The Complementary Role of Two Evaluation Methods in the Usability and Accessibility Evaluation of a Non-Standard System

Funmi Adebesin
Meraka Institute of the CSIR and School of Computing, UNISA
P.O. Box 395
Pretoria, 0001, South Africa
+27 12 841 3373
fadebesin@csir.co.za

Paula Kotzé
Meraka Institute of the CSIR and School of Computing, UNISA
P.O. Box 395
Pretoria, 0001, South Africa
+27 12 841 4791
paula.kotze@meraka.org.za

Helene Gelderblom
School of Computing, UNISA
P.O. Box 392, UNISA, 0003, South Africa
+27 12 429 6631
geldejh@unisa.ac.za

ABSTRACT
Usability, which is generally defined in terms of application effectiveness, efficiency and user satisfaction, is one of the focus areas of human-computer interaction (HCI). Accessibility is the design of systems that can be perceived, understood and used by people with varying abilities. Although accessibility concerns are aimed at making systems usable for people with disabilities, support for direct accessibility, the built-in redundancies in an application that enable as many people as possible to utilize it without system modifications, is beneficial to people with or without disabilities. Different usability evaluation methods (UEMs) are available. Selecting between the various methods can be influenced by the type of system being evaluated. The Digital Doorway (DD), a non-standard computer system deployed to promote computer literacy amongst underprivileged communities in South Africa, was evaluated using the heuristic evaluation method and a field usability study. The heuristic evaluation method revealed a large number of usability and direct accessibility-related problems, some of which could be classified as low-severity problems. The field study showed additional problems that affected the successful completion of user tasks. Since a number of these were a direct consequence of the context of use, they were not recognized as problems by expert evaluators. The study showed that the heuristic evaluation method can be optimized by complementing it with another method that involves user participation and is, preferably, carried out in the intended context of use.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *evaluation/methodology, User-centered design*.

General Terms
Design, Human Factors.

SAICSIT '10, October 11–13, 2010, Bela Bela, South Africa.

Keywords
Accessibility, field evaluation, heuristic evaluation, usability.

1. INTRODUCTION
Usability, one of the focus areas of human-computer interaction (HCI), is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [18].

The success of any interactive system is dependent on, amongst other factors, its utility, ease of use from the user's perspective, and user experience [5; 6; 28]. It is therefore crucial that designers incorporate usability design principles early on in the design process, and evaluate systems for usability before deployment. When the target user group has special needs this becomes even more important.

Accessibility used to be a concern of the built environment practitioners where legislations required cuts into street curbs to enable easy access by people in wheelchairs. Now, the explosive growth of the World Wide Web (WWW) and the Internet has put pressure on designers to make electronic information accessible to people with disabilities [19; 32]. Accessibility in the context of HCI is defined as the design of applications that "are perceivable, operable and understandable for people with a wide range of abilities" [14].

Some authors consider accessibility to be a subset of usability [13]. Others [31] claim that it is a prerequisite for usability. However, what is incontrovertible is that both design concepts contribute to good design. The specified users in the definition of usability for any given application may be diverse, including those with disabilities. Incorporating usability and accessibility design guidelines will result in an application that is usable by a wider spectrum of users.

The primary focus of this paper is on usability evaluation although we also address the issue of accessibility evaluation by considering the direct accessibility properties of the system being evaluated. Direct accessibility refers to the built-in redundancies in applications which enable as many people as possible to utilize it without system modifications [39]. Direct accessibility can enhance general usability, offering benefits to people with or without disabilities.

Different approaches to usability evaluation exists, among them heuristic evaluation, field observation, surveys and controlled usability testing. Some of these methods require expert evaluators while others rely on end-user involvement [6; 26;

30]. Several previous studies compared usability evaluation methods (UEMs), for example, those by Gray and Saltzman [10] and Molich and Dumas [25].

Combinations of usability evaluation methods may be required to offset the limitations inherent in the use of any single evaluation method, for example, the heuristic walkthrough evaluation method [34]. The focus of this paper is on the complementary role that the combination of expert and user-based evaluation methods can play, especially when the target system is a non-standard system[1].

The Digital Doorway (DD) is an example of such a system. It is a walk-up and use system deployed amongst underprivileged communities around South Africa as part of the effort to narrow the digital divide. Ever since the installation of the first DD in 2002, the DD project has mainly focused on providing physical computer access to underprivileged communities around the country. The systems were deployed without any formal usability evaluation of the software applications installed on them.

DDs are housed in rugged, custom-designed kiosks with multiple terminals that can be accessed simultaneously by users. Each of the terminals has a metal keyboard with reinforced touchpad for input [12]. The robust housing and metal keyboard help to minimize vandalism. Pre-loaded software applications and contents run on the Ubuntu Linux operating system, however, the interface does not follow any particular design standard or operating system interface. The systems are installed at venues such as schools, police stations and community centers. The project has seen 206 DDs installed around the country since its inception. A three-terminal DD is shown in Figure 1.

Figure 1. A three-terminal Digital Doorway

http://www.digitaldoorway.org.za/index_main.php?do=hardware

Despite the fact that the project has focused on hardware development, the hardware does not currently support the use of assistive devices, such as screen readers for visually impaired users. Furthermore, the environment of use of the DD sometimes impose additional restrictions on the use of the system e.g. noise and glaring of the sun. It has thus become essential for us to also consider the level of direct accessibility support in the evaluation of the DD, in addition to evaluating the usability of software installed on the system.

