Hermawati, Setia and Lawson, Glyn (2016) Establishing usability heuristics for heuristics evaluation in a specific domain: is there a consensus? Applied Ergonomics, 56 . pp. 34-51. ISSN 1872-9126

Establishing usability heuristics for heuristics evaluation in a specific domain: is there a consensus?

Setia Hermawati*, Glyn Lawson

Human Factors Research Group, The University of Nottingham, NG7 2RD, United Kingdom

Setia.hermawati@nottingham.ac.uk

Heuristics evaluation is frequently employed to evaluate usability. While general heuristics are generally suitable to evaluate most user interfaces, there is still the need to establish heuristics for specific domains to ensure that usability issues that are specific to the domains are identified. This paper presented a comprehensive review of 70 studies related to usability heuristics for specific domains. The aim of this paper is to review the processes that were applied to establish heuristics in specific domains and identify gaps in order to provide recommendations for future research and area of improvements. The most urgent issue found is the deficiency of validation effort following heuristics proposition and the lack of robustness and rigour of validation method adopted. There is an early indication that heuristics for specific domains are generally capable of identifying more issues than general heuristics. However, due to lack of validation quality and clarity on how to assess their effectiveness, it is not yet clear to what extent their advantages are. The lack of validation quality also affects effort in improving existing heuristics for specific domain as their weaknesses are not addressed.

1. Introduction

Usability is a key to ensure the success of a system (Markus and Keil 1994). ISO-9241-11 defined usability 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". To identify usability

issues and/or measure the degree of usability metrics compliance, usability evaluation is usually performed at different stages of system development (Nielsen 1994a). While the most robust approach for usability evaluation involves users, usability inspections are also commonly applied, especially at the early stages of a system's development. Heuristics evaluation, due to its simplicity and low cost (Jeffries et al. 1991; Tang et al. 2006; Hwang and Salvendy 2010), has gained popularity and is frequently employed in usability studies. This method was originally proposed by Nielsen and Molich (1990) and requires a number of experts to inspect usability based on "heuristics" which are essentially broad and non-specific rules of thumb.

Nielsen's (1994b) ten heuristics are normally used as the heuristics that guide usability evaluation. These heuristics consist of: 1) visibility of system status, 2) match between system and the real world, 3) user control and freedom, 4) consistency and standards, 5) error prevention, 6) recognition rather than recall, 7) flexibility and efficiency of use, 8) aesthetic and minimalist design, 9) help users recognise, diagnose, and recover from error, and 10) help and documentation. While these heuristics could be used to evaluate user interfaces for various domains, heuristics adjustment are needed to ensure that usability issues that are specific to user interfaces of certain domains are not overlooked (Nielsen 2015). As a result, various studies have attempted to establish usability heuristics that are specific for various domains.

The establishment of heuristics for specific domains could take three forms (Ling and Salvendy, 2005): 1) expansion of the heuristics set, 2) alteration of the evaluation procedure, and 3) alteration of the conformance rating scale. This paper focuses on the first form which generally involves two stages i.e. generation of the heuristics sets and their validation. To the extent of the authors' knowledge, there are only two studies that somewhat review how usability heuristics for specific domains were established i.e. Ling and Salvendy (2005) and van Gruenen et al. (2011). Both studies were limited in the number of studies that were included in the review and the scope of the review. Only 20 studies were included in Ling and Salvendy's (2005) review and only five in van Gruenen et

al. (2011). Both mainly focused on the process related to generation of heuristics sets and overlook the validation. Furthermore, both studies aimed for description rather than a critical review.

The main objective of this paper is to review and analyse existing heuristics for specific domains in three aspects: 1) the process that were applied to generate and express the heuristics sets, 2) the methods or approach that were applied to validate them, and 3) their effectiveness. This paper also aims to identify gaps with respect to establishing heuristics for specific domains and provide recommendations for future research and area of improvements. This paper is the first paper that provides a thorough review on the three aspects above. It begins by describing how studies that were included in the review were identified and how they were analysed. Next, the results of the review and analysis were explained in detail. The last section of the paper discusses recurrent issues or phenomenon from the reviewed studies and identifies emerging issues and future research questions that need to be addressed to advance the contribution of heuristics in specific domains. This paper's main contribution lies on the identification of gaps and how to address these gaps through recommendation for future research.

2. Method

The studies included in this review were identified primarily from six databases which were considered to be relevant to usability evaluation: 1) ZETOC, 2) IEEE-IET Electronics Library, 3) Scopus, 4) Science Direct, 5) ACM Digital Library, and 6) Abstracts in New Technology and Engineering. The term 'usability and heuristic' was used to perform the search. The broad search term ensured that no pertinent studies were overlooked. Only studies in English that were published beyond the year 2000 and allowed a full-text access from the University of Nottingham were considered. It should also be noted that only studies that contained propositions for usability heuristics in a specific domain were included. Studies that reported heuristics for other aspects (e.g. aesthetics, inclusiveness) were excluded. Studies that reported usability heuristics but were intended to be used as a guidelines during design activities were also excluded.

For each study that was included in the review, we identified the five themes as follows:

(1) The domain of the proposed usability heuristics.

(2) How the heuristics were created (based on either explicit or implicit description provided in a study).

(3) Whether or not the proposed heuristics were validated and if the proposed heuristics were validated, the type of approaches that were used to validate the proposed usability heuristics.

(4) How the proposed usability heuristics were described (e.g. as a checklist, adopting a formal format) and how different they are from Nielsen's (1994) heuristics which was the most frequently used in heuristics evaluation.

(5) The effectiveness of the proposed usability heuristics in a heuristics evaluation.

3. Results

Figure 1 provides a flow chart documenting the results of the study selection process, which resulted in the inclusion of a total of 70 studies (based on 90 articles) in this review. Appendix 1 and 2 provides an overview of all of the reviewed studies. In appendix 1, domains in which specific usability heuristics were proposed and how they were created and validated are listed. In appendix 2, an overview of the proposed heuristics and their effectiveness are detailed.

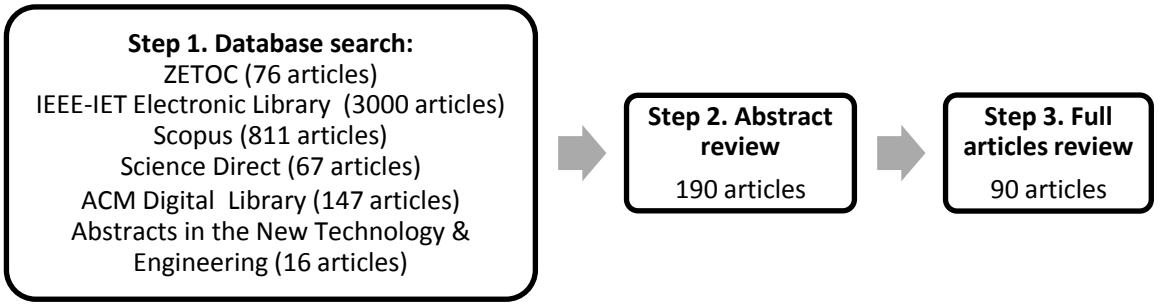

Figure 1. The results of articles selection process

Based on the 70 studies, the proposed usability heuristics for specific domains could generally be divided into three categories (see Figure 2): 1) usability for application or software – 83%, 2) usability

for devices – 14%, and 3) usability for buildings – 3%. In the first category, it was shown that 34.5% of studies were aimed for web-based applications albeit for different type of purposes e.g. e-learning (Reeves et al., 2002; Ardito et al., 2004, 2005; Dringus and Cohen, 2005), social networking (Alroobaea et al., 2013), and intercultural (Diaz et al., 2013). In addition to this, quite a few studies in this category were also aimed for mobile-based applications (e.g. Neto and Pimentel, 2013; Joyce and Lilley, 2014) as well as groupware applications (e.g. Herrmann, 2009; Karousos et al., 2010). On the contrary, in the second category, there was no obvious trend or pattern with a number of studies were aimed for medical devices (Katre et al., 2010; Zhang, et al., 2003), Human Robot Interaction systems (Clarkson and Arkin, 2007; Tsui et al., 2009), specific displays (Mankoff et al., 2003; Somervell et al., 2003) and input devices (Inostroza, et al., 2012a; Maike et al., 2014). In the third category, interestingly, usability heuristics were also applied to support buildings evaluation either in a physical form (Fink et al., 2010) or a conceptual form (Afacan & Erbug, 2009). The diversity of domains revealed in this study suggested the versatility of heuristics evaluation for usability evaluation. While the heuristics method was initially developed to evaluate user interface for application or software, it seemed that it could also be adopted for other things.

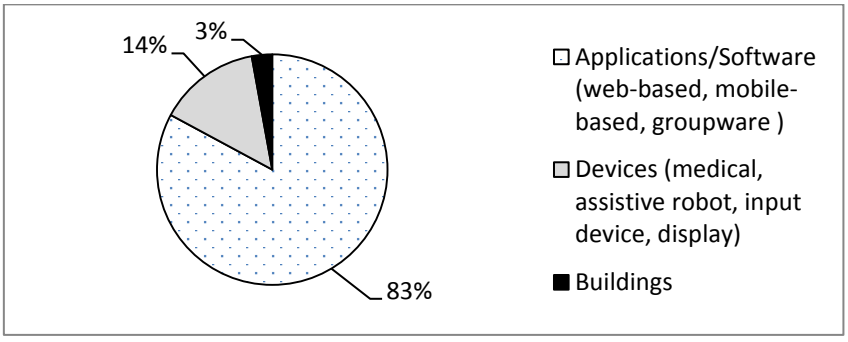

Figure 2. Domains in which specific usability heuristics were proposed

3.1. Creating usability heuristics

Based on a review of the studies, the creation of usability heuristics was shown to consist of two steps: 1) extraction of information, and 2) transforming the extracted information into heuristics. Four different methods were observed for extracting information (see Figure 3). These were: 1)

adopting one or more theories as a basis to identify aspects that are relevant for users' interactions; 2) studying the context of use and identifying aspects that are relevant for users, 3) studying and synthesising reported pertinent usability issues and/or existing heuristics/guidelines, and 4) developing a corpus of usability issues and identifying pertinent issues. It was also found that these studies chose to either apply one method (83%) or combine them (17%). However, while all of the observed studies performed a literature review to establish the state of the art of usability heuristics in their domain of interest, most of the studies did not provide specific reasoning on why one or more of these methods were adopted over the other. For instance, the third method would be suitable for an established domain in which a wealth of information is readily available while other methods would be more appropriate for a less established domain. Ideally, all methods should be considered and given the same weight as focussing on certain methods and disregarding the remaining might affect the final outcome of heuristic. For instance, Greenberg et al. (2000) and Baker et al. (2001, 2002) limited themselves on utilising the first method while Somervell et al. (2003) focused on utilising the second and fourth method. Both studies resulted in a rather limited scope of heuristics as the heuristics became too specific.

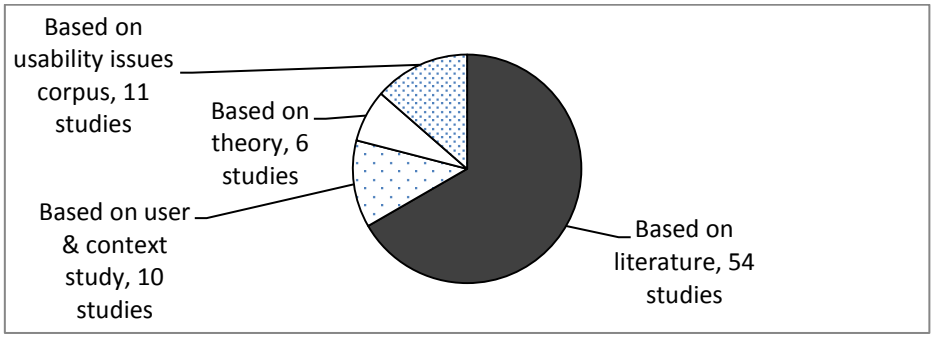

Figure 3. Domains in which specific usability heuristics were proposed

While there is a pattern on how information to establish heuristics was initially extracted, there was a less clear-cut on how extracted information transformed into heuristics. In fact, most studies did not provide clear information on how this process was achieved. For those that provided a more

detailed description, generally there were three approaches (see Figure 3). The first approach (e.g. Federoff, 2002; Zhang et al., 2003; Desurvire et al., 2004) involved listing extracted information (be it guidelines, usability issues, and existing heuristics), omitting any redundancies and irrelevancies, and then using the outcome as final set of heuristics. In the second approach (e.g. Yeratziotis, et al., 2011a, 2011b; AlRoobaea et al., 2013), extracted information that had gone through omission of redundancies and irrelevancies, was categorised to identify themes which then translated into heuristics. A variety of methods were used to aid categorisation of extracted information, ranging from requesting opinion of a number of experts to card sorting technique. The last approach (e.g. Rusu, et al., 2010; Paz et al., 2014; Inostroza, et al., 2012a; Muñoz et al., 2011) involved comparison of the listed extracted information with a general heuristics such as Nielsen (1994b) to identify required modification of existing and/or addition of new heuristics.

