratings and Averages.

The quantitative assessment of the heuristic evaluation results was a comparison between the two sets of methods, in particular, in terms of the areas of usability problems found in this experiment. These areas assisted in identifying how each method performed in each usability problem area. The eight expert evaluators discussed, agreed and decided on the categories to which the problems should belong in both sets, as Tables 9 and 10 illustrate. The overall results from both tables show that the two groups (and the three websites) revealed more usability problems by using DSI in all areas than HE, particularly in Learning process, Motivational factors, Design and media usability, User usability and Content information and process orientation, respectively. Three out of the ten HE heuristics performed more efficiently than four others and the three remaining failed to expose a sufficient number of usability problems. This suggests that the HE heuristics are rather too general, and are unlikely to encompass all the usability attributes of user experience and design in interactive learning systems.

Heuristic Evaluation	Skoool	AcademicEarth	BBC KS3bitesize
Visibility of system status	1	2	0
Match between the system and the real world	0	4	0
User control and freedom	1	3	0
Consistency and standards	1	0	0
Error prevention	0	1	0
Recognition rather than recall	3	1	0
Flexibility and efficiency of use	1	0	0
Aesthetic and minimalist design	2	1	0
Helps users recognize, diagnose and recover from errors	0	0	0
Help and documentation	1	1	2
Total problems	10	13	2

TABLE 9: Usability Problems Found by Category through HE.

Usability problem area	Skoool	Academic Earth	BBC KS3bitesize
User usability	4	5	2
Motivational factors	5	6	1
Content information and process orientation	4	3	0
Learning process	11	7	7
Design and media usability	9	8	3

TABLE 10: Usability Problems Found by Category through DSI.

6.1.3 UEM Performance Metrics

After employing the above formulae in terms of the UEM metrics, and as depicted in Table 11 below, the t-test was used to investigate the statistical differences between the DIS and HE. The measure of thoroughness for the DSI set in identifying the number of real problems was higher than for HE (0. 5 vs. 0.2). Also, the t-test revealed significant difference between them (t = -2.227, df = 19.208, p = 0.037). Also, the measure of validity for DSI was higher (in accurately identifying real usability problems) than for HE (0.7 vs. 0.4). There was a significant difference between them (t = -2.966, df = 20.705, p = 0.007). The effectiveness of DSI was also higher that for HE (0.3 vs. 0.1). Again, there was a significant difference between them (t = -2.212, df = 21.717, p = 0.038). Furthermore, the reliability values for DSI were slightly higher than for HE (0.5 vs. 0.4). It can now be concluded that there is agreement amongst the evaluators over the usability problems, and it is of a high level. Also, the t-test revealed a significant difference (t = 3.181, df = 11.326, p < 0.000). Finally, the average results for the cost of employing the two methods show that there is a slight difference (Table 12); (DSI = $1.240 vs. HE = $1.017).

Method	Skoool		Academic Earth		BBC KS3bitesize		Overall mean	
	HE	DSI	HE	DSI	HE	DSI	HE	DSI
Thoroughness	0.3	0.5	0.3	0.4	0	0.3	0.2	0.5
Validity	0.3	0.7	0.5	0.8	0	0.5	0.4	0.7
Effectiveness	0.1	0.4	0.2	0.3	0	0.2	0.1	0.3
Reliability	0.4	0.4	0.4	0.6	0.3	0.5	0.4	0.5

TABLE 11: Mean Score for UEM for the Two Methods.

Method	Skoool	Academic Earth	BBC KS3bitesize	Mean cost
Heuristic Evaluation (HE)	$1,020 This includes the time spent by 4 evaluators (1.6 hours) + hours collecting data from the evaluation sessions + 6 hours analysing data.	$990 This includes the time spent by 4 evaluators (1.3 hours) + hours collecting data from the evaluation sessions + 6 hours analysing data.	$1,040 This includes the time spent by 4 evaluators (1.8 hours) + 2.6 hours collecting data from the evaluation sessions +6 hours analysing data.	$1017
Domain Specific Inspection (DSI)	$1,130 This includes the time spent by 4 evaluators (2.7 hours) + 3 hours collecting data from the evaluation sessions + 6.6 hours analysing data.	$1,250 This includes the time spent by 4 evaluators (2.9 hours) + 3 hours collecting data from the evaluation sessions + 6.6 hours analysing data.	$1,240 This includes the time spent by 4 evaluators (2.8 hours) + 3 hours collecting data from the evaluation sessions + 6.6 hours analysing data.	$1206

TABLE 12: Cost of Employing Both Methods.

6.1.4 Post- test Questionnaire

- **Satisfaction score:** The researchers used the System Usability Scale (SUS) as previously mentioned. HE delivered a lower overall score, at 46, whereas DSI delivered a much higher score, at 71.