This paper describes the use of multi-category heuristics or evaluation criteria (both terms used interchangeably) for evaluating a non-standard computer system and how the method was complemented with a field usability evaluation that retained the system's context of use. The results of the two methods were compared to underscore their complementary role.

The rest of the paper is structured as follows: Section 2 briefly describes some of the UEMs, highlighting their benefits and limitations. In Section 3, we discuss how the heuristic evaluation method was applied in the evaluation of the DD and the results obtained from the evaluation. Section 4 describes the field usability study on the DD. In Section 5, we compare the results obtained from the two studies and conclude the paper in Section 6.

2. OVERVIEW OF USABILITY EVALUATION METHODS

This section provides a general overview of UEMs, highlighting the benefits and limitations of each.

UEMs are generally classified into two main groups – expert analysis and user evaluation. The following subsections briefly describe the heuristic method, cognitive walkthrough, direct field observations, interviews, questionnaires, and controlled usability testing.

2.1 Heuristic Evaluation

The heuristic evaluation technique, pioneered by Nielsen and Molich in 1990 [27], involves expert evaluators independently critiquing an interface using a set of evaluation criteria in order to identify potential usability problems. The heuristic method is an easy, flexible and cost effective method that can be used in formative and summative evaluations. However, to be of any value, the set of evaluation criteria must be appropriate for the specific application and at least three expert evaluators are required [20; 27; 30].

2.2 Cognitive Walkthrough

This is an inspection method that is based on cognitive science theory where experts step through a set of tasks. Cognitive walkthrough aims to assess the learnability of systems where the preferred method of learning to use the system is by exploration rather than going through user manuals [4; 6; 30; 40]. It is a flexible method that can be used for formative and summative evaluation before user testing [40], but it does not address other measures of usability like the application efficiency. The method assumes that the evaluator possesses cognitive theory skills [40].

2.3 Direct Field Observation

This involves the evaluator observing the users as they carry out normal or routine activities in the natural context of use either at home or the workplace. Because the context is retained, direct field observation can reveal details that are difficult to obtain using other evaluation methods. This natural environment may sometimes disrupt the evaluation process as a result of high level of noise and constant interruptions from colleagues. In addition, participants may alter their behavior when they become aware of being observed [7].

2.4 Interviews and Questionnaires

These are query techniques that can be used to elicit users' requirements for a proposed system or measure the extent to which an implemented system meets their expectations. Interviews can be structured, semi-structured or a combination of both. Interviews, especially unstructured interviews, are

[1] Non-standard in this context means systems that do not display standard operating system interfaces or use standard equipment.

beneficial since the questions can be varied to allow the evaluator to probe issues as they arise to obtain deeper understanding [21; 30]. Interviews can be highly subjective as they are typically used to measure users' opinion. While it may be impossible to avoid participant subjectivity, it is essential to be aware of them.

Questionnaires are similar to interviews in that the questions can be closed or open, though they are not as flexible because the questions are fixed and further probing is often impossible. While questionnaires can be used to reach large number of respondents in less time compared to interviews, their distribution and return rates can be problematic. The potential for generating flawed data is high when using questionnaire as the sole evaluation method. Hence it may not adequately reflect the actual usability of the application [26]. This is why evaluation methods are typically combined to triangulate data [26; 30].

2.5 Controlled Usability Testing

Usability testing (UT) is an evaluation method where the performance of typical users is measured as they carry out real, pre-defined tasks using the target application. The aim of UT is to test the usability of the system, not the users' ability. UT is expensive, requiring sophisticated usability laboratory equipped with monitoring cameras and equipment [3; 30; 33]. To be effective, UT is typically combined with think-aloud, where users are encouraged to verbalize their thoughts and the reasoning behind their actions as they carry out the tasks [3].

Table 1 provides a summary of the evaluation methods discussed above, highlighting the advantages and limitations of each method.

Table 1. Summary of usability evaluation methods

Technique	Description	Advantages	Limitations
Heuristic evaluation	Experts independently assess the interface using a set of evaluation criteria.	- Flexible, can be used for formative or summative evaluation. - Ability to provide quick feedback to designers. - Can reveal large numbers of potential usability problems	- Sometimes require the development of application-specific evaluation criteria. - Requires multiple evaluators. - Some problems that may affect real users may be overlooked.
Cognitive walkthrough	Experts step through a set of tasks to assess learnability of the system.	- Flexible, can be used in formative and summative evaluation.	- Setting up of representative tasks can be tedious. - Other measures of usability are not addressed.
Direct field observation	Users are observed while carrying out tasks in a natural context of use.	- Natural context is retained. - Ability to reveal specific usability problems which may impact user tasks.	- High level of disruptions. - Participants may alter their behavior.
Interviews	Used to obtain user requirements for a new system and measure extent to which a functional system meet their expectations.	- Flexible, can be used for formative or summative evaluation	- Participants' subjectivity.
Questionnaires	Used to obtain user requirements for a new system and measure extent to which a functional system meet their expectations.	- Large number of respondents can be reached in short time.	- Participants' subjectivity. - Low return rate.
Usability testing	Real users are observed while carrying out pre-specified tasks in a controlled environment.	- Ability to reveal specific usability problems which may impact user tasks.	- Expensive, requires sophisticated equipment. - Some participants may find thinking aloud unnatural.