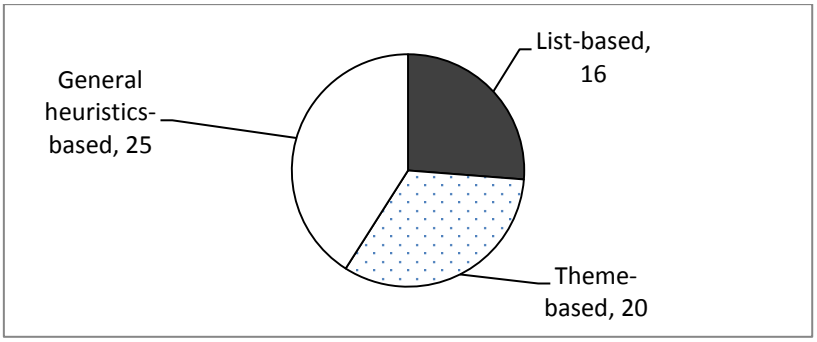

Figure 3. Domains in which specific usability heuristics were proposed

3.2. Validation of usability heuristics

In terms of validation, 34% of studies did not report any type of validation. The majority of remaining studies opted to perform validation at the end of the study which was aimed to investigate the heuristics' effectiveness. However, there were also a few of the studies that integrated validation as part of heuristics establishment in the sense that validation was used a means to assess and improve the effectiveness of heuristics (e.g. Väänänen-Vainio-Mattila and Wäljas, 2009; Reeves et al., 2002; Kuparinen et al., 2013). Either way, the methods varied from one study to another. Table 1 summarises the validation methods adopted from the most to the least common. Please note that some studies combined more than one method as part of the validation.

Table 1

No	Validation methods	No of studies	Description
1	Application of usability heuristics by experts	24	The usability heuristics were used to identify usability issues and their severity.
2	Comparison of the results of usability heuristics' application to other heuristics	20	The effectiveness of usability heuristics in specific domain was compared to existing heuristics. A wide range of data analysis was performed.
3	Comparison of the results of usability heuristics' application to user study	5	The effectiveness of usability heuristics in specific domain was compared to user study. Sometimes, issues found from a user study were also used to help creating master list of usability issues.
4	Other	1	Comparing the heuristics with the results of user requirements elicitation

As shown in Table 1, the application of usability heuristics by experts was shown to be the most popular. Unfortunately, while it is highly practical, this type of validation method does not provide in-depth information on the effectiveness of the usability heuristics. Although, in order to minimise this limitation, a small number of studies (e.g. Katre et al., 2010; Baker et al., 2002) theoretically compared their analysis of type and number of usability issues to other studies that adopted different set of usability heuristics. However, theoretical comparison presented many flaws such as no similarities between application/devices being evaluated, and lack of control on minimising effect of the evaluators. On the contrary, the second and third methods, despite being more time consuming and resource demanding, offer more opportunities to provide a complete picture on the effectiveness of usability heuristics. Interestingly, while the data obtained are more or less the same from one study to another, the types of analysis performed on the second methods varied. Only a small number of studies (i.e. Jaferian et al., 2011; AlRoobaea et al., 2013, Somervell & McCrickard,

2005) opted to adopt pre-defined measurements (Hartson et al., 2001) which allowed easy comparison among usability evaluation methods.

A majority of studies performed a wide range of data analysis on usability issues that were identified and used them as a basis to compare or assess the performance of heuristics in specific domain (see Table 2). While the variety of type of data analysis offered different perspectives, they presented difficulties for comparison of the effectiveness of usability heuristics, especially for studies that aim to improve existing usability heuristics from a specific domain.

Table 2

Type of analysis	Frequency
Difference on number of issues found	9
Qualitative feedback	7
Issue severity vs number of issues found, for each heuristic	4
Number of overlapped and specific issues	4
Issues found for each severity level	3
Known issues found vs number of evaluators	3
Measurements from Hartson et al. (2001)	3
Time	3
Description of usability issues per heuristics	3
Issues identification per heuristics	2
Number of issues with critical, major &minor problem	2
Non-applicable heuristics	1
Description of usability issues	1
Frequently used heuristics	1
Description on heuristics interpretation, redundancy, and conflict based on usability issues	1
Issues identified per subject	1
Efficiency per subject (number of issues/time)	1
Effectiveness per subject (number of issues/number of total known issues)	1

Description of usability issues with critical, major &minor problem	1
Severity of violation per heuristics	1
Severity of violation per issues	1
Average number of issues per heuristics	1
Average severity per heuristics	1
Evaluator effect	1
Number of evaluators required to perform evaluation	1
Cost	1
Themes of issues	1

3.3. Expressing usability heuristics

About 48.5% of studies proposed ≤ 10 usability heuristics. The maximum number of usability heuristics was 38, proposed by Carvalho's et al. (2009) study. 18.5% of the studies adopted a modular approach in which heuristics were assigned into different modules. Each module represents a specific aspect which was seen as being crucial within the context of use of an application or device. When a module approach is adopted, there also seems to be a tendency to introduce a rating or weighting system to judge the compliance level towards heuristics (e.g. Komarkova et al. 2007, Omar et al. 2010, Lynch 2011). The rating system allows identification of module with the biggest weaknesses and quantitatively compared heuristics compliance among applications or devices (e.g. Alsumait & Al-Osaimi 2010, Katre et al. 2010, Liao and Shen 2012). A modular approach also allows flexibility in a sense that certain module that are not relevant could be left out if necessary. Most of the heuristics were expressed in a short, succinct and clear sentence with additional description provided. The amount of additional description varied from short (similar to Nielsen's (1994)) to long. Some studies that provided long descriptions also adopted a more formal approach in which a certain format was followed (e.g. Rusu, et al. 2010, Muñoz et al. 2011, Inostroza, et al. 2012a, Diaz et al. 2013, Solano et al. 2013). In other studies, instead of

providing a lengthy description for each heuristic, checklists were introduced (e.g. Dringus & Cohen 2005, Kemp et al. 2008, Alsumait & Al-Osaimi 2010, Yeratziotis et al. 2011a, Al-Razgan et al. 2014).

About 83% of studies contained heuristics that were similar to some or all of Nielsen's (1994) heuristics. In some cases, the exact terminology of Nielsen's heuristics was used and only the description of an adapted heuristic was modified such that it fitted with the context of a domain (e.g. Inostroza, et al. 2012a, Zhang et al. 2003, Pang et al. 2005, Moraes & Silverira 2006, Collazos et al. 2009). While this approach was practical, it was also shown to cause misconception by experts and resulted in misidentification of heuristics violations (Joyce and Lilley, 2014). In other cases, both the terminologies and descriptions were modified to reflect the context of domain (e.g. Korhonen & Koivisto 2006, Aitta et al. 2008, Tsui et al. 2009). Only a small number of studies (13%) proposed heuristics that did not show an obvious link with Nielsen's (1994). The heuristics proposed in these cases were typically expressed in a lower level abstraction and aimed to evaluate features that are inherent of an application or device in a domain. In addition to this, these heuristics were proposed solely based on either a theory that was deemed to be fundamental to an application/device (e.g. Greenberg et al. 2000, Baker et al. 2002, Zuk et al. 2006a, Fink et al. 2010) or the results of studies with much emphasis on the end-user point of view (e.g. Somervell & McCrickard 2005, Herrmann 2009, Katre et al. 2010).

3.4. Effectiveness of usability heuristics

Out of 70 studies, only 17 studies could be used to provide indication on the effectiveness of usability heuristics in a specific domain. This was because most studies fell into one of the following categories: 1) did not perform any validation, 2) did not conduct validation in which effectiveness of usability heuristics in specific domains were compared to another heuristics set or usability evaluation method, or 3) did not analyse the comparison result quantitatively and focused solely on detailed textual descriptions of usability issues.

The reported findings from the 17 studies suggest that usability heuristics for specific domains found more usability issues than using existing heuristics that are more general in nature (Berry 2003,

Mankoff et al. 2003, Desurvire et al. 2004, Zhou et al. 2004, Bertini et al. 2006, Conte et al. 2009, Pinelle et al. 2009, Singh & Wesson 2009, Tsui et al. 2010, Rusu et al. 2011, Jaferian et al. 2014, Muñoz & Chalegre 2012, Inostroza et al. 2012b, Diaz et al 2013, Kuparinen et al. 2013, Neto & Pimentel 2013, Solano et al. 2013). However, following further investigation, the results are shown to be inconclusive. Only 6 studies (out of the 17 studies) performed any statistical comparison and of those that did, four studies reported a significant difference (i.e. Zhou et al. 2004, Conte et al. 2009, AlRoobaea et al. 2013, Jaferian et al. 2014) and the remaining two studies reported no significant difference (i.e. Berry 2003, Inostroza et al. 2012b).

Eleven studies (out of 17 studies) also investigated the effectiveness of heuristics in specific domains with respect to severity of usability issues. Two approaches were adopted: 1) averaged severity values for each and all of heuristics, 2) number of issues for each severity level. The results of the first approach are inconclusive with some studies suggesting that the heuristics set for specific domains identified more severe issues than general heuristics (e.g. Diaz et al 2013) and vice versa (e.g. Muñoz & Chalegre 2012). In the second approach, all studies adopted the severity rating scale proposed by Nielsen (1995 –http://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/) which differentiated severity issue into 5 categories: no issue, cosmetic, minor, major and catastrophe. Similar to the first approach, the results are also inconclusive. Some studies reported that heuristics set for specific domain performed better in identifying major & catastrophe issues than general heuristics (Mankoff et al. 2003, Bertini et al. 2006, Jaferian et al. 2014, Tsui et al. 2010) while others reported a mixed results (Kuparinen et al. 2013, AlRoobaea et al. 2013).

Only three studies (i.e. Jaferian et al. 2014; AlRoobaea et al. 2013; Somervell & McCrickard (2005)) adopted metrics proposed by Hartson et al. (2001) i.e. thoroughness, reliability, validity and effectiveness while comparing the effectiveness of heuristics for specific domain and general heuristics. From this small sample, comparison of the results was found to be quite straightforward as the same definition of metrics was applied. Based on the three studies, it was found that heuristics for specific domain performed better with respect to thoroughness (*proportion of real*

usability issues identified), reliability (*agreement between evaluators*), validity (*proportion of identified usability issues that are real issues*), and effectiveness (*thoroughness multiplied by validity*). Unfortunately, these three studies adopted different definition of what regarded as "real issues". In Jaferian et al (2014), usability issues identified by the two evaluators groups (who used Nielsen's heuristics and domain heuristics) were aggregated and four independent researchers identified major usability issues which were then used as "real issues". In AlRoobaea et al. (2013), usability issues were aggregated from three different groups (real end-users who performed usability evaluation and two evaluators groups that used Nielsen's heuristics and domain heuristics) and then the researchers determined severity ratings with major and severe usability issues used as "real issues". In Somervell & McCrickard (2005), pre-identified problem sets were used as the 'real' problem sets with no explanation given on how and who determined these problems. Thus, considering the findings above, it is difficult to establish whether or not domain heuristics actually perform better than conventional ones.

A further observation also revealed the variability when it comes to what considers being "experts" as heuristics evaluators. All but three studies (Conte et al. 2009, Singh and Wesson 2009, Jaferian et al. 2014) failed to provide sufficient information on how HCI or usability experts were determined i.e. whether it was based on formal educational training/qualification, professions. Furthermore, only few studies provided quantitative demographical information related to evaluators' expertise (Mankoff et al. 2003, Singh and Wesson, 2009, Jaferian et al. 2014). Some studies even did not supply any information on the level of experience of these experts with respect to the specific domains being evaluated (Berry 2003, Tsui et al. 2010, Diaz et al. 2013, Kuparinen et al. 2013, Inostroza et al. 2012b). Heuristic evaluation is heavily dependent on the performance of experts involved in the study (Hertzum and Jacobsen, 2001). Thus, the lack of information related to the experts involved in the reviewed studies introduced bias and complicated further the assessment on the effectiveness of heuristics in specific domain.