- **Opinion and attitudinal questions (Likert scale)**
  The expert evaluators completed the aforementioned Likert scale, and the scores were calculated for each statement of Likert questionnaire in order to obtain the overall results concerning the expert evaluators' opinion with the DSI and HE. A Likert score of 1-2 was regarded as a negative response, 4-5 a positive response, and 3 a neutral one. The Cronbach's Alpha test used to measure the reliability of responding and the result was 0.84. The Likert scores revealed that evaluators satisfied overall with the DSI and the results were significant differences between DSI and HE by using Mann-Whitney test as Table 13 shows.

Method	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Mann-Whitney U	3.000	8.000	11.000	5.000	14.500	12.000	.000
P-value (2-tailed)	.002	.009	.020	.004	.042	.023	$p < 0.001$

TABLE 13: Results of Mann-Whitney for Both Methods.

6.2 Quantitative Analysis for Usability Testing Result
6.2.1 Time Spent

Table 14 shows the time spent by each user on performing the experiment. The Skoool groups spent the longest time, more than the BBC KS3bitesize and Academic Earth groups, with 112, 96 and 88 minutes, respectively. This again was probably due to problems in navigation, structure and function in the three websites, which caused the users to spend more time in accomplishing their tasks. This was particularly so in the Skoool website, as some tasks were abandoned because the users had doubts about how to accomplish them. Also, in the BBC KS3bitesize

website, the group spent time thinking about how to perform some tasks, such as the 'registration' task and the 'post a question' task. The average time spent by each user in all three groups was more than 1.1 minutes. The efficiency formula used for UT for all the experiments, in terms of number of usability problem discovered over time spent, delivered a mean score of 0.4 (Skoool = 0.1, Academic Earth = 0.1, BBC KS3bitesize = 0.2). The One-way ANOVA test was used to determine any significant difference in terms of time spent; the results reveal significant a difference (F = 6.616, p = 0.003). However, there was no statistical difference in terms of efficiency (F = 0.109, p = 0.458).

Usability measure	Skoool	Academic Earth	BBC KS3bitesize
Total time spent by all users (in minutes)	112	88	96
Average time per user per task (in minutes)	1.4	1.1	1.2
Average time per user over four tasks	5.6	4.4	4.8

TABLE 14: Time Taken on Conducting the Evaluation.

6.2.2 User Satisfaction

It can be seen clearly that BBC KS3bitesize delivered the highest overall score, at 7, whereas Skoool delivered the second highest score, at 5, and Academic Earth delivered the lowest score among the three websites, at 3. This indicates that there were certain factors that influenced the users, which then affected the satisfaction rating for the tested website, as evidenced by the critical user comments on the design features of each website. These factors are the various activities, such as the test and revise functions that each website provided (or the games); also, the users were encouraged by simple and attractive designs. These results are similar to what evaluators stated.

6.2.3 Number of Usability Problems Discovered

Table 15 details the total numbers of usability problems found by user testing and their severity rating. All the redundant problems were removed. The usability problems detected in BBC KS3bitesize numbered 16, higher than in the Skoool and Academic Earth websites (13 vs. 12). The One-way ANOVA test was used, revealing no statistical difference amongst the numbers of problems found (p > 0.05). Pearson correlation was used, and the results reveal a positive relationship between time spent and problems discovered (the Sig value is 0.013). This result reveals that the users who spent more time were able to discover more usability problems.

Problem type	Skoool	Academic Earth	BBC KS3bitesize	Total no. of problems without duplication
	Total no. of usability problems	Total no. of usability problems	Total no. of usability problems	
Catastrophic	1	0	0	1
Major	3	3	2	8
Minor	2	2	5	9
Cosmetic	7	7	9	23
No. of problems	13	12	16	41

TABLE 15: Numbers of Usability Problems Discovered.

6.2.4 UEM Performance Metrics

By applying the UEM and reliability formulae, Table 16 details the thoroughness of UT in identifying real usability problems, with a mean of 0.3. The validity of UT in finding the known

usability problems was 0.2, and the effectiveness of UT in identifying usability problems related to the user interface was 0.1. The One-way ANOVA test was used to identify any significant difference between them in each website as a dependent factor. The results reveal that there are significant differences in regard to thoroughness (F = 9.873, p < 0.001), validity (F = 8.435, p = 0.001), effectiveness (F = 7.754, p =0.001) and reliability (F = 9.612, p < 0.001). The results for the costs of employing UT on each website were little different with an average $1,667, as shown in Table 176.

Metric\Websites	Skoool	Academic Earth	BBC KS3bitesize	Mean Total
Thoroughness	0.4	0.3	0.1	0.3
Validity	0.3	0.3	0.1	0.2
Effectiveness	0.1	0.1	0.01	0.1
Reliability	0.8	0.9	0.7	0.8

TABLE 16: The Mean Results for the UEM Metrics.