2.6 Choice of Evaluation Methods

The decision to select between the different evaluation methods is dependent on a number of factors, including the stage in the development life cycle at which evaluation is done (formative or summative), whether evaluation should be conducted in a controlled environment or natural setting and availability of resources [6]. For example, the heuristic evaluation method is flexible in that it can be used for formative or summative evaluation, provided appropriate evaluation criteria are used.

In this study, the main factor that influenced our choice of evaluation methods is the type of system being evaluated. Although we have access to a well-equipped usability testing laboratory, practical considerations make the use of controlled usability testing unfeasible since we could not physically move the DD to the usability laboratory. Furthermore, observation and logging software in the usability laboratory is only compatible with the Windows operating system. As stated in section 1, applications on the DD run on the Ubuntu Linux operating system. Because the study involved the evaluation of a fully functional system, the heuristic method is appropriate. To complement the heuristic evaluation method, a field usability evaluation was also done at a local school where the DD is installed.

3. HEURISTIC EVALUATION OF THE DIGITAL DOORWAY

In this section, we describe the application of the heuristic evaluation method in the evaluation of the DD.

The DD provides access to a large number of software applications and other resources, the majority of which are open

source or third-party applications. These includes the OpenOffice suites, educational games, scientific simulations, Wikipedia documents and Mindset applications, the latter being a South African curriculum-based educational program [11]. The content can be accessed by logging in as a guest (using the username 'dd1' for example) or as a registered user. A new user account is created by completing a simple electronic form, which is activated by typing 'new' in the username field.

Because it is impractical to evaluate all the applications on the DD, a selection of interfaces and applications developed in-house were evaluated in this study. These are: the login screen, the new user account registration form, the main desktop, and three educational games – *What-What Mzansi* (a quiz game), *OpenSpell* (an educational spelling game), and *Themba's Journey* (for developing life-skills).

3.1 Developing the Heuristics

As stated in Sections 2.1 and 2.6, the heuristic evaluation method can be used in formative or summative evaluation provided the evaluation criteria are suitable for the system being evaluated. To come up with evaluation criteria that provide adequate coverage of interfaces and applications to be evaluated, we used contextual analysis to examine the principles for usable interface design by Dix, Finlay, Abowd, and Beale [6], Gelderblom's [9] guidelines for the design of children' technology, Mayhew's guidelines for the design of form-fill interfaces [24], Nielsen's heuristics [27], the usability principles by Preece, Rogers and Sharp [30], the design principles by Norman [29], and Shneiderman's [36] golden rules for interface design, for their applicability to the DD system.

To address direct accessibility concerns, we studied the seven universal design principles by Story, Mueller and Mace [37], the web content accessibility guidelines (WCAG 1.0) by world wide web consortium (W3C) [41] (WCAG 2.0 was still in draft form when the heuristics were derived), the United States' standards for electronic information accessibility (Section 508) [38], and the IBM software accessibility checklists [17].

Because the applications evaluated included educational games, there was a need to study guidelines relating specifically to such applications. Game-specific guidelines studied are those proposed by Alessi and Trollip [2], Malone [22; 23] and Shelley [35].

Evaluation criteria were derived systematically over three iterations. During the first iteration, the first author of this paper tested the generated heuristics on the DD to assess their adequacy. The heuristics were then modified in another round of iteration and tested by the third author on the DD again. Further modifications were then made before producing a final set of evaluation criteria for the DD.

Not all the principles and guidelines examined were equally applicable to the DD. Examples of non-applicable principles are those relating to multithreading, task migratability and the use of markups and style sheets [22].

The process of refinement yielded a total of 77 heuristics. Page restrictions do not permit the inclusion of the complete heuristic set, but the interested reader can access it in [1]. A representative subset of the heuristics is provided in Table 2.

To aid readability and facilitate the analysis of identified problems, the criteria were organized into four categories, namely general usability, form-filling, direct accessibility, and game heuristics.

Table 2. Subset of evaluation criteria for the Digital Doorway

Category 1: General Usability Heuristics	
1.1	Provide information that will enable users understand how to interact with the DD using clear and simple terminologies.
1.2	Provide clear indication of what the next required action is.
1.3	Icons, labels and symbols should be intuitive and meaningful to users, taking into account user context and experience.
1.4	Follow and adhere to platform and industry standards and conventions.
1.5	Be consistent in the naming conventions used for icons, symbols, and objects.
1.6	Objects, options, and permissible actions should be visible so that users do not have to remember instructions.
1.7	Audio instructions should be given close to when the user is expected to act on them.
1.8	Feedback should be provided in clear and unambiguous terms.
1.9	Response to user action by the system should be instantaneous. Where this is not possible, the system should indicate that the task is in progress to avoid repeated clicking by user.
1.10	Prevent user error by using appropriate constraints at strategic points.
1.11	Error messages should be context-specific in relation to the action performed.
1.12	Ensure that information sequence and layout appear in natural and logical order.
1.13	The DD should not impose unnecessary constraints on user input method.
Category 2: Form-Filling Heuristics	
2.1	Provide visible cue by positioning the cursor in the first data field at start of the form.
2.2	Ensure that related items are grouped together to aid readability.
2.3	Provide visual reinforcement for element groups through efficient use of white spaces and borders.
2.4	Designate required fields in standard and consistent ways taking into account users' age and experience.
2.5	When input errors are detected, the cursor should be positioned in the error field with the field highlighted to attract user's attention.
Category 3: Heuristics to support direct accessibility	
3.1	Text size of instructions should be large enough to enable easy perception by users with low vision.
3.2	Provide audio equivalent of instructions and information to afford access by users who cannot read.
3.3	Provide quality speech output that enable users hear and comprehend their meanings.

