**Details of how reviewers’ concerns have been addressed**

|  |  |  |
| --- | --- | --- |
| **Reviewer B** | | |
|  | **Reviewer’s concerns** | **How they were addressed** |
| 1 | The reviewer was concerned with the large number of heuristics used for the HE (n=77). The reviewer mentioned HE is a discount UE method, and consequently, the number of heuristics should be fairly short. For example, Nielsen's heuristics comprise 10 heuristics and Shneiderman's heuristics comprise 8 heuristics. The heuristics proposed read more like design guidelines than heuristics. | The reviewer has a fairly limited view of heuristic evaluation and the extent to which it has developed since Nielsen published his heuristics. Application specific heuristics, also aimed at the use by non-usability experts, are always very large in quantity (even Nielsen advocates it now). For example, Travis’ 247 web usability guidelines, Xerox’s hundreds of heuristic to support Nielsen’s original heuristics and others, Sim et al.’s CAA heuristics, etc.  Even before Nielsen, large numbers of application specific heuristics were developed (e.g. Smith and Mosier).  We have updated the arguments in the paper to make this clearer. |
| 2 | Another concern is that HE is a task-based UE method, but no mention is made of which tasks were used for the HE. Were these the same tasks used for the field study or not? | The tasks used were explicitly stated in the revised paper. |
| 3 | The reviewer was concerned that the field study comprised 9 participants which is a very small number for a summative evaluation. For statistical purposes, at least 30 participants should have been used. | If the field study were used on its own and for the purpose of a summative evaluation of the DD software, the concern of the reviewer would have been valid. But it was not the case. The purpose of the evaluation was to test the applicability and coverage of the set of multi-category heuristics developed and not to be a summative evaluation of the DD.  We have strengthened the arguments in the paper to make the distinction clear. |
| 4 | A further concern was that the proposed heuristics were used as a basis for the user satisfaction questionnaire, rather than a standard user satisfaction questionnaire. As a result, the questionnaire was very long and only 6 participants completed the questionnaires. The terminology used in this questionnaire was appropriate for experts, but not for the participants who consisted of high school pupils. The results of a user study are only as reliable as the instruments used. | Only a sub-section of heuristics was adapted for the questionnaire. The language use was also adapted to be more appropriate to the user population in the field study. As stated in the paper the results of the questionnaires were never used as user satisfaction questionnaire or as a usability yardstick for the DD, and it is also not recommended for future use by the DD team on its own – due to the inherent problems many researchers found with the method; also confirmed in our study. |
| 5 | The conclusions make several statements relating to the usefulness of triangulating UE methods, but this is not new or original. The paper thus does not contribute any new knowledge, but merely comprises a usability evaluation of the DD, which is not research. | The use of triangulation was never claimed to be original – the paper just again confirmed the value of the use of triangulation.  The main purpose of the study was the development of the multi-category heuristics – this clearly constituted a major research effort. The actual usability evaluation of the DD was a by-product since it was only used to illustrate the applicability and coverage of the set of heuristic developed, and is not the main contribution of the paper.  Following the comments of the reviewer, we realised that we had to make sure that this fact is clearly stated throughout the paper and we therefore revised the argument of the paper throughout accordingly. |
| 6 | The paper also contains several editorial mistakes and the text in Figures 9 and 10 is too small to read. Several sections contain repetitions of earlier statements (for example see Section 5.2 and Section 4.2.1.1). | These have been attended to. |

|  |  |  |
| --- | --- | --- |
| **Reviewer C** | | |
|  | **Reviewer’s concerns** | **How they were addressed** |
| 1 | The reviewer was concerned about the lack of emphasis on accessibility evaluations, even though it is in the title and suggests that we should reconsider if accessibility should be in the title. | The reviewer is correct in noting that the primary focus of the study was not accessibility. We have changed the wording to make it more explicit. |
| 2 | The reviewer suggests we pay attention to flow between paragraphs throughout the paper | The flow of arguments between paragraphs have been revised and improved. |
| 3 | The reviewer suggests we use another picture shows that the DD indeed has multiple terminals | We have changed the picture that shows multiple users around the DD terminals |
| 4 | The reviewer suggests we use a better word for “resources” in the first paragraph of section 2 | We have revised the first two sentences in the paragraph into a more concise sentence and change “resources” to “content”. We have also changed all references for “resources” to “content” throughout the article. |