Evaluation method	Skoool	Academic Earth	BBC KS3bitesize	Mean cost
User Testing (UT)	$1,690 This includes the time spent by 20 users (1.9 hours) + 10 hours collecting data from the evaluation sessions + 5 hours analysing data.	$1,650 This includes the time spent by 20 users (1.5 hours) + 10 hours collecting data from the evaluation sessions + 5 hours analysing data.	$1,660 This includes the time spent by 20 users (1.6 hours) +10 hours collecting data from the evaluation sessions + 5 hours analysing data	$1667

TABLE 17: Cost of Employing UT in This Research.

6.3 Comparative Analysis to Evaluate the Adaptive Framework
This section represents comprehensive and comparative analysis between the three methods.

6.3.1 Types of Problems Found by UT in Relation to DSI and HE
Two independent expert evaluators were involved in discussing, agreeing on and deciding where the UT problems should be in HE, and to which category they should belong in DSI, as Tables 18 and 19 illustrate. The overall results from both tables show that all the UT problems were successfully classified into DSI, whereas just 11 problems out of 16 in the BBC KS3bitesize were classified into HE. This proves that HE is rather general, and is unlikely to encompass all user problems, such as usability problems in the Learning process area. Also, the tasks given to the users during the usability testing seem to have 'walked them through' the quality of the 'learning process', which could have increased the opportunity to discover problems. Furthermore, the findings confirm that User control and freedom, Motivational factors and Content information and Process orientation are a common weakness in dynamic websites (particular for educational websites). All the three websites found nearly equal numbers of usability problems related to navigation and visibility. In conclusion, UT worked better than HE because seven problems were not classified in it. However, all the users' problems were classified in the DSI.

Heuristic Evaluation	Skoool	Academic Earth	BBC KS3bitesize
Visibility of system status	6	1	4
Match between the system and the real world	2	0	3
User control and freedom	0	0	0
Consistency and standards	0	1	0
Error prevention	0	4	0
Recognition rather than recall	0	0	0
Flexibility and efficiency of use	1	0	0
Aesthetic and minimalist design	2	5	3
Helps users recognize, diagnose and recover from errors	0	0	0
Help and documentation	1	0	1
Total problems	**12**	**11**	**11**

TABLE 18: Usability Problems Found Compared with HE.

Usability problem area	Skoool	Academic Earth	BBC KS3bitesize
User usability	2	3	2
Motivational factors	1	0	1
Content information and Process orientation	1	1	1
Learning process	6	7	9
Design and media usability	3	1	3
Total problems	**13**	**12**	**16**

TABLE 19: Usability Problems Found Compared with the Domain Specific Inspection (DSI).

6.3.2 Performance of the Three Methods

Generally, Tables 20, 21 and 22 show how UT, HE and DSI revealed different types and numbers of usability problems. One-way ANOVA reveals that there is significant difference between the three methods in terms of discovering usability problems on the whole ($F = 13.447$, $p < 0.001$). UT, HE and DSI revealed 80%, 10% and 60% of the usability problems found in the BBC KS3bitesize website, respectively. One-way ANOVA-Tukey HSD was used and the results show that there is a strongly significant difference amongst the methods in finding usability problems on the BBC KS3bitesize website between HE and UT, where $p = 0.003$. In the Skoool website, UT, HE and DSI revealed 38%, 29% and 97% of the found usability problems, respectively. One-way ANOVA-Tukey HSD was used and the results show that there is a strongly significant difference amongst the methods in finding usability problems in Skoool (as a dependent factor), particular between HE and DSI and between DSI and UT, where $p < 0.001$. Finally, UT, HE and DSI revealed 34%, 37% and 83% of the found usability problems in Academic Earth, respectively. One-way ANOVA-Tukey HSD was used and the results show that there is a significant difference amongst the methods in finding usability problems in Academic Earth between HE and UT, where $p = 0.044$. The performance of HE in discovering usability problems during the experiment ranged from 10% to 37%. UT discovered usability problems ranging from 38% to 80%, while DSI discovered usability problems ranging from 60% to 97%. Also, UT and HE performed better in discovering major, minor and cosmetic real usability problems, but DSI was the best in discovering more catastrophic, major, minor and cosmetic real usability problems. Furthermore, 9 unique problems were discovered in all experiments on the three websites through UT (6 in BBC KS3bitesize and 3 in Academic Earth), whereas the remaining UT problems were discovered by DSI (although one was discovered by HE). Thus, it can be seen that DSI was the best in discovering real problems; this was followed by UT, and then finally HE.

Method Problem type	UT	HE	DSI	Total problems (no duplicates)
Catastrophic	0 (0%)	0 (0%)	1 (100%)	1
Major	2 (66%)	0 (0%)	3 (100%)	3
Minor	5 (100%)	0 (0%)	5 (100%)	5
Cosmetic	9 (100%)	2 (22%)	2 (22%)	11
No. of problems	16 (80%)	2 (10%)	12(60%)	20

TABLE 20: Findings in BBC KS3bitesize.