3.4	Provide feedback using multiple modes to facilitate access and comprehension.
3.5	Ensure that colour alone is not used to represent important information.
3.6	Ensure that background and text colours contrast well with each other.
3.7	Allow keyboard navigation for operations/tasks that do not essentially require use of the mouse.
3.8	Provide controls that enable users to pause, continue, or repeat audio-visual information.
3.9	Information should be accessible without undue physical efforts.
Category 4: Game-specific Heuristics	
4.1	Games should have clear goals and objectives.
4.2	Provide an easily accessible instruction on how to play the game.
4.3	Permissible actions and constraints should be clearly specified.
4.4	Users should be able to adjust the game's level of difficulty.
4.5	Performance feedback should not be given using negative or sarcastic statements.
4.6	Provide constructive and corrective feedback that will enable player learn from mistakes and improve future performance.
4.7	All control mechanisms should be visible and easily accessible.
4.8	Provide clear exit route to enable users leave the game at any stage.
4.9	Game should be accessible in different languages.

3.2 The Evaluation Process

The DD was voluntarily and independently evaluated by five usability/accessibility experts using the generated heuristics, part of which is provided in Table 2. Three of the experts have experience in usability evaluation while the other two have expertise in usability and accessibility evaluation. The evaluators were provided with the evaluation criteria, the procedure to be followed, and an overview of interfaces and applications to be evaluated, well in advance so they could familiarize themselves with the relevant documentation. Evaluators also signed informed consent forms that guarantee their anonymity and the confidentiality of information they provided.

For practical reasons (discussed in Section 2.6), evaluators had to go physically to the DD laboratory located at the CSIR for the evaluation. The five evaluation sessions were conducted over a 3-month period.

Using Nielsen's [27] procedure for conducting heuristic evaluation with some modifications the evaluators went through a two-pass session, with the first pass aimed at getting a feel of how the applications work. In the second pass they did the evaluation. Each session lasted approximately 2 hours although we did not set any time limit. The sessions were facilitated by one of the authors who acted as scribe.

Following the evaluation, an evaluation report was compiled by the scribe and mailed to the relevant evaluator within two days of the evaluation. The report was then verified by the evaluator to ascertain that it is a true reflection of the evaluation. In some cases, the verification resulted in modifications to the report.

Taking notes and compiling evaluation reports allowed us to make judicious use of evaluators' time and reduced the associated difficulty of note taking and evaluation while standing. Thus the evaluators were able to focus on the evaluation process.

3.3 Evaluation Results

Between the five evaluators, an aggregate of 71 usability and/or accessibility problems were identified. Thirty-four problems were identified by single evaluators (47.88% of the total problem set) while ten (14.08%) were recognized as such by two evaluators. Seven (9.85%) problems were identified by three evaluators, and a further seven (9.85%) were identified by four evaluators. All the five evaluators agreed that 13 (18.30%) of the total problem set were problems.

Within the login screen, a total of eight problems were identified by evaluators, four were located in the new user account registration form, and 13 on the main desktop. For the educational game applications, 17 problems were located in *What-What Mzansi*, 14 in *OpenSpell* and 15 in *Themba's Journey*. The number of problems located in each interface and application is shown in Figure 1.

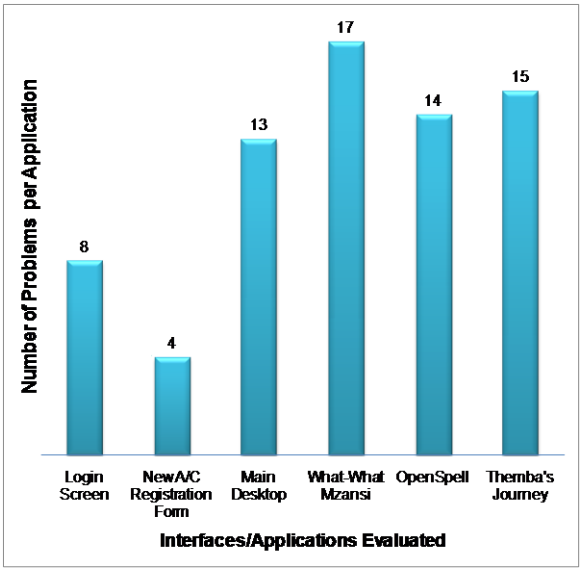

Figure 1. Number of problems identified in each interface and application

Due to page restrictions, the complete list of problem descriptions (available in [1]) could not be presented here. In section 5 we provide a subset of identified problems and compare them with actual problems experienced by users during field evaluation.

4. FIELD USABILITY EVALUATION OF THE DIGITAL DOORWAY

Designers typically make assumptions about potential users [8]. Evaluation with real users enables evaluators to assess the extent to which those assumptions are valid.

In this section, we discuss the field usability evaluation conducted to evaluate the DD in one of the centers where it is deployed.

4.1 The Evaluation Environment

We evaluated the DD at a local school where the system is installed in an open area on one of the school's corridors. This provides unrestricted access to the system. Children from surrounding homes also have access to the DD soon after the school closes until 18:00 late in the afternoons.

Prior to conducting the study, a formal approval was obtained from the school principal. Parents/guardians of participants also signed informed consent forms.