An aspect that is found to be similar nearly across all of the studies was that heuristics for specific domain only identified some of usability issues that were identified by general heuristics. It was found that this was likely caused by evaluators' tendency to ignore them as they were seen to be less problematic in comparison to other issues (Muñoz & Chalegre 2012, Inostroza et al. 2012, Neto & Pimentel 2013, Solano et al. 2013). There was only three studies reported comparison related to the time taken to perform usability evaluation. Two of the studies reported that heuristics for specific domain took a much longer time to complete than general heuristics (Bertini et al. 2006, Conte et al. 2009) while only one reported otherwise (Somervell & McCrickard 2005).

4. Discussion

Usability heuristics was originally proposed to ensure that user interface of a system is easy to use. However, as shown by the review related to application domain of usability heuristics, there appears to be the possibility that its use could be widened beyond user interface of a system. While current alternative examples on the review was limited to usability evaluation of bathroom and building plans, it could easily be extended to other areas, for instance during design and development of a public transportation system, especially related to ease of swapping between one mode of transportation to another, accessibility, etc.

The systematic review shows that, while there is an underlying pattern on how to establish usability heuristics for specific domains, the rigour and robustness of the reviewed studies vary from one to another. What is more surprising is the fact that, while some studies built upon previous works, there also seems to be frequent occasions where studies simply created a new set heuristics without taking into account the finding from available works related in the domain. It is highly likely that this was caused by the lack of validation of proposed heuristics. As reported in the results section, out of 70 studies, only 17 studies (≈ 24%) performed validation whereby the effectiveness of heuristics for specific domains was compared to general or existing heuristics. Through validation, the weaknesses of the proposed heuristics could be identified which in turn provide a starting point for other studies. In addition to this, only a few studies actually compared the effectiveness of the

proposed heuristics with existing heuristics in the same domain (e.g. Somervell & McCrickard 2005, Zuk et al. 2006). Effectiveness comparison with respect to well-known usability heuristics such as Nielsen and Molich (1990) is in a way also to provide opportunities for comparison across heuristics sets, provided that the methods employed for comparison and reporting outcome measures are comparable. However, as shown in the review results, this is not the case and thus omitting the possibility of effectiveness comparison for heuristics that are in the same domain.

As previously mentioned, it was evident from the review that robust and rigorous validation for heuristics in specific domain is lacking. Most studies terminated either as soon as proposition of heuristics for specific domains were made or once the proposed heuristics was employed to evaluate one or two applications. This, in turn could result in abundance of the proposition of a considerable number of heuristics in the same domain without any indication of their actual effectiveness. For those that went beyond these points, a wide variety of analysis of varying degree of rigour was employed among the studies. Unfortunately, only a few of studies show sufficient robustness and rigour and adopted a similar approach in the data analysis. Therefore, it is important that, not only validation method is robust and rigorous, but there should also be certain consensus on what to report as part of a validation exercise. For instance, Hartson et al (2001) proposed metrics that could be used as one option. In addition to this, severity distribution of usability issues could also be added. Furthermore, where it is appropriate, statistical comparison could also be performed.

There are also appears to be a tendency to integrate general heuristics as part of heuristics for specific domain. The current approach is to sort and judge the relevancy of existing heuristics and then either adopt the existing wording or terminologies and modify the description or alter the wording or terminologies such that they're relevant to specific domains. This appears to be successful as demonstrated by the overlap of identified usability issues using heuristics for specific domains and general heuristics. However, as indicated by the findings of the review, some usability issues were only found by using general heuristics. While it was claimed that these issues were

either mostly minor or issues that otherwise could be identified through heuristics for specific domains but chosen to be ignored by evaluators, this suggest that there is still room for improvement. Some studies suggest combining using both heuristics for specific domains and general heuristics. Unfortunately, this suggestion would likely demand more resources. Furthermore, simply using both heuristics also means that some time would be wasted on identifying the same usability issues. A possible remedy is to adopt a modular approach for the heuristics whereby one module is dedicated to identify usability issues that are likely found through general heuristics and another module is dedicated to identify usability issues that are specific to the domain.

When it comes to expressing heuristics, some studies utilised checklists as an alternative form to provide detailed description for each heuristic. This is mostly motivated by the opportunities to obtain scores for each statement in the checklist which can then be accumulated to provide the overall usability of a system. However, the problem with a checklist is that it could be time consuming, especially when there are a considerable number of statements in the checklist. Furthermore, they do not necessarily allow the capture of usability issues in detail and there is possibility that some issues would be referred more than once for more than one statement. Therefore, if the aim of establishing heuristics for specific domains is to identify usability issues, perhaps a high level of general statement would be more appropriate. In this case, the number of heuristics should be not so many such that evaluators can remember all of the proposed heuristics. In addition to this, heuristics statements should be made clear, especially if modified from existing heuristics, such that misunderstandings could be minimised.

Based on the review of existing studies, the following recommendations are made for future studies:

1. *Robust and rigorous validation and adoption of standard measures as indicators of heuristics' effectiveness.* Establishing heuristics for specific domains should not stop once the heuristics are proposed. It needs to be followed by performing a robust and rigorous validation in which the effectiveness of heuristics for specific domains is

compared to other heuristics (general heuristics and/or other existing heuristics for specific domain). In addition to this, standard measures should be adopted to indicate the heuristics' effectiveness. For instance, a combination of Hartson et al. (2001) metrics and severity distribution of identified issues could be adopted as minimum standard measures.

2. *Building on heuristics that already exist in a domain.* At the moment, this was mostly performed during the creation of heuristics for specific domain. However, as a result of lack of robust and rigorous validation, the efforts to build on existing heuristics were limited to simply reviewing the heuristics without knowledge on their effectiveness. As a result, only a few studies show the continuation of heuristics from one study to another even if they are in the same domain. Therefore, some effort should also be aimed to validate existing heuristics and not simply focused on proposing new heuristics.
3. *Better definition of what constitutes as experts with respect to usability and specific domain.* It is of importance to provide sufficient and detailed definition related to experts that are involved as part of establishing the heuristics for specific domain. At the moment, this information was found to be inconsistent, even scarce, across different studies. In addition to this, as heuristics for specific domain require knowledge in usability and specific domain, a fair representation of expertise from both is imperative and should be adhered throughout the process i.e. not only limited to the creation of usability heuristics stage.

5. Conclusion

The review of 70 studies related to usability heuristics for specific domains showed that the need for heuristics in specific domains is widely acknowledged across various domains. There is generally a consensus on how to create usability heuristics for specific domains (i.e. extraction of information, and transforming the extracted information into heuristics). However, adopted approaches for each

step vary across different studies and there is yet guidance on the best approaches for each step. Guidance is clearly needed as it ensures that proposed heuristics is created based on thorough and rigorous process that could be replicated. The review also shows that there is yet established proof across the different domains with respect to the effectiveness these heuristics. While some studies showed usability heuristics for specific domains found more usability issue, the usability issues identified were not necessary "real" usability issues (i.e. encounter by real end users) as usability studies were really seldom used as the standard to compare the performance between specific and general heuristics. Furthermore, as most of the heuristics for specific domain proposed heuristics that showed an obvious link with Nielsen's heuristics, this poses a further question on the real contribution of heuristics for specific domains. In other words, due to lack of validation quality, it is not yet clear to what extent the advantages of heuristics for specific domains are. The lack of validation quality also affects effort in improving existing heuristics for specific domain as their weaknesses are not addressed.

Acknowledgement

This work was supported through the funding from "CloudFlow" research project under grant agreement no. 609100ICT-285176, which is funded by the European Commission's 7th Framework programme (FP7-2013-NMP-ICT-FoF).

References

Aitta, M-R., Kaleva, S., Kortelainen, T., 2008. Heuristic evaluation applied to library web services. New Library World. 109 (1/2), 25-45.

Afacan, Y., Erbug, C., 2009. An interdisciplinary heuristic evaluation method for universal building design. Applied Ergonomics. 40, 731-744.

Al-Azawi, R., Ayesh, A., Kenny, I., Al-Masruri, K. A., 2013. A generic framework for evaluation phase in games development methodologies. In Proceedings of Science and information Conference, 237-243.

Albion, P.R., Heuristic evaluation of educational multimedia: from theory to practice. Proceedings of 16th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education, ASCILITE, 1999.

Alotaibi, M.B., 2013. Assessing the usability of university websites in Saudi Arabia: a heuristic evaluation approach. In: Proceedings of the 10th International Conference on Information Technology: New Generations, 138-142.

AlRoobaea, R. S., Al-Badi, A.H., Mayhew, P., 2013. A framework for generating domain-specific heuristics for evaluating online educational websites. International Journal of Information Technology & Computer Science. 8(2), 75-84.

Al-Razgan, M. S., Al-Khalifa, H. S., Al-Shahrani, M. D., 2014. Heuristics for evaluating the usability of mobile launchers for elderly people, in: Marcus, A (Ed.), DUXU 2014 Part I LNCS 8517, pp. 415–424.

Alsumait, A., Al-Osaimi, A., 2010. Usability heuristics evaluation for child E-learning applications. Journal of Software. 5(6), 654-661.

Amar, R., Stasko. J. A., 2004. Knowledge task-based framework for design and evaluation of information visualizations. In: Proceedings of IEEE InfoVis Conference, 143–149.

Ang, C.S., Avni, E., Zaphiris, P., 2008. Linking pedagogical theory of computer games to their usability. International Journal on E-Learning. 7(3), 533-558.

Albion, P.R., Heuristic evaluation of educational multimedia: from theory to practice. Proceedings of 16th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education, ASCILITE, 1999.

Ardito, C., Marsico, M.D., Lanzilotti, R., Levialdi, S., Roselli, T., Rossano, V., Tersigni, M., 2004. Usability of E-learning tools. In: Proceedings of the Working Conference on Advanced Visual Interfaces, 80-84.

Bowie, J.L., 2012. Sound usability? Usability heuristics and guidelines for user-centered. Communication Design Quarterly Review. 13(2), 15-24.

Baker, K., Greenberg, S., Gutwin, C., 2002. Empirical development of a heuristic evaluation methodology for shared workspace groupware. In: Proceedings of Conference on Computer Supported Cooperative Work, 96-105.

Baker, K., Greenberg, S., Gutwin, C. 2001. Heuristic evaluation of groupware based on the mechanics of collaboration, in: Reed, M., Nigay, L. (Eds.) LNCS 2254, pp. 123–139.

Bertin, J., 1983. Semiology of Graphics, The University of Wisconsin Press, WI.

Bertini, E., Gabrielli, S., Kimani, S., 2006. Appropriating and assessing heuristics for mobile computing. In: Proceedings of the working conference on Advanced Visual Interfaces, 119-126.

Berry, B. 2003. Adapting heuristics for notification systems. In: Proceedings of 41st Annual ACM Southeast Conference, 144-149.

Carvalho, C.J., Borycki, E.M., Kushniruk, A., 2009. Ensuring the safety of health information systems: using heuristics for patient safety. Healthcare Quaterly, 12, 49-54.

Clarkson, E., Arkin, C. 2007. Applying heuristic evaluation to Human-Robot Interaction systems. In: Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference, 44-49.

Collazos, C.A., Rusu, C. , Arciniegas, J.L., Roncagliolo, S., 2009. Designing and evaluating interactive television from a usability perspective. In: Proceedings of Second International Conferences on Advances in Computer-Human Interactions, 381-385.

Conte, T., Massolar, J., Mendes, E., and Travassos, G.H., 2007. Usability evaluation based on web design perspectives. In: Proceedings of the First International Symposium on Empirical Software Engineering and Measurement, 146-155.

Conte, T., Massolar, J., Mendes, E., and Travassos, G.H., 2009. Web usability inspection technique based on design perspectives. IET Software. 3(2), 106-123.

Costabile, M. F., De Marsico, M., Lanzilotti, R., Plantamura, V. L., and Roselli, R., 2005. On the usability evaluation of E-learning application. In: Proceedings of the 38th Hawaii International Conference on System Sciences, 1-10.

Desurvire, H., Caplan, M., and Toth, J. A., 2004. Using heuristics to evaluate the playability of games. In: Proceedings of Human Factors in Computing Systems, 1509-1512.

Desurvire, H., Wiberg, C., 2009. Game usability heuristics (PLAY) for evaluating and designing better games: the next iteration, in: Ozok, A.A., Zaphiris, P. (Eds.), Online Communities LNCS 5621, pp. 557–566.

Diaz, J., Rusu, C., Pow-Sang, J.A., Roncagliolo, S., 2013. A cultural – oriented usability heuristics proposal. In: Proceedings of the Chilean Conference on Human-Computer Interaction, 82-87.

Dringus, L.P., Cohen, M.S., 2005. An adaptable usability heuristic checklist for online courses. In: Proceedings of 35th Annual Conference Frontiers in Education, T2H-6.

Federoff, M.A., 2002. Heuristics and usability guidelines for the creation and evaluation of fun in video games. Master Thesis. Indiana University.