Method Problem type	UT	HE	DSI	Total problems (no duplicates)
Catastrophic	1 (25%)	2 (50%)	4 (100%)	4
Major	3 (30%)	2 (20%)	6 (89%)	7
Minor	2 (29%)	3 (43%)	11 (85%)	11
Cosmetic	7 (54%)	3 (23%)	12 (92%)	12
No. of problems	13 (38%)	10 (29%)	33 (97%)	34

TABLE 21: Findings in Skoool.

Method Problem type	UT	HE	DSI	Total problems (no duplicates)
Catastrophic	0 (0%)	1 (33%)	3 (100%)	3
Major	3 (50%)	3 (50 %)	4 (66%)	6
Minor	2 (17%)	7 (58 %)	11 (92%)	12
Cosmetic	7 (50%)	2 (14%)	11 (79%)	14
No. of problems	12 (34%)	13 (37%)	29 (83%)	35

TABLE 22: Findings in Academic Earth.

6.3.3 Overlapping and Unique Problems

Many researchers recommend conducting UT together with HE because they have found that each method discovers unique problems [36], so when they are conducted together, they can reveal and present all the problems in the targeted website. Again, this experiment may confirm or deny this recommendation, depending on the following results. Table 23 shows the performance of the three methods on a unique performance basis for the three websites. DSI was able to discover 7 catastrophic, 12 major, 20 minor and 16 cosmetic problems that were not revealed by the other methods. HE was not able to identify any catastrophic problems alone; however, it was able to identify 1 major, 2 minor and 3 cosmetic problems. UT was not able to discover any major problems; however, it discovered 1 catastrophic, 3 minor and 5 cosmetic problems.

Figure 6 also shows the overlapping real usability problems discovered by the three methods. In fact, each method revealed different types of problem (both unique and overlapping). However, DSI revealed the majority of real usability problems, indicating those with high severity ratings, and DSI also appeared to work fruitfully for the expert evaluators, who then revealed more real problems, both unique and overlapping. For example, DSI found 62% unique problems out of the total number of real usability problems (n = 55 out of 89). HE found 7% unique problems out of the total number of real usability problems (n = 6 out of 89), and UT identified 10% unique problems out of the total number of real usability problems (n = 9 out of 89). 19 (21%) real problems out of 89 were discovered as 'overlapping' by the three methods. The clear superiority of DSI was due to involving user inputs in drawing up the method (as in one stage of the adaptive framework), and due to DSI having characteristics that are appropriate to the educational domain.

Problem types	HE (unique)	DSI (unique)	UT (unique)	HE & UT (overlap)	DSI & UT (overlap)	DSI & HE (overlap)	Total number of problems in three websites (unique)
Catastrophic	0	7	1	0	1	3	8
Major	1	12	0	2	6	3	13
Minor	2	20	3	7	7	6	33
Cosmetic	3	16	5	9	27	7	35
Total	6	55	9	18	29	19	89

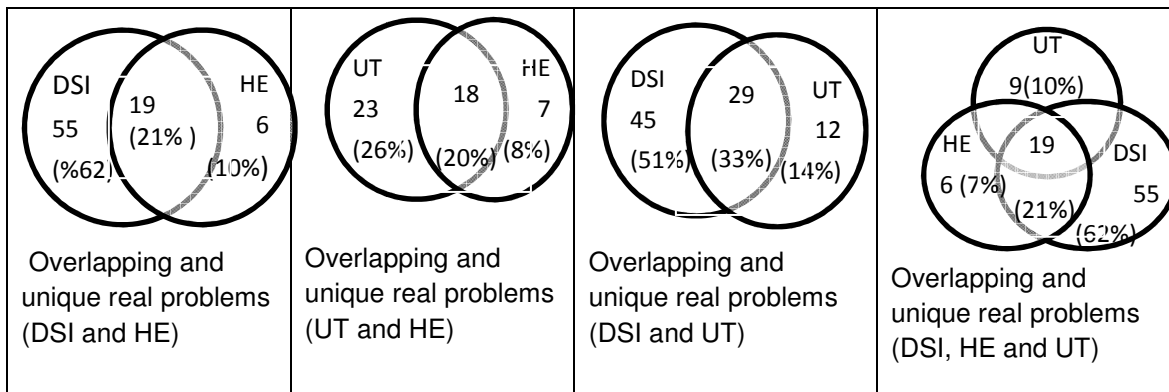**TABLE 23:** Each Method's Performance with Severity Rating.

FIGURE 6: Each Method's Performance, Uniquely and Working in Pairs.