Six learners participated in the evaluation, with two participants each using one of the educational games *What-What Mzansi*, *OpenSpell* or *Themba's Journey*. Participants were also required to register a new user account before accessing the applications unless they had a valid account. Participants with disabilities were not included in the evaluation, since the DD does not support the use of assistive devices such as a screen reader. Hence, it will be pointless to include a blind user, for example.

4.2 The Evaluation Process

Rather than use the conventional field study where users are observed while using the system, participants were given pre-defined tasks (shown in Figure 2) to complete. This enabled us to focus the evaluation on the specific interfaces and applications identified for the evaluation.

Digital Doorway Evaluation -Task list

1. Read the screen instruction on how to register as a new Digital Doorway user if you are not a registered user.
2. Complete the registration form if you are not a registered user, otherwise proceed to step 3.
3. Start the Digital Doorway by providing the requested information.
4. Search for the quiz game '**What-What Mzansi'**.
5. Remember to provide verbal feedback all the time.
6. Search for and read the instruction on how to play the game.
7. Proceed to play the quiz game.
8. Choose how difficult you want the game to be.
9. Change the volume to suit your need.
10. Remember to provide verbal feedback all the time.
11. Close the Digital Doorway when you are done.

Figure 2. Task list for field usability evaluation (using the application 'What-What Mzansi)

To avoid disruptions to learning activities as much as possible, and minimize distractions from noise, evaluation sessions took place in the afternoons well after the official closing hour of the school.

Using the cooperative evaluation style, participants were encouraged to ask questions and assistance whenever they got stuck with any activity. This approach is justified since the DD is not a transaction processing system where the speed of task completion is a measure of usability.

As a secondary instrument to triangulate data, each participant was given a semi-structured questionnaire to evaluate the DD after the evaluation sessions. The questions in the questionnaires were based on a selection of the derived heuristics, discussed in section 3, although they were rephrased using simpler terminology to aid comprehension by novices.

4.3 Evaluation Results

Various usability and direct accessibility-related problems were encountered by users during the evaluation. These ranged from inability to locate the required application, to lost data due to lack of error tolerance by the system.

The total number of actual user problems experienced by the six participants was 29, as illustrated in Figure 3. Eight of these were additional problems not recognized by expert evaluators while the others formed part of the problems identified by the experts. Three problems were related to the login screen, six were in the new user account registration form while five involved the main desktop. In the educational game applications, three problems related to the quiz game, *What-What Mzansi*, six problems involved *OpenSpell* while a further six related to *Themba's Journey*.

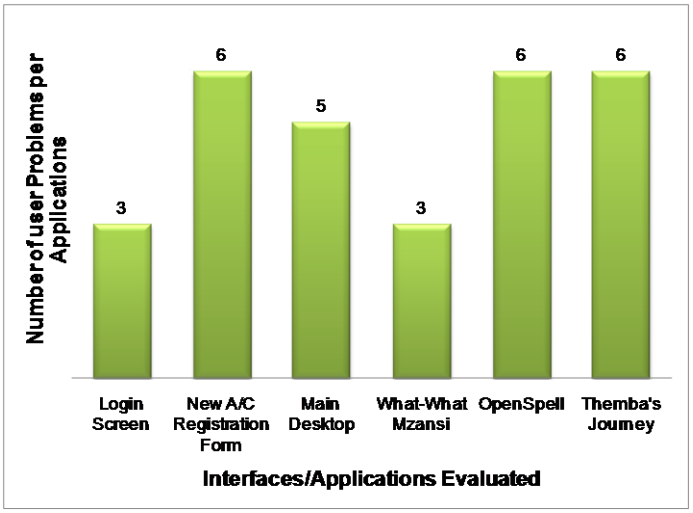

Figure 3. Number of actual user problems per interface/application

Only 4 of the 6 questionnaires were returned. Analysis of the questionnaires showed discrepancies between participants' responses and their actual behavior during observations. For example, in response to the statement, *I am able to determine the meaning and purpose of signs and symbols used in the Digital Doorway*, two participants responded with 'Agree'. However, one of these participants was unable to locate the volume control and exit buttons on the desktop without assistance while the other could not set the level of difficulty in OpenSpell (designated with *s).

Although the questionnaire method was not a primary method of evaluation in the study, the responses from participants highlighted some of the problems associated with the use of questionnaires as mentioned in Table 1. In Section 5, we provide the usability and direct accessibility-related problems found during the field evaluation and compare them with those identified by expert evaluators.

5. COMPARISON OF RESULTS FROM HEURISTIC AND FIELD USABILITY EVALUATIONS

As discussed in section 2, each evaluation method has benefits and limitations. Combining an expert evaluation method like heuristic evaluation with one that involves user participation, especially in a natural environment of use, allows one method to offset the shortcomings in the other, for example by revealing problems that were not picked up while using the other method.

Table 3 provides the description of the problems identified by expert evaluators and compare them with actual user problems during field evaluation to emphasize this complementary role.