Fink, N., Park, R., Battisto, D., 2010. Developing a usability evaluation tool to assess the patient room bathroom. HERD. 3(3), 22-41.

Fitzpatrick, G., 1998. The locales framework: understanding and designing for cooperative work. PhD Thesis, Department of Computer Science and Electrical Engineering, The University of Queensland.

Forsell, C., Johansson, J., 2010. An heuristic set for evaluation in information visualization. In: Pproceedings of Advanced Visual Interfaces, 199-206 .

Greenberg, S., Fitzpatrick, G., Gutwin, C., Kaplan, S., 2000. Adapting the locales framework for heuristic evaluation of groupware. Australasian Journal of Information Systems. 7(2), 102-108

Grice, R.A., Bennet, A.G., Fernheimer, J.W., Geisler, C., Krull, R., Lutzky, R.A., Rolph, M.G.J., Search, P., Zappen, J.P., 2013. Heuristics for broader assessment of effectiveness and usability in technology-mediated. Technical Communication. 60(1), 3-27.

Gutwin, C. Greenberg, S., 2000. The Mechanics of collaboration: developing low cost usability evaluation methods for shared workspaces. In: Proceedings of the 9th IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, 98-103.

Hargis, G., Carey, M.,Hernandez, A.K.,Hughes,P., Longo, D., Rouiller, S., and Wilde, E. Developing Quality technical information, 1998, Prentice Hall; New Jersey.

Hartson, H. R., Andre, T.S., Williges, R. C., 2001. Criteria for evaluating usability evaluation methods. International Journal of Human Computer-Interaction. 13, 373-410.

Hermann, T., 2009. Design heuristics for computer supported collaborative creativity. In: Proceedings of the 42nd Hawaii International Conference on System Sciences, 1-10.

Hertzum, M., Jacobsen, N. E. 2001. The evaluator effect: a chilling fact about usability evaluation methods. International Journal of Human Computer Interactions. 13(4), 421-443.

Hub, M., Čapková, V., 2010. Heuristic evaluation of usability of public administration portal. In Proceedings of the International Conference on Applied Computer Science, 234-239.

Hwang, W., Salvendy, G., 2010. Number of people required for usability evaluation: The 10±2 rule. Communications of the ACM. 53(5), 130–133.

Inostroza, R., Rusu, C., Roncagliolo, S., Jimenez, C., Rusu, V., 2012a. Usability Heuristics for touchscreen-based mobile devices. In: Proceedings of the 9th International Conference on Information Technology: New Generations, 662–667.

Inostroza, R., Rusu, C., Roncagliolo, S., Jimenez, C., Rusu, V., 2012b. Usability heuristics validation through empirical evidences: a touchscreen–based mobile devices proposal. In: Proceedings of the 31th International Conference of the Chilean Computer Science Society, 61-68.

Inostroza, R., Rusu, C., Roncagliolo, S., Rusu, V., 2013. Usability heuristics for touchscreen-based mobile devices: update. In: Proceedings of the 1st Chilean Conference of Computer-Human Interaction, 24-29.

Inostroza, R., Rusu, C., 2014. Mapping usability heuristics and design principles for touchscreen-based mobile devices. In: Proceedings of the 7th Euro American Conference on Telematics and Information Systems, article no 27.

Jaferian, P., Hawkey, K., Sotirakopolous, A., Velez-Rojas, M., and Beznosov, K., 2014. Heuristics for evaluating IT security management tools. Human–Computer Interaction. 29(4), 311-350.

Jaferian, P., Hawkey, K., Sotirakopolous, A., Beznosov, K., 2011. Heuristics for evaluating IT security management tools. In: Proceedings of CHI, 1633-1638.

Jeffries, R., Miller, J. R., Wharton, C., Uyeda, K. M., 1991. User interface evaluation in the real world: a comparison of four techniques. In Proceedings of ACM CHI 1991 Conference on Human Factors in Computing Systems, 119-124.

Jiménez, C., Rusu, C., Gorgan, D., Inostroza, R., 2013. Grid applications to process, supervise and analyze earth science related phenomena: what about usability? In: Proceedings of Chilean Conference of Computer-Human Interaction, 94-97.

Jo, S., Choi, W., Postic, S., Kim, H., Lee, H., 2009. A study on heuristics guideline for MMOG UI. In: Proceedings of the 6th International Conference on Information Technology and Applications, 157-160.

Joyce, G., Lilley, M., 2014. Towards the development of usability heuristics for native smartphone mobile applications, in: Marcus, A. (Ed.), DUXU 2014 Part I LNCS 8517, 465–474.

Kantner, L., Shroyer, R., Rosenbaum, S., 2002. Structured heuristic evaluation of online documentation. In: Proceedings of Professional Communication Conference, 331-342.

Kaptelinin, V., and Nardi, B. A., Acting with technology: Activity theory and interaction design, 2006, MIT Press; Cambridge, MA.

Karousos, N., Papaloukas, S., Kostaras, N., Xenos, M., Tzagarakis, M., Karacapilidis, N., 2010. Usability evaluation of web-based collaboration support systems: the case of CoPe_it!. In: Lytras, M.D., Pablos, P.O.D., Ziderman, A., Roulstone, A., Maurer, H., Imber, J.B. (Eds.), Knowledhe Management, Information Systems, E-Learning and Sustainability Research, 248–258.

Katre, D., Bhutkar, G., Karmakar, S., 2010. Usability heuristics and qualitative indicators for the usability evaluation of touch screen ventilator systems. In Katre, D., Orngreen, R., Yammiyavar, P., Clemmensen, T. (Eds.), Human Work Interation Design: Usability in Social, Cultural and Organizational Context, pp. 83–97.

Kemp, E. A., Thompson, A-J., Johnson, R., 2008. Interface evaluation for invisibility and ubiquity – an example from E-learning. In: Proceedings of International Conference on Human-Computer Interaction, 31-38.

Koivisto, E. M. I., Korhonen, H., 2006. Mobile game playability heuristics. Nokia Corporation.

Komarkova, J., Visek, O., Novak, M., 2007. Heuristic evaluation of usability of geo websites, in Ware, J.M., Taylor, G.E. (Eds.), Web and Wireless Geographical Information Systems, 264–278.

Komlodi, A., Caidi, N., Wheeler, K. 2005. Cross-cultural usability of digital libraries, in: Chen, Z., Chen, H., Miao, Q., Fu, Y., Fox, E., Lim, E. (Eds.), Digital Libraries: International Collaboration and Cross-Fertilization, 584-593.

Korhonen, H., Koivisto, E. M. I., 2006. Playability heuristics for mobile games. In: Proceedings of Mobile Human-Computer Interactions, 9-16.

Kuparinen, L., Silvennoinen, J., Isomäki, H., 2013. Introducing usability heuristics for mobile map applications. In Proceedings of the 26th International Cartographic Conference, 1-11.

Ling, C., Salvendy, G., 2005. Improving the heuristic evaluation method: a review and reappraisal. Ergonomia: An International Journal of Ergonomics and Human Factors. 27(3), 179-197.

Liao, Y. H., Shen, C-Y., 2012. Heuristic evaluation of digital game based learning a case study. In: Proceedings of the 4th IEEE International Conference On Digital Game And Intelligent Toy Enhanced Learning, 192-196.

Lynch, K. R., 2011. Weighted heuristic evaluation and usability testing of Ohio area agency on aging websites for older adults. MSc Thesis. Faculty of the Russ College of Engineering and Technology of Ohio University.

Maike, V. R. M. L., Neto, L. de S.B., Baranauskas, M.C.C., Goldenstein, S.K., 2014. Seeing through the Kinect: A Survey on Heuristics for Building Natural User Interfaces Environments, in: Stephanidis, C., Antona, M. (Eds.), Universal Access in Human Computer Interaction: Design and Development Methods for Universal Access, pp. 407–418.

Mankoff, J., Dey, A., Hsieh, G., Kientz, J., Lederer, S., Ames, M., 2002. Heuristic evaluation of ambient displays. In: Proceedings of the Conference on Human Factors in Computing Systems, 169-176.

Markus, M. L., Keil, M., 1994. If We Build It, They Will Come: Designing Information Systems That People Want to Use. Sloan Management Review. 35(4), 11–25.

Moraes, M. C., Silveira, M. S., 2006. How am I? Guidelines for Animated Interface Agents Evaluation. In: Proceedings of International Conference on Intelligent Agent Technology, 200-203.

Muñoz, R., Barcelos, T., Chalegre, V., 2011. Defining and validating virtual worlds usability heuristics. In: Proceedings of the 30th International Conference of the Chilean Computer Science Society, 171-178.

Muñoz, R. and Chalegre, V., 2012. Defining virtual worlds usability heuristics. In: Proceedings of the 9th International Conference on Information Technology – New Generations, 690-695.

Neto, O. M., Pimentel, M. da G., 2013. Heuristics for the assessment of interfaces of mobile devices. In: Proceedings of WebMedia, 93-96.

Nielsen, J., 1994a. Enhancing the explanatory power of usability heuristics. In Proceedings of ACM CHI'94 Conference, 152-158

Nielsen, J., 1994b. Heuristic evaluation. In Nielsen, J., and Mack, R.L. (Eds.), Usability Inspection Methods, John Wiley & Sons, New York.

Nielsen J., Severity Ratings for Usability Problems, 1995, https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/, (accessed on 18,03.16).

Nielsen, J. 2000. Designing Web Usability: The Practice of Simplicity. New Riders Publishing, Indianapolis.

Nielsen, J., 2015. How to conduct a Heuristic Evaluation. http://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/ Accessed on 1 July 2015.

Nielsen, J., Molich, R., 1990. Heuristic evaluation of user interfaces. In Proceedings of ACM CHI'90 Conference, 249-256.

Nokelainen, P., An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students. Educational Technology & Society, 9 (2), 2006, 178-97.

Omar, H. M., Jafar, A., 2010a. Heuristics evaluation in computer games. In: Proceedings of Information Retrieval & Knowledge Management, 188 - 193.

Omar, H. M., Jafar, A., 2010b. Challenges in the evaluation of educational computer Games. In: Proceedings of International Symposium in Information Technology, 1-6.

Omar, H. M., Jafar, A., 2011. AHP_HeGES: tools to evaluate usability of educational computer game (UsaECG). In: Proceedings of International Conference on User Science and Engineering, 73-76.

Omar, H.M., Yusof, Y. H.M., Sabri, N.M., 2010. Development and potential analysis of heuristic evaluation for courseware. In: Proceedings of the 2nd International Congress on Engineering Education, 128-132.

Pang N. L. S., Cao, S., Schauder, D., Klein, R. R., 2005. A hybrid approach in the evaluation of usability for multimedia objects: case Study of the media assets management platform for an advertainment production project toward Beijing Olympics 2008. In: Proceedings of the 3rd International Conference on Information Technology and Applications, 82-87.

Papaloukas, S., Patriarcheas, K., Xenos, M., 2009. Usability assessment heuristics in new genre videogames. In: Proceedings of the 13th Panhellenic Conference on Informatics, 202-206.

Paz, F., Paz, F. A., Pow-Sang, J. A., Collantes, L., 2014. Usability heuristics for transactional web sites. In: Proceedings of the 11th International Conference on Information Technology, 627-628.

Pinelle, D., Wong, N., Stach, T., Gutwin, C., 2009. Usability heuristics for networked multiplayer games. In: Proceedings of International Conference on Supporting Group Work. 169-178.

Pinelle, D., Wong, N., and Stach, T., 2008. Heuristic evaluation for games: usability principles for video game design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1453-1462.

Quiñones, D., Rusu, C., and Roncagliolo, S., 2014. Redefining usability heuristics for transactional web applications. In: Proceedings of 11th International Conference on Information Technology: New Generations, 260-265.

Reeves, T., Benson, L., Elliott, D., Grant, M., Holschuh, D., Kim, B., Kim, H., Lauber, E. and Loh, S., Usability and instructional design heuristics for e-learning evaluation. Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications AACE, 2002, 1615-21.

Röcker, C., Haar, M., 2006. Exploring the usability of video game heuristics for pervasive game development in smart home environments. In: Proceedings of the 3rd International Workshop on Pervasive Gaming Applications, 124 – 131.

Rusu, C., Roncagliolo, S., Tapia, G., Hayvar, D., Rusu, V., Gorgan, D., 2010. Evaluating the usability of intercultural collaboration platforms: grid computing applications. In: Proceedings of the 3rd ACM International Conference on Intercultural Collaboration, 179-182.

Rusu, C., Roncagliolo, S., Tapia, G., Hayvar, D., Rusu, V., Gorgan, D. 2011. Usability heuristics for grid computing applications. In: Proceedings of the 4th International Conference on Advances in Computer-Human Interactions, 53-58.