It can also be seen that combining the results of DSI with either HE or UT offers better performance in terms of catastrophic, major, minor and cosmetic problems, whereas combining HE with either DSI or UT offers quite good results in terms of cosmetic problems. Combining UT with either DSI or HE offers better results in terms of minor and cosmetic problems. To sum up, the result of the comparison between UT and HE confirms conduct UT with HE in order to overcome the shortcomings of each, because each one is indeed complementary to the other. On the other hand, DSI (as created from the adaptive framework) refutes this recommendation.

6.3.4 Usability Problem Areas: As evident in Table 24, DSI identified large numbers of real usability problems in all usability areas on the three websites (74). However, HE overall worked slightly better in discovering 25 real usability problems related to three usability problem areas but it failed in exposing any usability problems in two main usability problems areas, which are Motivational factors and Learning process, and it failed to identify a sufficient number of usability problems in the Content information and process orientation area. Furthermore, UT worked better in discovering usability problems (41) in three usability areas but it failed to identify a sufficient number of usability problems in the Content information and process orientation area.

Usability problem areas	UT	DSI	HE
User usability	✓ (14 problems)	✓ (20 problems)	✓ (15 problems)
Motivational factors	✳ (1 problem)	✓ (8 problems)	-
Content information and process orientation	✳ (4 problems)	✓ (7 problems)	✳ (2 problems)
Learning process	✓ (6 problems)	✓ (20 problems)	-
Design and media usability	✓ (16 problems)	✓ (19 problems)	✓ (8 problems)
Total number of problems	**41**	**74**	**25**

TABLE 24: Number of Usability Problem Areas Identified by the Three Methods.

6.3.5 Comparison between the Three Methods in UEM Performance Metrics

It can be seen from Table 25 that DSI was more efficient, thorough and effective in terms of identifying the total number of real problems to total time spent, and its ability to identify usability problems related to the user interface than the other two methods. UT is the second best method but it is more reliable than DSI or HE. HE delivered the worst result in terms of identifying real problems; however, it is the cheapest to use. Moreover, DSI is slightly more expensive than HE, and is cheaper that UT. One-way ANOVA reveals that there is significant difference between the methods used in terms of the UEM metrics results, as shown in Table 26.

Metric / Method	Efficiency	Thoroughness	Validity	Effectiveness	Reliability	Cost
HE	0.7	0.2	0.4	0.1	0.4	**$1017**
DSI	1	0.5	0.7	0.3	0.5	**$1206**
UT	0.4	0.3	0.2	0.1	0.7	**$1667**

TABLE 25: Comparing the Metrics between the Three Methods.

Metric	F	Sig. (*p*-value)
Efficiency	7.613	0.001
Thoroughness	3.950	0.023
Validity	3.525	0.034
Effectiveness	4.369	0.016
Reliability	38.571	$p < 0.000$

TABLE 26: One-way ANOVA Results between the Three Methods.

6.3.6 Advantages and Disadvantages of the Three Methods

This study has addressed the relative effectiveness of three methods for evaluating user interfaces, and it now offers some insights into each (see Table 27). Overall, DSI, as applied here, produced the best results; it found the most real problems, including more of the most serious ones, than did HE and UT, and at only a slightly higher cost. HE missed a large number of the most severe problems, but it was quite good in identifying cosmetic and minor problems. UT is the most expensive method and it missed some severe problems; however, it helps in discovering general problems and it assists, as does DSI, in defining the users' goals.

Method	Advantages	Disadvantages
Usability Testing	• Helps define and achieve the users' goals • Identifies the users' real problems • Identifies recurrent and general real problems	• Misses some severe real problems • High cost • Takes more time • Conducted under lab conditions
Heuristic Evaluation	• Identifies a few real problems • Low cost	• Misses some severe problems • Too general • Not readily applicable to many new domains
Domain Specific Inspection (DSI)	• Identifies many more real problems • Identifies more serious, major, minor and cosmetic real problems • Improves the evaluator's performance • Identifies the users' real problems and helps define and achieve the users' goals	• A little higher in cost than HE and cheaper than UT • Slightly higher in time than HE

TABLE 27: Summary of the Study's Findings.