Table 3. Problems identified by expert evaluators versus actual user problems

Heuristics violated	Problems identified by expert evaluators	Actual user problems
Login Screen		
1.8	There was no feedback whatsoever when an incorrect username and/or password was provided.	Incorrect username and/or password were a common user error. The system returned the same screen over and over without an indication of what the mistake was. Most of the time the field observation facilitator had to inform participants that the problem was with the username or password they are providing.
1.1	The login instruction was quite confusing, not sure how to handle the choice between creating a new user account and using the system as a guest.	The system did not provide login information for users who have just registered or those with existing accounts. The only prompt on the screen is addressed to guest users and those wanting to create new user accounts. Some participants actually typed in 'new' or 'dd1' to login, while others asked what should be done next before being told that the newly chosen username and password should be used to log in.
1.1, 1.12	Instructions for creating new user account and that for a guest log in are lumped together in the same text box.	None.
1.3	None	Users sometimes confused their surname with a 'username' and type their surname in the username field, though this was not the chosen username.
3.1	The font sizes of the instructions on how to login/create user account in four other languages, namely Xhosa, Afrikaans, Sotho and Venda were small.	None. Participants in the field evaluation only read the English version of screen instructions. The font size of instructions in the other languages could have been a problem if the participants did not understand English.
1.2	After entering the username, there was no indication of what to do next.	User problems related to this involved the hardware i.e. the keyboard. Some participants confused the <Enter> key on the keyboard with the key designated for producing a 'mouse click' effect because the keys were not labelled. However, after pressing one without the desired effect they then pressed the other.
New User Account Registration Form		
2.1	When the form application is activated, the cursor is not positioned in the first data field. The user is required to place the cursor in the first field	Some participants began typing their names only to realise later that the input was not being accepted and needed to place the insertion point within the first field before typing again.
2.4	There is no indication of which fields are compulsory to be filled and which ones are optional	Participants typically kept the home language and preferred language fields empty only to have error messages urging them to fill the fields.
3.7	User cannot use the <Tab> key on the keyboard to select female for the gender field	None of the participants used the <Tab> key to navigate the form. They positioned the insertion point over relevant fields before typing or clicking to make their selections. This was to be expected since they are not expert users.
1.10	None	Two participants accidentally clicked on the <Cancel> button. This inadvertent user error resulted in the form being closed without any warning to the user thereby erasing all the data fields input thus far.
2.5	None	The form did not facilitate the location of an error field. A participant erased his input in the password field accidentally, while trying to correct the name field entry following an error message. The insertion point remained in the password field after clicking on the <Register user> button. Without the user realising this, he pressed the backspace button several times and erased the wrong field.
Main Desktop		

1.1, 1.3	The functionality of the volume control slider is not clear from its look.	Only three of the six participants were able to locate the volume control button on the desktop, the other three required assistance after several failed attempts.
1.3	One folder on the desktop has the caption 'new_content'. This is not descriptive of the applications it contains. The game applications *What-What Mzansi*, *Themba's journey* and *OpenSpell* are hidden inside the folder 'new_content'.	Only two of the participants found the location of the game applications on their own. Other participants unsuccessfully searched for the applications within the <Game> submenu, located in the Resource menu, before they were told where to find them.
3.6	The level of contrast between the dark blue background and the grey foreground used to label icons is low. The contrast between the word 'Digital Doorway' and the dark blue background is poor.	Three of the participants found the background colour to be too dark. On several occasions, they had to shield their faces and the screen with their hands while using the DD to overcome the extent of reflection of the sun on the dark background. The dark background was significantly worse than that experienced in the close-up laboratory used by expert evaluators. The reflection worsened the contrast issue.
1.5	The caption of the element labeled 'Bluetooth_ saver' has the first character capitalized while the captions for all other icons and folders are in lowercase.	None
1.10	The locations of the following icons on the taskbar are too close to one another: the right-pointing arrow button ⇒, <System> <Volume control>, and Volume control slider. Users could easily click on the right-pointing ⇒ button while trying to use the volume control slider, thereby closing the system unintentionally.	A user accidentally clicked on the ⇒ button used to exit the system while trying to locate the volume control button and the system was shut down without any warning.
What-What Mzansi		
1.6	At the start of the application, some of the control buttons and the character that reads out instructions and questions are hidden from user's view. A full screen mode is activated by clicking on an icon which does not indicate this function.	None of the two participants who used this application knew how to get a full screen view of the game.
1.6	None	One of the terminals (the third terminal) used for the evaluation sessions had unusually large icons. This resulted in non-visibility of a number of control buttons, in this particular instance, a right pointing arrow '>' used for forward progression. This made it impossible for participants to repeat the level which they had just completed as required following poor performance.
4.5	The performance feedback 'don't make me laugh' after a poor performance is cheeky and not encouraging. Some users might find it offensive.	The two participants who used this application were indifferent to the performance feedback.
OpenSpell		
1.3	The use of the labels <Say>, <Guess>, and <Spell> are not descriptive of their functionalities.	The two participants who used this program selected the <spell> menu option when asked to learn the spelling of a few words. However, this functionality is provided within <say> menu option.
1.13, 3.7, 4.3	When the <Spell> option is selected, the user cannot use the keyboard to provide input but must use onscreen keyboard. There is no instruction stipulating this restriction.	When asked to do some spelling exercises, both participants first attempted to use the keyboard to provide their input, only to realise later that they can only use the onscreen keyboard.
1.3	The * symbols used to represent the level of difficulty are not intuitive	Only one of the two participants was able to associate the * symbols with the level of difficulty. The other user did not know how to set the difficulty level.
Themba's Journey		
3.2, 4.9	The narration voice is only in Xhosa. Non-Xhosa users who cannot read cannot use the application.	None
3.9, 4.9	To access an English version, the user must hover the mouse on the speech bubble. This can be	Both participants did not know how to get the English version of the application until they were told. Much effort was required by participants

	problematic for users with limited use of their hands. The information provided under <Help> did not specify this.	to move pointer around the speech bubbles in order to read English versions.
3.6	None of the expert evaluator identified the dark background of Themba's Journey as potential problem. This is due to the close-up environment where the evaluation was conducted.	Application background was very dark. Both participants had to shield their faces and screen with hands. The dark background was made worse because the DD is located in an open space with excessive natural lighting and glaring sun.