| 5 | The reviewer was concerned that in paragraph 4 of section 2, there was no indication of the number or percentage of unregistered users that access content on the DD relative to total user access. | We have modified the section to reflect the percentage of registered users that access content on the DD, based on a 2010 study, as follows: “A 2010 study by Gush and De Villiers [[23](#_ENREF_23)] on overall usage patterns of the DD revealed that the majority of logins were by unregistered users (76%), while registered users accounted for only 24% of the total logins. Only demographic information of registered users can be collected. The ratio between the male/female registered users was 77:23”. |
| 6 | The reviewer requests that we clarify the purpose of the “home language” field in section 2.2 | We have revised the section as follows: “Information provided by users is automatically transferred to a central server on a daily basis. Currently this information is only used by the DD team to gather demographic data and application usage patterns [20]. The information provided is not used to customize the applications. For example, the language fields do not affect the default language setting in *OpenSpell* (see section 2.5)”. |
| 7 | The reviewer requests that we clarify, in section 2.5, whether the selection of any of the 11 languages in OpenSpell has any effect on the interface and content of the application | We have modified the paragraph as follows: “*OpenSpell* is an educational spelling game. It is available in all of the eleven South African official languages and provides three levels of difficulty designated with \*, \*\*, \*\*\*. When the user launches the application, the default language is English and the difficulty level is the simplest form. The user can select another language which then changes the interface and content to the language of choice; however, it reverts back to English when the application is restarted”. |
| 8 | The reviewer requests that we use a bigger picture in Figure 9 and give a better explanation of the phases shown in the figure | We have reformatted formatted figure 9 to display over one column and improved our discussion of the phases. |
| 9 | In section 4.1, the reviewer provided more pointers on the danger of poor usability | As suggested by the reviewer, we have included new reference to the danger of poor usability as follows: “…without usability, there is a risk that user communities might not benefit from content as they can’t get to it. Further, a complex interface can lead to users perceiving themselves as unable to use a computer, or low computer self-efficacy [[8](#_ENREF_8)], which can in turn result in people abandoning the DD altogether”. |
| 10 | The reviewer is concerned that the discussion on the different usability evaluation methods provided in section 4.2 is too detailed. | We have now revised the section extensively to provide a more concise summary of the different evaluation methods. |
| 11 | The reviewer questioned the use of heuristic evaluation, since it is usually performed by experts. Although the reviewer acknowledges that heuristic evaluation was used in conjunction with user-based evaluation, there was a concern regarding the number of user who participated in the study. | We have strengthened the argument on the use of heuristics by experts and non-experts considerably. We agree that the first version of the paper did not clearly distinguish between the purpose of developing the heuristics and the evaluation results presented in the paper. Since the purpose of the evaluation was not to be a full-blown usability study, the number of participants is in line with that used in other research projects of the same nature. |
| 12 | The reviewer questioned the value of added by Figure 11 | We agree with the reviewer that the initial Figure 11 did not add value to the paper and have removed it from the revised version. |
| 13 | The reviewer suggests we use percentages of total problems rather than raw data in Figures 12 and 13. It was also suggested that we use a line chart for these Figures instead of columnar charts | As suggested by the reviewer, the revised version of the paper now has a line chart (Figure 12), reflecting the relationship between the problem percentages per evaluator and the heuristics category. |
| 14 | In section 6.2.2, the reviewer requests that we provide specific numbers for the results of the field evaluation instead of using ‘some’. | We have revised the entire section so that the numbers of problems are explicitly stated. |
| 15 | The reviewer requests that we rename the section 7 heading by omitting the word ‘Conclusion’. | We have renamed the section heading as follows: “Reflection on the use of instrument in evaluating the digital doorway”. |
| 16 | The reviewer would like to know if there is anything else that can be seen with a comparison of heuristics and questionnaire | Section 7 has been modified to include a new sub-section 7.1.3, titled “Comparing heuristic evaluation and responses to questionnaires”. |
| 17 | The reviewer wanted to know whether we would consider evaluation of the learnability and memorability of the DD interfaces, since it is a long term, occasional use type of device. | The multi-category heuristics includes aspects of learnability, although issues of memorability were not addressed. Future studies would consider issues of memorability, but would require sorting out the unique identification of users first. |