7. DISCUSSION AND FINDINGS

This section explores the results of this experiment and highlights the main findings. It then draws out the lessons learned from the research. The main objective of this experiment was to evaluate the adaptive framework through its ability to generate new method, and specifically the specific inspection method for Educational domain\ (DSI) by comparing its results with usability testing (UT) and Heuristic Evaluation (HE). It has been clearly shown that the hypotheses were accepted and DSI was able to find all the real problems that were discovered by the UT and HE, but with greater efficiency, thoroughness and effectiveness. Also, DSI was better at discovering catastrophic, major, minor and cosmetic real problems. It seemed to guide the evaluators' thoughts in judging the usability of the website through clear guidelines that included all aspects of the educational quality of the website, which were represented in Content information and process orientation, Management of learning process, Quality of design and media, and Motivational factors to learn. As a result, it is unsurprising that the DSI method revealed a number of problems not discovered by the other two methods. HE method did not perform as well as either DSI or UT, based on the number of usability problems discovered during this experiment. The experts that used HE seemed to have their confidence undermined whilst performing the evaluation, for example, when they performed the evaluation, they found no readily applicable heuristic within HE for performing some of the main functions in these educational websites, such as Educational process and management. Consequently, HE performed poorly in discovering problems. The UT method performed modestly against DSI, and well against HE, based on the number of problems identified. Thus, the findings indicate that it is not essential to conduct UT in conjunction with HE, in order to address the shortcomings of these methods; rather, to avoid wasting money, an alternative that is well-developed, context-specific and capable, such as what has been generated here for educational domain, should be employed. Furthermore, the adaptive framework provided optimal results regarding the identification of comprehensive 'usability problem areas' on the educational websites, with minimal input in terms of cost and time spent in comparison with the employment of usability evaluation methods. The framework was used here to generate DSI, which helped to guide the evaluation process as well as reducing the time that it would have taken to identify these usability issues through current evaluation methods. In terms of the definition of missed problem given by [15], we can consider that the problems that were found by any one method and not found by the others as missed problems. From this standpoint, DSI missed discovering 15 real usability problems. However, HE and UT missed 64 and 61 real usability problems, respectively.

The above findings facilitate decision-making with regard to which of these methods to employ, either on its own or in combination with another, in order to identify usability problems on

educational websites. The selection of the method or methods will depend on the types of problem best identified by each of them.

8. CONCLUSION AND FUTURE WORK

Contrary to most of the efforts to construct and test enhanced usability methods, our work here has made explicit the process for so doing. The adaptive framework includes the views of users and usability experts to help generate a context-specific method for evaluating any chosen domain. The work presented here illustrates and evaluates this process for the generation of the DSI method to assess the usability of educational websites. DSI outperformed both HE and UT, even when taken together. This clearly represents a step in the right direction. Further validation of the use of our adaptive framework will indicate whether it is indeed applicable across domains. In order to consolidate and confirm the findings, future research could include testing the adaptive framework by developing DSI for different fields such as e-commerce and healthcare systems.

In conclusion, this research contributes to the advancement of knowledge in the field. Its first contribution is the building of an adaptive framework for generating a context-specific method for the evaluation of whichever system in any domain (Figure 1). The second contribution is the introduction of DSI that is specific for evaluating educational systems (Table 2). The third contribution is the identification of usability problem areas in the educational domain (five areas in Table 2).

9. ACKNOWLEDGEMENTS

10. REFERENCES

[1]   Abuzaid, R. (2010). Bridging the Gap between the E-Learning Environment and E-Resources: A case study in Saudi Arabia. *Procedia-Social and Behavioral Sciences*, 2(2): 1270-1275.

[2]   AcademicEarth, (2012), *AcademicEarth*, Accessed on 3/4/2012, Available at: [http://academicearth.org/]

[3]   Alias, N., Siraj, S., DeWitt, D., Attaran, M. & Nordin, A. B. (2013), Evaluation on the Usability of Physics Module in a Secondary School in Malaysia: Students' Retrospective. *The Malaysian Online Journal of Educational Technology*, 44.

[4]   Alkhattabi, M., Neagu, D. and Cullen, A. (2010). Information Quality Framework for E-Learning Systems. *Knowledge Management & E-Learning: An International Journal (KM&EL)*, 2(4): 340-362.

[5]   Alrobai, A. AlRoobaea, R. Al-Badi, A., Mayhew, P. (2012). Investigating the usability of e-catalogue systems: modified heuristics vs. user testing, *Journal of Technology Research*.

[6]   Alshamari, M. and Mayhew, P. (2008). Task design: Its impact on usability testing. In Internet and Web Applications and Services, 2008, ICIW'08. Third International Conference on, pages 583-589. IEEE**.**

[7]   Ardito, C., Costabile, M., De Angeli, A. and Lanzilotti, R. (2006). Systematic evaluation of e-learning systems: an experimental validation. In Proceedings of *The 4th Nordic Conference on Human-Computer Interaction: changing roles*, pp. 195-202. ACM.

[8] BBC KS3bitesize, (2012), *BBCKS3bitesize*, Accessed on 3/4/2012, Available at: [http://www.bbc.co.uk/schools/ks3bitesize/]

[9] Brooke, J. (1996). SUS - A quick and dirty usability scale. *Usability Evaluation in Industry*, pages 189-194.

[10] Chattratichart, J. & Lindgaard, G., (2007). Usability testing: what have we overlooked?. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1415-1424). ACM.

[11] Chattratichart, J. and Lindgaard, G. (2008), A comparative evaluation of heuristic-based usability inspection methods, In the proceeding of *CHI'08 extended abstracts on Human factors in computing systems*, 2213-2220

[12] Chattratichart, J. and Brodie, J. (2002), Extending the heuristic evaluation method through contextualisation. Proc. *HFES2002*, HFES (2002), 641-645.