Rusu, C., Roncagliolo, S., Figueroa, A., Rusu, V., Gorgan, D., 2012. Evaluating the usability and the communicability of grid computing applications. In: Proceedings of the 5th International Conference on Advances in Computer-Human Interactions, 204-207.

Salvador, V. F. M., Moura, L. de A., 2010. Heuristics evaluation for automatic radiology reporting transcription systems. In: Proceedings of the 10th International Conference on Information Science, Signal Processing and their Applications, 292-295.

Scholtz, J. 2002. Evaluation methods for human-system performance of intelligent systems. In: Proceedings of Workshop in Measuring the Performance and Intelligence of System, 1-6.

Sheridan, T. 1997. Eight ultimate challenges of human-robot communication. In: Proceedings of International Workshop on Robot and Human Communication, 9–14.

Shneiderman, B., 1998. Designing the User Interface, third ed. Addison-Wesley, Reading, MA.

Shneiderman. B., 1996. The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings of the IEEE Symposium on Visual Languages, 336–343.

Sim, G., Read, J. C., Holifield, P., 2008. Heuristics for evaluating the usability of CAA applications. In: Proceedings of the 12th International Computer Assisted Assessment Conference, 283-294.

Sim, G., Read, J. C., Cockton, G., 2009. Evidence based design of heuristics for computer assisted assessment, in: Gross, T., Gulliksen, J., Kotze, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M (Eds.), Human-Computer Interaction INTERACT 2009, 204–216.

Singh, A., Wesson, J., 2009. Evaluation criteria for assessing the usability of ERP system. In: Proceedings of Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, 87-89.

Sivaji, A., Abdullah, A., Downe, A.G., 2011. Usability testing methodology: effectiveness of heuristic evaluation in E-government website development. In Proceedings of the 5th Asia Modelling Symposium, 68-72.

Solano, A., Rusu, C., Collazos, C., Arciniegas, J., 2013. Evaluating interactive digital television applications through usability heuristics. Ingeniare. 21(1), 16-29.

Somervell, J. McCrickard, D. S., 2005. Better discount evaluation: illustrating how critical parameters support heuristic creation. Interacting with Computers. 17, 592-612.

Somervell, J., Wahid, S., McCrickard, D., 2003. Usability heuristics for large screen information exhibits. In: Proceedings of the 9th International Conference on Human Computer Interaction, 904-907.

Somro, S., Ahmad, W. F. W., Sulaiman, S., 2012. A preliminary study on heuristics for mobile games. In: Proceedings of International Conference on Computer & Information Science, 1030-1035.

Sutcliffe, A.G., and Kaur, K.D., Evaluating the usability of virtual reality user interfaces, Behaviour and Information Technology, 19 (6), 2000, 415–426

Sutcliffe, A., Gault, B., 2004. Heuristic evaluation of virtual reality applications. Interacting with Computers. 16, 831-849.

Tang Z., Johnson, T.R., Tindall, R. D., Zhang, J., 2006. Applying heuristic evaluation to improve the usability of a telemedicine system. Telemedicine Journal and e- Health. 12(1), 24-34.

Tsui, K. M., Abu-Zahra, K., Casipe, R., M'Sadoques, J., Drury, J. L., 2009. A process for developing specialized heuristics: case study in assistive robotics. University of Massachusetts Lowell Technical Report 2009-11. Department of Computer Science.

Tsui, K. M., Abu-Zahra, K., Casipe, R., M'Sadoques, J., Drury, J. L., 2010. Developing heuristics for assistive robotics. In: Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction, 193-194.

Tufte, E. R., 2001. The Visual Display of Quantitative Information, 2nd ed. Graphics Press, Cheshire.

Väänänen-Vainio-Mattila, K., Wäljas, M., 2009. Developing an expert evaluation method for user experience of cross-platform web services. In: Proceedings of MindTrek, 162-169.

van der Geest, T., Spyridakis., 2000. Developing heuristics for web communication: an introduction to this special issue. Technical Communication, 3rd quarter: 301-310.

Van Greunen, D., Yeratziotis, A., Pottas, D., 2011. A three-phase process to develop heuristics. In: Proceedngs of the 13th Annual Conference on World Wide Web Appplications, 5-23.

Ware, C., 2004. Information Visualization: Perception for Design, second ed., Morgan Kaufmann Publishers, Massachusetts.

Yeratziotis, A., Pottas, D. and van Greunen, D., A Three-Phase Process to Develop Heuristics. Proceedings of the 13th ZA-WWW conference, 14–16 September 2011, Johannesburg, South Africa, 2011a.

Yeratziotis, A., Pottas, D. and van Greunen, D., Recommendations for Usable Security in Online Health Social Networks. Proceedings of the joint conference of the 2011 6th International Conference on Pervasive Computing and Application (ICPCA) and the 2011 3rd International Symposium of Web Society (SWS), 26-28 October 2011, Port Elizabeth, South Africa, 2011b.

Yeratziotis, A., Pottas, D., van Greunen, D., 2012. A usable security heuristic evaluation for the online health social networking paradigm. International Journal of Human-Computer Interaction. 28(10), 678-694.

Zaibon, S. B., Shiratuddin, N., 2010. Heuristics evaluation strategy for mobile game-based learning. In: Proceedings of the 6th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technologies in Education, 127-131.

Zhang, J., Johnson, T. R., Patel, V. L., Paige, D. L., Kubose, T., 2003. Using usability heuristics to evaluate patient safety of medical devices. Journal of Biomedical Informatics. 36, 23-30.

Zhou, A. T., Blustein, J., Zincir-Heywood, N. 2004. Improving intrusion detection systems through heuristic evaluation. In: Proceedings of Canadian Conference on Electrical and Computer Engineering, 1641-1644.

Zuk, T. Carpendale, M.S.T., 2006a. Theoretical analysis of uncertainty visualizations. In: Proceedings of SPIE & IS&T Conference in Electronic Imaging, 606007.

Zuk, T., Schleiser, L., Neumann, P., Hancock, M. S., Carpendale, S., 2006b. Heuristics for Information Visualization Evaluation. In: Proceedings of AVI workshop on Novel Evaluation Methods for Information Visualization, 1-6.

Appendix 1. Detail of the 70 studies included in the review

No	Study	Domain	How heuristics were created	How heuristics were validated
1	Geest & Spyridaki (2000)	Web site	Employed four steps: 1) Inventory & review of existing literature, 2) Formulating draft heuristics, 3) Reviewing draft heuristics - each heuristic reviewed by a student group & overall overview by web professionals after use, 4) Revise and finalise.	N/A
2	Greenberg et al. (2000)	Groupware	Derived heuristics from the locales framework (Fitzpatrick, 1998) i.e. principles related to the nature of team work and how a locale (or place) can support it.	Authors evaluated a team work application. The strengths of the application & major and minor interface issues were identified. However, this was not extended to severity identification.
3	Baker et al. (2001), Baker et al. (2002)	Groupware with shared visual workspace	Derived the heuristics from the mechanics of collaboration (Gutwin & Greenberg, 2000) which was developed specifically for shared workspace.	Novice & expert evaluated two groupware applications. The followings were observed: 1) unique, out of scope & false positive usability issues; 2) average & consistency performance of individual evaluator, 3) aggregate performance, identification of each issue & identification of major vs minor issues.
4	Federoff (2002)	Video games	Used literature review to collate existing game heuristics. These were then compared to Nielsen 10 heuristics and the	N/A

			results of observation & interview with a game development team.	
5	Kantner et al. (2002)	Online documentation	Used the following to create heuristics: 1) Nielsen's 10 heuristics, 2) Existing guidelines for online documentation, 3) Learnability, and 4) Context of use (book metaphor and searching)	N/A
6	Reeves et al. (2002)	E-learning	Employed two steps: 1) Modified and expanded Nielsen's 10 through critical analysis & "brainstorming" sessions. 2) Applying the heuristics to evaluate an e-learning application.	Applying the heuristics to evaluate an e-learning application.
7	Berry (2003)	Notification system	Employed steps: 1) Collectively thinking usability issues related to notification system, 2) Categorise the issues into heuristics.	Experts evaluated three notification systems' usability using Nielsen's and author's heuristics were compared. The number of usability issues identified.
8	Mankoff et al. (2003)	Ambient display	Employed these steps: 1) Eliminated & modified Nielsen's (1994) heuristics and added new heuristics to suit ambient displays. 2) Running a pilot survey with ambient display designers, 3) Experts survey.	16 experts evaluated two ambient displays in a between subjects study design which compared Nielsen's and author's heuristics. It was unclear whether it was in between or within subject comparison. The followings were observed: 1) number of issues found, 2) severity of issues found.
9	Somervell et al.	Large screen	Employed critical parameters through the use scenarios-based	The followings were carried out: 1) The authors' heuristics were

	(2003), Somervell & McCrickard (2005)	Information exhibits	(to explore the possible design of five large screen information exhibits), producing claims which were then synthesised to generate heuristics.	compared with Nielsen (1994) and Berry (2003) and measurements (based on Hartson et al., 2001) was obtained. 21 experts, divided in three groups, were presented with usability issues of three large screen and was asked to rate the applicability of each heuristic to each issue found. 2) The heuristics were used by students & education experts (non-usability expert) to evaluate usability issues.
10	Zhang, et al. (2003)	Medical device	Combined usability heuristics from Shneiderman (1998)) and Nielsen (1994).	Four evaluators evaluated two medical devices. Usability issues and their severity were identified.
11	Ardito et al. (2004), Costabile et al. (2005)	E-learning (module & platform)	Employed the following steps: 1) Pilot study on e-learning to identify usability problems experienced by end-users, 2) Establishing four dimension based on the pilot study and literature study, taking into the context of use of e-learning, 3) Identification of criteria that characterise the effectiveness and efficiency of each dimension.	N/A
12	Desurvire et al. (2004)	Computer game	The heuristics were based on literature and review by playability experts and game designers.	The results of usability evaluation using the heuristics and user study were compared. Comparison on type of usability issues, number of issues & severity for both approaches were made.

13	Sutcliffe and Gault (2004)	Virtual reality	Derived from Nielsen (1994), taking into the context of VR, and their previous work on VR design principles (Sutcliffe and Kaur, 2000).	Experts used the heuristics to evaluate two systems in three studies. Usability issues and severity ratings were identified.
14	Zhou et al. (2004)	Intrusion detection system (IDS)	Employed the following steps: 1) Identified the primary goal of IDS and usability issues by conducted surveys on users i.e. network security administrators, 2) Modified and compared Nielsen (1994) heuristics to the corpus of usability issues.	The authors' and Nielsen's (1994) heuristics were used simultaneously by 12 experts to evaluate two applications. The usability issues were identified and compared with a master list of usability issues (with severity ratings) which were separately identified by the authors.
15	Dringus and Cohen (2005)	E-learning	Employed the following steps: 1) Collecting a corpus of usability issues based on evaluation from end-users and independent inspection of authors, 2) Combined the corpus of usability issues and identified heuristics categories and their specifics.	N/A
16	Pang et al. (2005)	Media asset management platform	Combined Nielsen's (1994) 10 heuristics and cultural guideline for usability (Komlodi et al., 2005).	Three experts used the authors' heuristics to evaluate a system and the results were compared to the results of scenario-based design and claim analysis by experts. Usability issues were identified.
17	Bertini et al.	Mobile	Employed the following steps: 1) Collecting previous studies	8 experts evaluated two applications in a between subjects study

	(2006)	computing	that documented usability issues related to mobile computing, 2) Individual and team analysis (by three experts) of usability issues & their categorisation to an abstraction level that were appropriate for developing/generating heuristics, 3) Harmonising the categorisation and their terminologies, 4) Individual and team further synthesis with Nielsen's (1994) heuristics and mobile computing context of use to produce heuristics; 5) Obtaining feedback from other experts on the adequacy of the heuristics proposed to refine the heuristics.	design which compared Nielsen's and author's heuristics (between subjects comparison). The followings were observed: 1) number of issues found, 2) severity of issues found, 3) time needed for evaluation.
18	Korhonen & Koivisto (2006)	Mobile games	Employed the following steps: 1) Apply modular approach (which includes usability as one of the modules), 2) For usability module, results of the mobile game context analysis, a review of Nielsen's heuristics and game design guidelines were used to create heuristics, 3) comparison of proposed heuristics to usability problems identified in a mobile game, 4) refinement of modular approach and their heuristics.	Between 2-4 evaluators evaluated five games using the authors' heuristics. Usability issues were identified.
19	Moraes & Silveira (2006)	Animated Agents	Nielsen heuristics were used as a basis. For each heuristics, the context of Artificial Intelligence was used to extend it by	The authors' heuristics were used by two evaluators to evaluate four interface agents.