The two problem sets were also matched with the evaluation heuristics in Table 2.

The results presented in Table 3 illustrate the value of using a combination of heuristic evaluation and field usability evaluation methods. The two methods each yielded errors that were overlooked in the other method. The heuristic evaluation method highlights what experts believe could constitute potential problems for users while the field evaluation revealed actual problems that impacted on users' tasks. There was also high variation in the nature of the usability/accessibility problems recognized by evaluators, with 48% of problems reported by single evaluators. The differences in the type of problems identified by evaluators has been termed the evaluator effect by authors like Horbaek and Frojaer [16] and Hertzum and Jacobsen [15].

We also found that the heuristic evaluation method in general, yielded the kind of usability errors that can be seen as predictable; evaluation with real users produced additional errors that are generally unpredictable. An example of this occurred in the use of the terms 'username' and 'surname', where the experts could distinguish between the two, whilst some users confused the two.

The heuristic evaluation method identified a large number of usability and/or accessibility problems (71), some of which were low-severity problems that may not necessarily affect users' tasks. Examples of these includes the lack of consistency in the use of upper and lower case letters for interface elements captions and the layout of log-in instructions. Eight additional problems which surfaced in the field study were not recognized by expert evaluators. For example, the low support for users to identify and locate the fields in the registration form that caused error messages. This resulted in an unintended deletion of input data.

Although some of the concerns raised by expert evaluators were inconsequential for the participants in this study (e.g. the small instruction font size and the sarcastic performance feedback), they could constitute problems for other users. Attention to as many of the potential problems as possible will improve general usability for a wider spectrum of users.

The use of questionnaires as a secondary instrument during the field study to triangulate data showed the kind of inconsistencies that may arise between what users say and what they actually do. For example, responding that they understood the meaning of symbols and icons when in reality they were unable to determine their purpose.

One of the shortcomings of the heuristic evaluation method is that of overlooking real user problems. A field usability evaluation on the other hand may not necessarily reveal all the problems that potential users may encounter due to the relatively small number of participants and their level of expertise in the particular study. By combining the heuristic evaluation method with a field usability evaluation, we were able to address these limitations. Furthermore, the kind of disruptions that is characteristic of field evaluation settings was minimized by scheduling the evaluations after the official closing hour of the school.

Overall, both the heuristic evaluation method, using a set of heuristics specifically derived to suit such environment, and the field usability evaluation methods were found to be appropriate methods for assessing the usability of a non-standard interactive system.

6. CONCLUSION

In this paper we presented the use of multi-category heuristics in the evaluation of a non-standard system, the DD. We described how the heuristic evaluation method was complemented with field usability evaluation at a local school where the real context of the system's use is retained. We showed that despite the limitations inherent in different evaluation methods, using a combination of evaluation methods can offset such limitations.

Future research will involve field evaluation with a larger and more diverse user group to assess the extent to which the problems that were identified by expert evaluators, and which were not experienced by participants in this study, are reflected in the use of the DD by other user groups.

7. ACKNOWLEDGMENT

We thank the DD team at Meraka Institute for allowing us to evaluate the system. We also wish to express our gratitude to the expert evaluators for making time out of their busy schedule to evaluate the DD, as well as the learners who participated in the field usability study.

8. REFERENCES

[1] Adebesin, T. F. 2010. Report on the Usability and Accessibility Evaluation of the Digital Doorway. Available from http://hufee.meraka.org.za/Hufeesite/links/files/Technical%20Report%20-%20Digital%20Doorway%20Evaluation-Protected.pdf/view

[2] Alessi, S. M., & Trollip, S. R. 2001. Multimedia for Learning: Methods and Development (3rd ed.). Massachusetts: Allyn & Bacon.

[3] Barnum, C. M. 2002. Usability Testing and Research. Massachusetts: Allyn & Bacon.

[4] Cockton, G., Woolrych, A., & Lavery, D. 2008. Inspection-Based Evaluations. In A. Sears & J. A. Jacko (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (2nd ed.). New York: Lawrence Erlbaum Associates.

[5] Davis, F. D. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13(3), 319-340.

[6] Dix, A., Finlay, J., Abowd, G. D., & Beale, R. 2004. Human-Computer Interaction (3rd ed.): Pearson Education Ltd.

[7] Dumas, J. S. 2003. User-Based Evaluation. In J. A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook*. Mahwah: Lawrence Erlbaum Associates.

[8] Gardner-Bonneau, D. 2010. Is Technology Becoming More Usable -or Less - and With What Consequences. Journal of Usability Studies, 5(2), 46-49.

[9] Gelderblom, J. H. 2008. Designing Technology for Young Children: Guidelines Grounded in a Literature Investigation on Child Development and Children's Technology. PhD Thesis, UNISA. Retrieved from http://etd.unisa.ac.za/ETD-db/theses/available/etd-09172008-132111/unrestricted/thesis.pdf

[10] Gray, W. D., & Salzman, M. C. 1998. Damaged Merchandise? A Review of Experiments That Compare Usability Evaluation Methods. Human-Computer Interaction, 13, 203-261.

[11] Gush, K., Cambridge, G., & Smith, R. 2004. The Digital Doorway - minimally invasive education in Africa. Paper presented at the Proceedings of the ICT in Education Conference, Cape Town.