[13] Chattratichart, J. & Brodie, J. (2004), Applying user testing data to UEM performance metrics. In *CHI'04 extended abstracts on Human factors in computing systems* (pp. 1119-1122). ACM.

[14] Chen, S. Y. and Macredie, R. D. (2005), The assessment of usability of electronic shopping: A heuristic evaluation*, International Journal of Information Management*, vol. 25 (6), pp. 516-532.

[15] Cockton, G. and Woolrych, A. (2002). Sale must end: should discount methods be cleared off HCI's shelves? *Interactions*, 9(5): 13-18. ACM.

[16] Coursaris, C. K. & Kim, D. J. (2011), A meta-analytical review of empirical mobile usability studies. *Journal of usability studies*, 6(3), 117-171.

[17] Dumas, J. and Redish, J. (1999). *A practical guide to usability testing*. Intellect Ltd.

[18] Doubleday, A., Ryan, M., Springett, M., & Sutcliffe, A. (1997). A comparison of usability techniques for evaluating design. In *Proceedings of the 2nd conference on Designing interactive systems: processes, practices, methods, and techniques* (pp. 101-110). ACM.

[19] Fernandez, A., Insfran, E. and Abrahão, S. (2011), Usability evaluation methods for the web: A systematic mapping study*, Information and Software Technology*.

[20] Gutwin, C. & Greenberg, S. (2000), The mechanics of collaboration: Developing low cost usability evaluation methods for shared workspaces. In *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2000. (WET ICE 2000). Proceedings. IEEE 9th International Workshops on* (pp. 98-103). IEEE.

[21] Hertzum, M. and Jacobsen, N. (2001). The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human-Computer Interaction*, 13(4): 421-443.

[22] Holsapple, C. W. and Wu, J. (2008), Building effective online game websites with knowledge-based trust, *Information Systems Frontiers*, vol. 10 (1), pp. 47-60.

[23] Holzinger, A. (2005), Usability engineering methods for software developers*, Communications of the ACM*, vol. 48 (1), pp. 71-74.

[24] Hart, J., Ridley, C., Taher, F., Sas, C. and Dix, A. (2008), Exploring the Facebook experience: a new approach to usability. In Proceedings of the *5th Nordic Conference on Human-Computer Interaction: Building Bridges*, pages 471-474. ACM.

[25] Hasan, L. (2009), Usability evaluation framework for e-commerce websites in developing countries.

[26] ISO (1998), *ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs): Part 11: Guidance on Usability*.

[27] Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. (1991, March). User interface evaluation in the real world: a comparison of four techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 119-124). ACM.

[28] Khajouei, R., Hasman, A. and Jaspers, M. (2011), Determination of the effectiveness of two methods for usability evaluation using a CPOE medication ordering system*, International Journal of Medical Informatics*, vol. 80 (5), pp. 341-350.

[29] Liljegren, E. (2006), Usability in a medical technology context assessment of methods for usability evaluation of medical equipment*, International Journal of Industrial Ergonomics*, vol. 36 (4), pp. 345-352.

[30] Liljegren, E., & Osvalder, A. L. (2004). Cognitive engineering methods as usability evaluation tools for medical equipment. *International Journal of Industrial Ergonomics*, *34*(1), 49-62.

[31] Ling, C. and Salvendy, G. (2005), Extension of heuristic evaluation method: a review and reappraisal*, Ergonomia IJE & HF*, vol. 27 (3), pp. 179-197.

[32] Law, L. and Hvannberg, E. (2002). Complementarily and convergence of heuristic evaluation and usability test: a case study of universal brokerage platform. In Proceedings of *The second Nordic conference on human-computer interaction*, pages 71-80, ACM.

[33] Magoulas, G. D., Chen, S. Y. and Papanikolaou, K. A. (2003), Integrating layered and heuristic evaluation for adaptive learning environments, In the proceeding of *UM2001*, 5-14.

[34] Molich, R. and Nielsen, J. (1990), Improving a human-computer dialogue, *Communications of the ACM*, vol. 33 (3), pp. 338-348.

[35] Muir, A., Shield, L. and Kukulska-Hulme, A. (2003), The pyramid of usability: A framework for quality course websites, In the proceeding of *EDEN 12th Annual Conference of the European Distance Education Network, The Quality Dialogue: Integrating Quality Cultures in Flexible, Distance and eLearning*, 188-194. Rhodes, Greece.

[36] Nielsen, J. (1992), Finding usability problems through heuristic evaluation, In the proceeding of *ACM CHI'92 Conference* 373-380. Monterey, CA, USA, May 3-7.

[37] Nielsen, J. (1994a), Heuristic evaluation, *Usability inspection methods*, vol. 24, pp. 413.

[38] Nielsen, J. (1994b), *Usability engineering*, Morgan Kaufmann.