			providing specific guidelines.	
20	Rocker and Haar (2006)	Pervasive game in smart home environment	The HEP (Desurvire et al., 2004) was analysed to identify elements, that are independent from the platform (and thus could be transferred for the specific domain) and those which are not (and thus be omitted or modified).	The heuristics were compared (mapped) to the result of focus groups with potential end-users who were involved as part of requirements elicitation. The heuristics were not revealed to end-users during the focus groups.
21	Zuk et al. (2006a), Zuk et al. (2006b)	Information visualisation	Employed an analysis of a variety of theories and framework of information visualisation (Bertin, 1983; Tufte, 2001; Ware, 2004) to abstract the heuristics.	Four experts evaluated a system and each used one of the three different sets of heuristics i.e. Shneidermans's (1996), Amar and Stasko's (2004), and Zuk et al. (2006a). Usability issues were identified and used to investigate interpretation, redundancy, and conflict in heuristics between the sets.
22	Clarkson and Arkin (2007)	Human-robot interaction (HRI) system	Employed the following steps: 1) Creating an initial list of HRI heuristics via brainstorming and synthesizing existing lists of potentially applicable heuristics (Nielsen, 1994; Scholtz, 2002; Sheridan, 1997; Baker et al., 2002; Mankoff et al., 2003), 2) Modifying the initial list based on pilot studies, consultation with other domain experts, and other informal techniques, 3) Refining the heuristics based on the findings from its	N/A

			application by experts.	
23	Conte et al. (2007), Conte et al. (2009)	Website	Employed the following steps: 1) Using web design perspective to expand on existing heuristics (Nielsen's (1994) heuristics), 2) Refining the heuristics by using it as part of an evaluation of a system and obtaining qualitative & quantitative feedback from experts.	Experts were divided into three groups (control group, group using the author's heuristics, and group using Nielsen's heuristics) to evaluate a system with a limited time. Usability issues found, qualitative feedback on heuristics, and time needed to evaluate were collected. Comparison was made on the performance of the groups.
24	Komarkova et al. (2007)	Web-based GIS	No description was provided.	The heuristics were used to evaluate 14 geo-websites. Scores of usability were obtained.
25	Aitta et al. (2008)	Library web service	Previous studies which focused on factors affecting the usability of library web sites, studies concerning usability in general (e.g. Nielsen, 1994) and experiences gained in authors' empirical studies were used to create heuristics.	The heuristics were informally used by experts to evaluate up to 15 library web sites. Usability issues were identified and overlapped with the results of user studies for one websites were also identified.
26	Kemp et al. (2008)	E-learning	Employed the following steps: 1) Expanded Nielsen's (1994) heuristics to include aspects of commercial web-sites (Nielsen, 2000) and issues particularly relevant to the learning appliance ubiquitous computing (attention/focus, adoption/flexibility, trsust/privacy, etc.), 2) creating a review checklist for each	N/A

			heuristics containing specific questions to inspect the system compliance with each heuristics., 3) refinement of the heuristics and checklists by using it (one evaluator) to inspect an E-learning application.	
27	Pinelle et al. (2008), Pinelle et al. (2009)	Single and Networked multiplayer games	Employed the following steps: 1) Collecting PC game reviews from a popular gaming website from various genres, 2) Identifying common usability issues and categorised the problems, 3) create heuristics from the categories of usability issues (by inversing the problem categories).	Experts used the heuristics to evaluate a game and usability issues and theirs severity issues were obtained (Pinelle et al., 2008). In Pinelle et al. (2009), 10 experts evaluated two games in a between subjects study design which compared Baker et al.'s heuristics and author's heuristics.
28	Sim et al. (2008), Sim et al. (2009)	Computer-based assessment application	Employed the following steps: 1) Identification of frequently encountered or severe issues from a corpus of usability problems (obtained from authors' previous studies) and literature studies, 2) Mapping the issues related to themes, 3) Extracting the heuristics and its description from the themes.	N/A
29	Afacan and Erbug (2009)	Building	Adopting universal design principles (The centre for Universal Design, 1997) as heuristics.	The heuristics were validated using three scenarios at a shopping mall and involved five experts who were asked to analyse the construction drawings.
30	Carvalho et al.	Health	Employed the following steps: 1) A systematic review of	An expert used the heuristics to evaluate a HIS application.

	(2009)	information system (HIS)	research articles to identify published research that outlines any potential harms from using HIS, 2) Thee HIS experts discussed the finding of the first phase and generate the heuristics, 3) Group the heuristics under categories.	Usability issues (violation of heuristics) were identified.
31	Collazos et al. (2009)	Interactive television (iTV)	Heuristics were proposed based on guidelines to design iTV application (no description was provided on how to create heuristics from the guidelines). The guidelines were taken from existing literature and the context of use of iTV.	N/A
32	Desurvire & Wiberg (2009)	Computer games	Employed the following steps: 1) Used the work of Desurvire et al. (2004) and discussion with game developers to generate 166 principles; 2) Conducted a survey, in which participants were required to experience both good and bad computer games, to assess the validity of the proposed principles.	N/A
33	Herrmann (2009)	Creativity groupware	The heuristics were created based on interviews with 12 people of various background and roles. The heuristics were created using a theme-based type analysis from the interview results.	N/A
34	Jo et al. (2009)	Massive	Employed the following steps: 1) Reviewed existing heuristics	N/A

		multiplayer online game (MMOG)	for game user interface, fun and flow factors, 2) Identified modules and key issues	
35	Papaloukas et al. (2009)	Computer games	Employed the following steps: 1) identification and classification of usability issues found in game reviews and existing studies, 2) Observation of end-users, 3) Categorisation of usability issues from step 1 and 2, 4) Creating heuristics that addresses the usability issues.	The heuristics were used by three experts to evaluate the user interface of a game and the recording of users (logging combined with thinking aloud protocol) while playing the game.
36	Singh and Wesson (2009)	Enterprise resource planning (ERP) application	Employed the following steps: 1) identified usability issues from previous studies and publications, 2) Identified common usability criteria from previous studies and categorise them, 3) Mapping usability issues to the categories and prioritising the categories as heuristics, 4) Expands each heuristics to include details to inspect.	The heuristics was evaluated by comparing its effectiveness with general usability heuristics (Nielsen). Three experts were involved and asked to review the application against each heuristics. They were also asked to identify usability problems.
37	Tsui et al. (2009, 2010)	Assistive robotic	The heuristics were created based on literature review related to existing guidelines/heuristics for people with disabilities, social robotics and framework based on model-human processor.	The heuristics was validated against Nielsen (1994) heuristics for a specific task. A total of four evaluators were involved, with two evaluators for the authors' and Nielsen's heuristics. Number of total non-duplicative errors and errors identified were compared.

38	Väänänen-Vainio-Mattila and Wäljas (2009)	Mobile web site	Employed the following steps: 1) Creating the heuristics based on a synthesis of a literature review and on informal analysis of existing three services, 2) Using heuristics to evaluate three applications (three experts per application), 3) Identifying main theme of usability problems and used this to update the heuristics.	Using heuristics to evaluate three applications (three experts per application).
39	Alsumait & Al-Osaimi (2010)	E-learning (for children)	Context was used as a means to create three categories of heuristics (usability, child usability, e-learning usability). For each category, existing guidelines, heuristics and checklist were adapted and consolidated.	The validation was performed by comparing the application of heuristics in evaluation to user testing on two E-learning systems
40	Fink et al. (2010)	Building	Employed the following steps: 1) reviewing of existing material, 2) interview with six nurses to gather perception of current patient room bathroom and problems they had experienced, 3) focus groups of nine nurses (role-play with a mock-up patient bathroom performing five tasks followed by a group discussion), 4) creation of heuristics and their description & details.	N/A
41	Forsell and	Information	Employed the following steps: 1) Identified existing heuristics	N/A

	Johansson (2010)	visualisation	and guidelines for information visualisations, including those from general usability (Shneiderman, 1996; Nielsen, 1994), 2) Listed the heuristics, 3) Gathered usability issues in the form of text and screen shots, 4) Six experts sifted through the usability issues, compared them with the heuristics list and judged how well each heuristic explained each problem (scale 0-5), 5) identified heuristics that were capable of identifying the highest percentage of usability problem.	
42	Hub and Čapková (2010)	Public administration portal	Employed the following steps: 1) identification of basic set of heuristics, 2) analysis of environment - familiarising with the concrete type of user interface and its specificity, 3) identifying individuals and team that would be assigned for creating heuristics, two teams should be created, the first team does step 4 and both teams do step 5, 4) heuristics creation, 5) evaluation of heuristics (aggregation of importance rating, rate of correspondence between the teams), and 6) refinement for final set of heuristic.	N/A
43	Karousos et al.	Groupware	Employed the following steps: 1) analysis of existing studies	N/A

	(2010)		related to usability issues, principles and system particularities, 2) identification of further usability issues through user observations, 3) description of how usability issues can be resolved through the heuristics, 4) categorisation of heuristics.	
44	Katre et al. (2010)	Touch screen-based ventilator system	The heuristics were proposed based on the results of interview with experts and also direct observation of how the ventilator was used etc.	The heuristics was used as part of evaluation with four evaluators and reliability of usability heuristics was established (comparing the four evaluators).
45	Omar et al (2010)	E-learning	Employed the following steps: 1) Initial study, 2) Propose heuristics by modifying existing heuristics (Nielsen, 1994; Nokelainen, 2006; Albion, 1999), 3) Expert reviews on heuristics usefulness by means of questionnaires, 4) Refinement of heuristics and weight assignment.	N/A
46	Omar and Jafar (2010a, 2010b, 2011)	Educational computer game	Existing heuristics that were related to computer games and issues related to pedagogical elements in e-learning were reviewed. This was then used to create five categories and assign heuristics in each category.	Five experts used the heuristics to identify usability issues and their severity.
47	Rusu, et al.	Grid	Employed the following steps: 1) Collect bibliography related to	Two studies were performed: 1) An expert inspected an

	(2010, 2011) and Jimenez et al. (2013)	computing	application domain, its characteristics and general or specific usability heuristics, 2) Identify important information from step 1, 3) Identify characteristics of usability heuristics the domain should have - use Nielsen (1994) if no heuristics existing, 4) Formally specify the heuristics & apply a standard template, 5) Validate the heuristics, 6) Refine the heuristics	application and usability issues identified were then mapped to Nielsen's and the authors' heuristics, 2) The same expert inspected an application using the authors' heuristics to see if the heuristics helped identified issues which were not found in study 1. Severities of usability issues were also obtained.
48	Salvador & Moura (2010)	Radiology transcription system	The heuristics were based on reinterpretation of Nielsen's (1994) heuristics, non-functional requirements of applications in the domain, and good practices in the domain.	The heuristics were used by one evaluator to evaluate an application.
49	Zaibon and Shiratuddin (2010)	Educational computer game	The heuristics were adapted from Korhonen and Koivisto (2006) and Koivisto and Korhonen (2006) by adding a new component related to learning.	The heuristics were used by end-users to evaluate an application. Scores for each component was obtained.
50	Jaferian et al. (2011), Jaferian et al. (2014)	IT security management (ISTM) application	Existing guidelines were compiled and existing studies were analysed based on activity theory (Kaptelinin and Nardi, 2006). The theory allowed interpretation of the rationale behind each guideline and abstraction of guidelines into heuristics.	28 experts evaluated an application in a between subjects study design which compared Nielsen's (1994) and author's heuristics. Usability issues and severitywere identified. Throughness, realibility, and validity (Hartson et al., 2001) and performance of evaluators were compared.
51	Lynch (2011)	Website (for	The heuristics were created based on a literature review of	Task performance (time and error) of end-users on three website

		elderly)	guidelines for website that wer aimed for general website and older end-users. Based on suvey from end-users, scoring system for heuristics were then developed.	(good, medium and low usability) were compared with heuristics scores obtained by two experts evaluation.
52	Muñoz et al. (2011), Muñoz & Chalegre (2012)	Virtual world	See study no. 47	6 experts evaluated two applications in a between subjects study design which compared Nielsen's (1994) and author's heuristics. Usability issues and severitywere identified.
53	Sivaji et al. (2011)	E-government website	Nielsen's heuristics were used as basis for heuristics expansion. No detail explanation was given.	The heuristics was used by experts to evaluate e-Government website at different stages (three) with two to three experts being involved in each stage.
54	Yeratziotis, et al. (2011a, 2011b)	Security in online Health Social Networks	Employed the following steps: 1) reviewing existing literatures and identify emerging themes, 2) naming high level heuristics from identified themes, 3) modifying existing heuristics to cratea checklist items, 4) grouping checklist items under the heuristic, 5) reviewing and refining by presenting the heuristics and checklist to experts.	N/A
55	Bowie (2012)	Podcast (sound)	The heuristics were developed based on inference of usability definition, existing heuristics and podcasts context of use. The	N/A

			heuristics were then expanded to include guidelines by evaluating 11 parts of a podcasts.	
56	Inostroza, et al. (2012a), Inostroza et al. (2012b), Inostroza et al. (2013), Inostroza & Rusu (2014)	Touch screen-based mobile device	See study no. 47	Validation of heuristics performed in two stages resulted in refinement of heuristics at each stage. In the first stage, four experts inspected an application in a between subjects study design which compared Nielsen's (1994) and author's heuristics. Number of issues, severity and type of issues were compare. The heuristics were also evaluated w.r.t. easines to understand by 27 experts in an online survey. In the second stage, design principles which were extracted from the websites of four operating system developers were mapped to heuristics.
57	Soomro et al. (2012)	Mobile games	Employed the followint steps: 1) Identify major issues and compared heuristics for computer and mobile games, 2) Categorised issues into categories by Korhonen (2006), 3) Questionnaire and issues to explore the issues from game users perspective, 4) Creation of heuristics	N/A
58	Liao and Shen (2012)	Educational computer	Heuritics were proposed based on existing guidelines (Ang et al., 2008) and modified based on feedback from experts. No	The heuristics were used by experts to evaluate a game. Usability issue and compliance rating was obtained.