[12] Gush, K., De Villiers, M. R., Smith, R., & Cambridge, G. In Press. Digital Doorways. In J. Steyn, J. Belle & E. M. Villeneuva (Eds.), *Development Informatics and Regional Information Technologies: Theory, Practice and the Digital Divide* (Vol. 2). Pennsylvania: IGI.

[13] Henry, S. L. 2002. Understanding Web Accessibility. In J. Thatcher, P. Bohman, M. R. Burks, S. L. Henry, B. Regan, S. Swierenga, M. D. Urban & C. D. Waddell (Eds.), *Constructing Accessibility Web Sites*. Birmingham: Glasshaus.

[14] Henry, S. L. 2007. Just Ask: Integrating Accessibility Throughout Design. Retrieved from http://www.uiaccess.com/JustAsk/

[15] Hertzum, M., & Jacobsen, N. E. 2003. The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. International Journal of Human-Computer Interaction, 15(1), 183-204.

[16] Hornbaek, K., & Frokjaer, E. 2008. A Study of the Evaluator Effect in Usability Testing. Human Computer Interaction, 23(3), 251-277.

[17] IBM. 2009. IBM Software Accessibility Checklist Retrieved 24 November 2009, from http://www-03.ibm.com/able/guidelines/software/accesssoftware.html

[18] International Organization for Standardization. 1998. ISO 9241-11 Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) - Part 11: Guidance on Usability.

[19] Iwarsson, S., & Stahl, A. 2003. Accessibility, Usability and Universal Design - Positioning and Definition of Concepts Describing Person-Environment Relationships. Disability & Rehabilitation, 25(2), 57-66.

[20] Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. 1991. User Interface Evaluation in the Real World: A Comparison of Four Techniques. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

[21] Lindgaard, G. 1994. Usability Testing and System Evaluation: A Guide for Designing Useful Computer Systems. London: Chapman & Hall Computing.

[22] Malone, T. W. 1980. What Makes Things Fun to Learn? Heuristics for Designing Instructional Computer Games. Paper presented at the 3rd ACM SIGSMALL Symposium and 1st SIGPC Symposium on Small Systems.

[23] Malone, T. W. 1981. Toward a Theory of Intrinsically Motivating Instruction. Cognitive Science, 5(4), 333-369.

[24] Mayhew, D. J. 1992. Principles and Guidelines in Software User Interface Design. New Jersey: Prentice Hall, Inc.

[25] Molich, R., & Dumas, J. S. 2008. Comparative Usability Evaluation (CUE-4). Behaviour & Information Technology, 27(3), 263-281.

[26] Nielsen, J. 1993. Usability Engineering. Boston: Academic Press, Inc.

[27] Nielsen, J. 1994. Heuristic Evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability Inspection Methods*. New York: John Wiley & Sons, Inc.

[28] Nielsen, J. 2003. Usability 101: Introduction to Usability Retrieved 20 July 2009, from http://www.useit.com/alertbox/20030825.html

[29] Norman, D. A. 2001. The Design of Everyday Things. London: MIT Press.

[30] Preece, J., Rogers, Y., & Sharp, H. 2007. Interaction Design: Beyond Human-Computer Interaction (2nd ed.). Chichester: John Wiley & Sons Ltd.

[31] Pühretmair, F., & Miesenberger. 2005. Making sense of Accessibility in IT Design – Usable Accessibility vs. Accessible Usability. Paper presented at the Database and Expert Systems Applications (DEXA'05).

[32] Rogoff, R. 2001, 24-27 Oct 2001. Making Electronic Information Accessible to Everyone. Paper presented at the Professional Communication Conference, (IPCC 2001).

[33] Rubin, J. 1994. Handbook of Usability Testing: How to plan, design, and conduct effective tests. New York: John Wiley & Sons.

[34] Sears, A. 1997. Heuristic Walkthroughs: Finding the Problems Without the Noise. International Journal of Human-Computer Interaction, 9(3), 213-234.

[35] Shelley, B. 2001. Guidelines for Developing Successful Games Retrieved 10 March 2010, from http://jnoodle.com/careertech/files/GuidelinesDevelopingSuccessfulGames.pdf

[36] Shneiderman, B. 1998. Designing the User Interface: strategies for Effective Human-Computer Interaction (3rd ed.). New York: Addison Wesley.

[37] Story, M. F., Mueller, J. L., & Mace, R. L. 1998. The Universal Design File: Designing for People of All Ages and Abilities Retrieved 17 November 2009, from http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/19/ac/11.pdf

[38] United States Access Board. 2000. Electronic and Information Technology Accessibility Standards (Section 508) Retrieved 09 November 2009, from http://www.access-board.gov/sec508/standards.htm

[39] Vanderheiden, G. C. 1994. Application Software Design Guidelines: Increasing the Accessibility of Application Software to People with Disabilities and Older Users Retrieved 04 November 2009, from http://trace.wisc.edu/docs/software_guidelines/software.htm

[40] Wharton, C., Braffort, J., Jeffries, R., & Franzke, M. 1992. Applying Cognitive Walkthroughs to More Complex User Interfaces: Experiences, Issues, and Recommendations. Paper presented at the SIGCHI Conference on Human factors in Computing Systems.

[41] World Wide Web Consortium. 1999. Web Content Accessibility Guidelines 1.0 Retrieved 7 August 2009, from http://www.w3.org/TR/WAI-WEBCONTENT/