[39] Nielsen, J. and Loranger, H. (2006), *Prioritizing web usability,* New Riders Press, Thousand Oaks, CA, USA.

[40] Nielsen, J. and Molich, R. (1990), Heuristic evaluation of user interfaces, In the proceeding of *SIGCHI conference on Human factors in computing systems: Empowering people*, 249-256.

[41] Nielsen, J. (2000), HOMERUN Heuristics for Commercial Websites, in www.useit.com.

[42] Nielsen, J. (2001), "Did poor usability kill e-commerce", in www.useit.com.

[43] Nayebi, F., Desharnais, J. M. & Abran, A. (2012), The state of the art of mobile application usability evaluation. In *Electrical & Computer Engineering (CCECE), 2012 25th IEEE Canadian Conference on* (pp. 1-4). IEEE.

[44] Oztekin, A., Kong, Z. J. and Uysal, O. (2010), UseLearn: A novel checklist and usability evaluation method for eLearning systems by criticality metric analysis, *International Journal of Industrial Ergonomics*, vol. 40 (4), pp. 455-469.

[45] Sears, A. (1997), Heuristic walkthroughs: Finding the problems without the noise, *International Journal of Human-Computer Interaction*, vol. 9 (3), pp. 213-234.

[46] Shackel, B. and Richardson, S. J. (1991), *Human factors for informatics usability*, Cambridge University Press.

[47] Skoool, (2012), *Skoool*, Accessed on 3/4/2012, Available at: [http://lgfl.skoool.co.uk//]

[48] Stracke, C. and Hildebrandt, B. (2007). Quality Development and Quality Standards in e-Learning: Adoption, Implementation and Adaptation. In Proceedings of *World Conference on Educational Multimedia, Hypermedia and Telecommunication 2007*, pp. 4158-4165.

[49] Thompson, A. and Kemp, E. (2009), Web 2.0: extending the framework for heuristic evaluation. In Proceedings of *The 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction*, pp. 29-36. ACM.

[50] Wilson, C. (2007). Taking usability practitioners to task. *Interactions*, 14(1): 48-49.

[51] Van den Haak, M., de Jong, M. and Schellens, P. (2004), Employing think-aloud protocols and constructive interaction to test the usability of online library catalogues: a methodological comparison. *Interacting with computers*, 16(6): 1153-1170.

[52] Zaharias, P. & Poylymenakou, A. (2009), Developing a usability evaluation method for e-learning applications: Beyond functional usability. *Intl. Journal of Human-Computer Interaction*, *25*(1), 75-98.

INSTRUCTIONS TO CONTRIBUTORS

The International Journal of Human Computer Interaction (IJHCI) publishes original research over the whole spectrum of work relevant to the theory and practice of innovative and interactive systems. The journal is inherently interdisciplinary, covering research in computing, artificial intelligence, psychology, linguistics, communication, design, engineering, and social organization, which is relevant to the design, analysis, evaluation and application of human computer interactive systems.

IJHCI addresses the cognitive, social, health, and ergonomic and other aspects of interactive computing including cognitive ergonomics, collaborative design, ergonomics & human factors, legal, ethical & social aspects of IT, human computer interaction, ergonomics, human computer interaction and user interface. IJHCI emphasizes the human element in relation to the systems in which humans function, operate, network, and communicate, including software, hardware, and their various contexts of use.

To build its International reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJHCI.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Started with Volume 4, 2013, IJHCI appears with more focused issues. Besides normal publications, IJHCI intend to organized special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

We are open to contributions, proposals for any topic as well as for editors and reviewers. We understand that it is through the effort of volunteers that CSC Journals continues to grow and flourish.

IJHCI LIST OF TOPICS
The realm of International Journal of Human Computer Interaction (IJHCI) extends, but not limited, to the following:

- Agent models Co-ordination and communication
- Agent-based Computing
- Computer Mediated Communication
- Computer Supported Cooperative Work
- Design and Evaluation of Innovative Interactive Sy
- Empirical Studies of User Behaviour
- Graphic Interaction
- Human language Technologies and Machine Learning i
- Human-Computer Interaction Theory
- Innovative Interaction Techniques
- Intelligent Tutoring, Coaching and Debugging Syste
- Intelligent user Interfaces
- Interaction in Mobile and Embedded Systems
- Interactive Decision Support Systems
- Interface Design and Evaluation Methodologies
- Knowledge Acquisition, Discovery, Modeling and Man
- Mixed and Augmented Reality
- Multimodal Interaction
- Natural language interaction
- Peer to Peer Communication Between

- Presence
- User Interface Prototyping for Interactive Systems

- Intelligent Sys
- Speech Interaction
- Virtual Reality

CALL FOR PAPERS

Volume: 4 - **Issue: 3**

i. Paper Submission: September 30, 2013   **ii. Author Notification:** October 31, 2013

iii. Issue Publication: November 2013

CONTACT INFORMATION