		game	detailed description was provided.	
59	Alotaibi (2013)	Website	No detail information was given on how the heuristics list was generated (except mentioning that these were based on existing liteature).	The heuristics were used to evaluate 12 universities websites and involved 30 evaluators for each set of heuristics. The fulfillment of websites for each heuristic was rated against the following four-point scale (0-4): not applicable, not fulfilled, partially fulfilled and fully fulfilled.
60	Al-Azawi et al. (2013)	Computer game	No clear description was provided.	N/A
61	AlRoobaea et al. (2013)	Social networking website	Employed the following steps: 1) reviewed past literatures related to heuristics evaluation in the target domain, 2) Mini usability studies with end-users to identify users' perspective, requirements and expectations, 3) Focus group with experts' especially related to findings in step 1 and 2 to establish usability issues, 4) Development of heuristics by analysing outcome from the previous three steps.	Two studies were performed, involving experts and end-users. In experts study (eight expert divided into two groups) which compared the use of Nielsen's and authors' heuristics, the following were compared: usability issues & their severity, time and cost required, and measures from (Hartson et al., 2001), ease of use using SUS. One group used Nielsen's on the first two websites and authors' heuristics on the last website, while the other group used author's heurisics for the first 2 websites and Nielsen's heuristics on the last website. For the user study, usability issues were identified and would be compared with the

				result of experts study.
62	Diaz et al. (2013)	Intercultural website	See study no. 47	6 experts evaluated two applications in a between subjects study design which compared Nielsen's (1994) and author's heuristics. Usability issues and severitywere identified.
63	Grice et al. (2013)	Website	Existing heuristics (Nielsen, 1994; Hargis et al., 1998) were expanded iteratively until they fully described the usability and user experience in five broad use cases. For each case study, experience and usability issues that were not covered by heuristics were identified, and then additional heuristics were proposed and agreed.	N/A
64	Kuparinen et al. (2013)	Mobile map application	Employed the following steps: 1) performed literature review on general heuristics (e.g. Nielsen, 1994), existing heuristics Pinelle et al. 2008, Alsumait & Al-Osaimi 2009, Jaferian et al. 2011) & methods used to generate them for the domain, 2) Applied context of used (cartography) to existing heuristics to generate heuristics, 3) Four experts evaluated an application in a between subjects study design which compared Nielsen's (1994) and author's heuristics. Applicability and intelligibility of	Four experts evaluated an application in a between subjects study design which compared Nielsen's (1994) and author's heuristics.

			heuristics were collected and used to refine the heuristics.	
65	Neto & Pimentel (2013)	Mobile-based application	Employed the following steps: 1) Collected usability issues by inspections of four existing applications by experts, 2) grouped the identified issues into categories and associated/approximated them with Nielsen's (1994) where possible, 3) Enriched the heuristics by adding more details	10 experts evaluated an application in a between subjects study design which compared Nielsen's (1994) and author's heuristics. Usability issues were identified.
66	Solano et al. (2013)	Interactive television	See study no. 47	10 experts evaluated an application in a between subjects study design which compared Nielsen's (1994) and author's heuristics. Usability issues were identified. Usability test with 5 users were conducted to investigate the severity of issues that were found specifically using Nielsen's and authors' heuristics.
67	Al-Razgan et al. (2014)	Mobile-based application (for elderly)	Employed the following steps: 1) Converting existing guidelines into usability issues, 2) Group similar issues into categories, 3) Translated usability issues into heuristics which were inspired by existing relevant heuristics, 4) Elaborate heuristics into questions as checklists	The heuristics were used by four experts to inspect six applications with two persona assigned as end-users. Usability issues and feedback on heuristics were obtained.
68	Joyce and Lilley (2014)	Smartphone mobile	Employed the following steps: 1) Performed a literature review to establish the baseline of what heuristics should be like, 2)	N/A

		application	Proposing the heuristics and performed survey on experts to gauge each heuristics usefulness and relevance, 3) Modification of heuristics	
69	Maike et al. (2014)	Natural User Interface (touch-less, body-based interface)	Employed the following steps: 1) Performed a systematic literature review and identified possible heuristics which was then entered into a predefined format, 2) identified patterns of heuristics and categorised them	N/A
70	Paz et al. (2014), Quiñones et al. (2014)	Transactional website	See study no. 47	Six groups of 2-4 evaluators were involved to evaluate independently a transactional website. The followings were observed: correct and incorrect association of usability problems to heuristics, distribution of usability problems across heuristics, common problems identified among groups.

Appendix 2. Detail of the proposed heuristics and their effectiveness

No	Study	The proposed heuristics	Effectiveness
1	Geest & Spyridaki (2000)	Expressed in two forms, a long form (contained explanation, support and examples) and a quick list. Details of heuristics were not provided.	N/A
2	Greenberg et al. (2000)	Short description was given for the five proposed heuristics. As the focus of heuristics was on teamwork, no overlap with Nielsen's (1994) heuristics was found.	No comparison was made with other method.
3	Baker et al. (2001), Baker et al. (2002)	Explanation of each heuristics (a total of 8) & how it is typically realised & supported in groupware was provided. As the focus of heuristics was on teamwork, no overlap with Nielsen's (1994) heuristics was found.	The results of validation were theoretically compared to Nielsen's heuristics evaluation published outcome. The same level of performance was shown.
4	Federoff (2002)	Heuristics (a total of 40) were divided into three categories (Interface, Mechanics and Play) and for each category heuristics were stated without additional explanation. Nearly all of Usability heuristics (of a total of 14) were in accordance with Nielsen's (1994) heuristics.	N/A
5	Kantner et al. (2002)	For each heuristic (a total of 10), the definition is expanded by adding details under four different dimensions (structure, presentation, dynamics, content) as appropriate. 7 of the heuristics were in	N/A

		accordance with Nielsen (1994) heuristics.	
6	Reeves et al. (2002)	For each heuristic (a total of 15), short definition and sample questions to be asked by evaluators were given. 8 of the heuristics were in accordance with Nielsen's (1994) heuristics.	N/A
7	Berry (2003)	Detailed description was given for each heuristic (a total of 8). 4 of the heuristics were in accordance with Nielsen's (1994) heuristics.	There was no significant difference in effectiveness between both sets of heuristics. However, none of author's heuristic was seen as inapplicable by evaluators.
8	Mankoff et al. (2003)	Short description was given for each of the eight heuristics. 5 of the heuristics were in accordance with Nielsen (1994).	In average, an evaluator using domain specific heuristics found more usability problems than that using Nielsen. However, this was not significant. To cover all usability issues, half of Nielsen's and all but one author's heuristics were needed. None of authors' heuristic was seen as inapplicable by evaluators.
9	Somervell et al. (2003), Somervell & McCrickard (2005)	For each heuristic (a total of 8) details of what should or should not be done were given. No overlap with Nielsen's (1994) heuristics was found as the author's heuristics were at much lower level.	The authors' heuristic showed higher thoroughness, validity, effectiveness and reliability than Nielsen's. Both novice and education experts could identify usability issues by using the heuristics even although the education experts initially found it difficult.
10	Zhang, et al. (2003)	Semantic tags, names, general description and specific information	Violations of usability heuristics were identified using the

		(i.e. things to look out for) were provided for each of the 14 heuristics. 10 and 4 of heuristics mapped accordingly to Nielsen's (1994) and Shneiderman (1998).	proposed heuristics.
11	Ardito et al. (2004)	Four dimensions (presentation, hypermediality, application proactivity, user's activity) with 16 and 8 criteria for e-learning platform and module, respectively. For each criterion, detail information (guideline) was given. Only a significantly small number of criteria & guideline were in accordance with Nielsen's (1994) heuristics.	N/A
12	Desurvire et al. (2004)	There were a total of four dimensions (Game Play, Game Story, Mechanics and Usability). Majority of the heuristics for Usability dimension (from a total of 12) were in accordance with Nielsen's (1994) heuristics albeit specifically phrased to suit the context of computer game.	The heuristics identified more usability issues than user testing with some overlapped observed between the two. The heuristics identified general interface design issues while the user testing identified specific problems.
13	Sutcliffe and Gault (2004)	A total of 12 heuristics with a short description for each. Seven of the heuristics were in accordance with Nielsen's (1994) heuristics.	Increasing number of evaluators did not discover many more error. Difficulties in interpreting the heuristics and unsuitability of some of heuristics were reported.
14	Zhou et al. (2004)	Six heuristics were proposed. Only the name of heuristics was	The authors' heuristics identified more usability issues than

		provided. Four of them were in accordance with Nielsen's (1994) heuristics.	Nielsen's (1994) heuristics, especially for moderate to highly severe issues.
15	Dringus and Cohen (2005)	A total of 13 heuristics were created with each heuristic was expanded into a set of further checklist. Eight of them were in accordance with Nielsen's (1994) heuristics.	N/A
16	Pang et al. (2005)	A total of 14 heuristics were created, no description was provided for each heuristic. Ten of them were taken from Nielsen's (1994) heuristics and three of them taken from Komlodi et al. (2005).	Scenario based design and claims analysis identified causal relationships between system features and the usability of a user's interaction but was more time consuming.
17	Bertini et al. (2006)	A total of 8 heuristics were created with description provided for each heuristics. Eight of Nielsen's (1994) heuristics were adopted, expanded and modified to match mobility computing.	The authors' heuristics detected more issues than Nielsen's heuristics but also more time consuming. Nielsen's heuristics was better at identifying severe issues whereas authors' heuristics identified more minor and major issues.
18	Korhonen & Koivisto (2006)	Heuristics were contained within three modules (mobility, gameplay and game usability). Game usability consisted of 12 heuristics with 8 of heuristics reflected Nielsen's (1994) heuristics. Detailed description of each heuristic was not provided.	The results indicated that these heuristics are useful in identifying playability (including usability) issues in mobile games.
19	Moraes & Silveira (2006)	A total of ten heuristics (exactly the same terminologies as Nielsen's (1994) heuristics) were proposed with each heuristic contained	The usability heuristics allowed identification of which agents were better than others.

		detailed information in the form of specific guidelines.	
20	Rocker and Haar (2006)	Only some of heuristics were provided in the study, of those provided, a clear description was given. Majority of the heuristics for Usability dimension (from a total of 12) were in accordance with Nielsen's (1994) heuristics albeit specifically phrased to suit the context of game.	The results show that some of the focus groups results could be mapped with the heuristics while some were not.
21	Zuk et al. (2006a), Zuk et al. (2006b)	A total of 13 heuristics were proposed, no detailed description was provided. As the focus of heuristics was on information visualisations, no overlap with Nielsen's (1994) heuristics was found.	The comparison showed that the evaluation process and results have a high dependency on the heuristics and the types of evaluators chosen.
22	Clarkson and Arkin (2007)	Eight heuristics were created and description for each heuristic was provided. Six of the heuristics were modification of existing heuristics (Nielsen, 1994; Scholtz, 2002; Mankoff et al., 2003).	N/A
23	Conte et al. (2007, 2009)	A total of 10 heuristics which were adopted and modified from Nielsen's (1994) were proposed. Description of each heuristics and relevant design perspectives and their hints were provided.	The authors' heuristics were as nearly twice as effective as Nielsen's (1994) heuristics and as efficient as Nielsen's (1994) heuristics.
24	Komarkova et al. (2007)	There were a total of 9 high level heuristics with each heuristics expanded further into a set of checklist (in the form of specific questions) totalling into 138 checklists. 5 of the high level heuristics	Total points were used to compare the usability of GIS applications. Applications with most usability issues reached the same position in the usability testing and in heuristic evaluation.

		were modification of Nielsen's (1994) heuristics. A range of scores were assigned for each checklist.	
25	Aitta et al. (2008)	Nine heuristics were created and really detailed description was provided for each heuristic. Five of the heuristics were adapted from Nielsen's (1994) heuristics.	There was 37% overlapped of usability issues identification with user studies.
26	Kemp et al. (2008)	A total of 18 heuristics were proposed with 10 heuristics were adopted from Nielsen's (1994). Specific questions were not provided.	N/A
27	Pinelle et al. (2008), Pinelle et al. (2009)	10 heuristics were created in Pinelle et al. (2008) with description provided for each heuristic. Five of the heuristics were in accordance with Nielsen's (1994). 10 heuristics were also created in Pinele et al. (2009) with description provided for each heuristics. None of the heuristics corresponded with Nielsen's (1994).	Pinelle et al. (2009) found that their heuristics performed better in identifying more issues. The severity rating of issues for both sets was similar.
28	Sim et al. (2008), Sim et al. (2009)	A total of 11 heuristics were created and description of each heuristic was provided. 3 were based on Nielsen's original set, 2 were modifications, and 6 were new heuristics specific to CAA.	N/A
29	Afacan and Erbug (2009)	A total of 7 heuristics were proposed with description and design considerations for each heuristics provided. Two of the heuristics were similar to Nielsen's (1994).	Issues that could not have been detected solely through an analysis of the construction drawings were found. There was a substantial evaluator effect. Therefore, more than one and

			interdisciplinary evaluators were needed.
30	Carvalho et al. (2009)	A total of 38 heuristics divided into four categories (workflow, content, safeguards function) were proposed. Some of heuristics were re-used in different categories. Heuristics were so specific, no overlap with Nielsen's (1994) heuristics was found.	Usability issues were identified.
31	Collazos et al. (2009)	A total of 13 heuristics were proposed and description for each heuristic was provided. Ten of the heuristics were adopted and modified from Nielsen's (1994).	N/A
32	Desurvire & Wiberg (2009)	The heuristics (broken down further into 48 principles) were intended for three genres of computer games (real time strategy, action adventure and first person shooter) and divided into three categories (Game Play, Coolness/entertainment/humour/emotional immersion, Usability & game mechanics). 7 of the heuristics (from a total of 9) were in accordance with Nielsen's (1994) heuristics.	N/A
33	Herrmann (2009)	Five heuristics were proposed. No overlap with Nielsen's (1994) heuristics was found due to their focused in supporting creativity collaboration.	N/A
34	Jo et al. (2009)	Eight heuristics were categorised into three modules (usability,	N/A

		playability, and enjoy ability). Each heuristics was expanded into a list of checklist (in the form of questions). Five of the heuristics were adopted from Nielsen's (1994).	
35	Papaloukas et al. (2009)	A total 10 heuristics were proposed and description for each heuristic was provided. Four of the heuristics were in accordance with Nielsen's (1994).	The heuristics helped identification of usability problems. However, overlapping between different heuristics and broad coverage of some heuristic were reported by experts.
36	Singh and Wesson (2009)	Five heuristics were identified and description for each heuristics was provided. Specific directions on what to inspect for each heuristic were provided.	Based on severity ratings of heuristics, the authors' heuristics and Nielsen's (1994) identified and focused on different usability issues.
37	Tsui et al. (2009, 2010)	A total of nine heuristics were proposed with each heuristic was expanded with detailed questions that were based on the results of the literature review. Two of the heuristics were adopted from Nielsen's (1994).	The authors' heuristics found more issues than Nielsen's heuristics, with only 18% overlapped between the two.
38	Väänänen-Vainio-Mattila and Wäljas (2009)	A total of nine heuristics were proposed (updated and modified from initial of six. Description for each heuristics, containing purpose, applicability and examples of its pragmatics and hedonic aspect was provided.	N/A
39	Alsumait & Al-Osaimi	Nielsen's usability was adopted for usability category. The description	There was 32.3% overlapped issue between heuristics evaluation

	(2010)	for each heuristic was adjusted and modifed to suit the domain.	and user testing. 83.3% issue found using heuristic were not found using usability testing.
40	Fink et al. (2010)	Six heuristics were identified and checklists (in the form of questions) were created for each heuristic. No overlap with Nielsen's (1994) heuristics was found due to their high focus on the domain.	N/A
41	Forsell and Johansson (2010)	A total of 10 heuristics were proposed with description for each heuristic kept original. Four of heuristics were adopted from Nielsen's (1994).	N/A
42	Hub and Čapková (2010)	A total of seven heuristics were proposed. Heuristics were expanded in details and resulted in 92 specific questions that are relevant to the domain. Six of the heuristics matches with Nielsen's.	N/A
43	Karousos et al. (2010)	The heuristics contained two major parts, user interface and system particularities. For user interface, Nielsen's (1994) were adopted without any modification. For system particularities, nine heuristics were proposed. Short description was provided for each heuristics.	N/A
44	Katre et al. (2010)	16 heuristics, assigned into five categories, were proposed. Each heuristic was really specific and categorical responses (e.g. yes, no) were required. No overlap with Nielsen's (1994) heuristics was found	No significant difference on reliability was found between the author's heuristics and reported results from other methods (Zhang et al., 2003)

		due to their high focus on the domain.	
45	Omar et al (2010)	The usability heuristics were categorised into 4 (interface, pedagogical, content, and suitability). For each category, relevant heuristics were presented. No further description of each heuristic was provided. Three of the heuristics under interface category were adapted from Nielsen's (1994).	N/A
46	Omar and Jafar (2010a, 2010b, 2011)	Heuristics were divided into five categories (interface, pedagogical, content, multimedia and playability). No further description of each heuristic was provided. Three of the heuristics under interface category were adapted from Nielsen's (1994).	N/A
47	Rusu, et al. (2010), Rusu et al. (2011), Rusu et al. (2012), Jimenez et al. (2013)	A total of 12 heuristics were proposed. In addition to description, examples, benefits, etc., each heuristic was also expanded to include a list of checklists. Seven of the heuristics were in accordance with Nielsen's (1994) heuristics.	Some specific domain usability problems were not detected by using only Nielsen's heuristics.
48	Salvador & Moura (2010)	A total of 6 heuristics were provided. Each heuristic had its description and checklists. Three of the heuristics were in accordance with Nielsen's (1994) heuristics albeit titles were modified.	Usability issues were identified.
49	Zaibon and Shiratuddin	A total of 27 heuristic were proposed and divided into four	Scores for each heuristics in each component were reported.

	(2010)	components (usability, mobility, game play and learning). No description was provided for each component.	
50	Jaferian et al. (2011), Jaferian et al. (2014)	A total of 7 heuristics were provided with description for each heuristic provided. Only one heuristic was in accordance wth Nielsen's (1994).	The authors' heuristics resulted in identification of more severe issues, fewer overlaps between indvidual experts, than Nielsen's.
51	Lynch (2011)	A total of 32 heuristics were categorised into 4 themes: navigation, accessibility, readability and content. No desription was provided for each heuristic.	The scores from heuristics was in accordance with end-user performance i.e. high score showed shorter task completion time and error.
52	Muñoz et al. (2011), Muñoz & Chalegre (2012)	A total of 16 heuristics were proposed. In addition to definitions, examples, etc., each heuristic was also expanded into a set of checklist (a total of 53 items). Seven of the heuristics were in accordance with Nielsen's (1994) heuristics.	The authors' heuristics identified more issues thanNielsen's, although some issues identified in Nielsen's were not found using authors' heuristics.
53	Sivaji et al. (2011)	A total of 12 heuristics were proposed with description for each provided. Five of the heuristics were in accordance with Nielsen's (1994).	Heuristics that were specific for the domain identified 17% of the overall usability issues.
54	Van Gruenen et al. (2011), Yeratziotis et al.	A total of 13 high-level heuristics with individualized checklist items that help examine usable security	N/A

	(2012)		
55	Bowie (2012)	A total of 7 heuristics were proposed. Description for each heuristics was provided and checklists (guidelines) for each heuristics for different part of podcast were also provided.	N/A
56	Inostroza, et al. (2012a), Inostroza et al. (2012b), Inostroza et al. (2013), Inostroza & Rusu (2014)	A total of 14 heuristics were proposed. Each heuristic had its own description and other detailed information. Twelve of the heuristics were in adopted from Nielsen's (1994) with only minor modification performed on their titles.	In the first stage, the new heuristics showed that it could identified slightly more usability problems than Nielsen and were easy to understand by experts. In the second stage, six of the 43 design principles could not be associated with any heuristic and two new heuristics, "user experience" and "cognitive load", were added.
57	Soomro et al. (2012)	A total of 10 heuristics, divided into four categories were proposed. Short description was provided for each.	N/A
58	Liao and Shen (2012)	A total of 36 heuristics were divided into 6 categories. Two categories (game interface and navigation) contained two of heuristics from Nielsen's (1994).	Compliance rating to heuristics was reported.
59	Alotaibi (2013)	The heuristics contains seven heuristics with 58 questions (checklist). For each heuristic, a short description was provided. Two of the heuristics were in accordance with Nielsen's (1994) heuristics.	The usability rating for websites and each heuristic was identified.

60	Al-Azawi et al. (2013)	Heuristics were divided into five categories (quality, playability, usability, enjoyment and mobility) with a total of 26 heuristics for usability. The usability criteria contained heuristics that were adopted from existing heuristics (e.g. Korhonen & Koivisto, 2006; Desurvire, 2004; Federoff, 2002).	N/A
61	AlRoobaea et al. (2013)	A total of 26 heuristics were proposed and categorised based on 7 types of usability issues areas. No additional description was provided for each heuristic.	The authors' heuristics identified all *real* problems that were discovered by user study and expert study using Nielsen's (1994) heuristics with greater efficiency, thoroughness effectiveness, validity and reliability. Researchers identified real problem by extracting issues identified in three methods, removed all false positive ('not real') problems, 'evaluator subjective' problems and duplicated problems. The issues agreed on were added as unique master problem list and any problems on which the evaluators disagreed were removed.
62	Diaz et al. (2013)	A total of 13 heuristics were proposed. For each heuristics, detailed information was provided (description, examples, benefits, checlist, etc.). 10 of the heuristics were adapted (with the same titles) from Nielsen's (1994).	The authors' heuristics found more problem than Nielsen's and higher averaged severity. Checklists of heuristic which did not help identified any problem was combined with other heuristics.

63	Grice et al. (2013)	A total of 10 heuristics were proposed with each heuristic contained several sub heuristics (a total of 36). For each sub heuristics, metrics (product, behavioural, survey) were added. A small number of of the sub-heuristics were adopted from Nielsen's (1994).	N/A
64	Kuparinen et al. (2013)	Ten heuristics were proposed with each had its own description. 9 of the heuristics were adapted from Nielsen's (1994).	N/A
65	Neto & Pimentel (2013)	A total of 11 heuristics were proposed with short description and the literature source provided. Six of the heuristics were adapted from Nielsen's (1994).	The authors' heuristics found more issues than Nielsen in all categories (cosmetic, minor, major and catastrophic).
66	Solano et al. (2013)	14 heuristics were proposed with each heuristic contain specific information on what it covers. A checklist was created to support the usability heuristics. Nine of the heuristics were adapted from Nielsen's (1994), some titled were modified.	The authors' heuristics found more issues than Nielsen. Some issues were identified by both heuristics and some were not.
67	Al-Razgan et al. (2014)	A total of 12 heuristics were proposed with each has a set of questions (checklists). Six of the heuristics were adapted from Nielsen's (1994) albeit with different titles.	The authors' heuristics identified issues and was easy to understand and relevant for the domain.
68	Joyce and Lilley (2014)	A total of 13 heuristics were proposed with each had its own description. Five of the heuristics (titled changed_ were adapted from	N/A

		Nielsen's 91994)	
69	Maike et al. (2014)	A total of 23 heuristics, divided into four categories were proposed. For each category, description and source were given.	N/A
70	Paz et al. (2014), Quiñones et al. (2014)	A total of 15 heuristics were proposed. Additional information (ID, name, short description, examples of violation and compliance etc.) for each heuristic was provided. 12 of the heuristics were adopted from Nielsen's (1994)	Some problems were wrongly associated with heuristics. Therefore, a more clear and detailed definition of some heuristics were needed, checklist would likely help addressing this.