

Heuristic Evaluation for Serious Immersive Games and M-instruction

Neil Gordon, Mike Brayshaw and Tareq Aljaber

Department of Computer Science, University of Hull,
Hull, HU6 7RX, United Kingdom
m.brayshaw@hull.ac.uk;
n.a.gordon@hull.ac.uk;
t.aljaber@2013.hull.ac.uk

Abstract. Two fast growing areas for technology-enhanced learning are serious games and mobile instruction (M-instruction or M-instruction). Serious games are ones that are meant to be more than just entertainment. They have a serious use to educate or promote other types of activity. Immersive Games frequently involve many players interacting in a shared rich and complex – perhaps web-based - mixed reality world, where their circumstances will be multi and varied. Their reality may be augmented and often self-composed, as in a user-defined avatar in a virtual world. M-instruction and M-instruction is learning on the move; much of modern computer use is via smart devices, pads, and laptops. People use these devices all over the place and thus it is a natural extension to want to use these devices where they are to learn. This presents a problem if we wish to evaluate the effectiveness of the pedagogic media they are using. We have no way of knowing their situation, circumstance, education background and motivation, or potentially of the customisation of the final software they are using. Getting to the end user itself may also be problematic; these are learning environments that people will dip into at opportune moments. If access to the end user is hard because of location and user self-personalisation, then one solution is to look at the software before it goes out. Heuristic Evaluation allows us to get User Interface (UI) and User Experience (UX) experts to reflect on the software before it is deployed. The effective use of heuristic evaluation with pedagogical software [1] is extended here, with existing Heuristics Evaluation Methods that make the technique applicable to Serious Immersive Games and mobile instruction (M-instruction). We also consider how existing Heuristic Methods may be adopted. The result represents a new way of making this methodology applicable to this new developing area of learning technology.

Keywords: Heuristic Evaluation; Serious Games; M-instruction

1 Introduction

Throughout education, there is a growing focus ways to improve student engagement [2], which may be through utilising different pedagogic approaches [3] or technologies

[4, 5, 6, 7, 8]. Serious Games are one approach to improve engagement through the benefits of a typically fun based, interactive environment where the learning is embedded within playful activity; the game has the aim of delivering some knowledge and/or skill as the student progresses through the game [4].

The technological developments – primarily through computer and related IT – have led to Technology Enhanced Learning (TEL) [9], which encompasses software of computer based learning approaches, as well as the internet based eLearning.

One particular approach to technology enhanced learning is to utilise immersive environments – ones in which the learner is placed in a simulation of the real world [10], so that the student behaves as though in a real world context which is identified as improving learning in various ways. Combining the notion of serious games and immersive learning, serious immersive games utilise game-like (or actual game) environments to provide learning opportunities.

Computer Based Instruction in its general sense leads on to the more specific concept of Mobile-instruction (M-instruction) - also known as mobile learning (M-learning). In this paper, we consider such M-instruction as learning that utilises mobile devices [11]. As these have become increasingly sophisticated, the distinction between M-instruction and other e-learning is reducing. An example of the convergence of eLearning and M-instruction is that tablet and mobile devices can now routinely access server based material. Modern learning environments support mobile friendly formats and interfaces. However, there are some distinct characteristics – particularly when considering the interface specific elements of eLearning designed for a computer (desktop or laptop), as opposed to the types of interaction that are more suited to mobile devices.

The game based and mobile approaches to software to support learning are examples of ubicomp. Ubiquitous computing and applications (also known as pervasive computing, or as ubicomp) relates to the concept that computing (and in particular, computer science and software engineering) appears everywhere [12]. Users may interact with this in a variety of forms, and the issue of evaluating such ubicomp applications becomes complex – since there are a myriad of platforms and instances of use. Moreover, as noted by Gordon [13], where ubicomp is applied to give students choice (i.e. flexible education), the location and time of study, so the context of use, will vary. Thus, the issue of evaluation becomes a multi-dimensional one. One potential approach that is explored in this paper is heuristic evaluation.

To support changes and innovations in teaching and learning, software and systems have been, and continue to be, developed. For software and system development, there are a number of potential ways to help in designing and evaluating them, from functional behaviour through to the acceptability to the users. Regarding the learning software considered in this paper, this may be evaluated as general software, through usability requirements, or, as considered in Brayshaw et al [1], may be evaluated through heuristic evaluation. Usability testing is well established in software engineering, that focusses on the user interaction with the system [14]. This typically requires that a scenario be created for the users to work within, carrying out specific activities, which can be observed and measured to indicate the usability of the system. The clear difficulty here, when considering ubicomp – is that there is a multitude of possible scenarios, and each may have its own characteristics.

Heuristic Evaluation [14] offers an approach that uses evaluators to identify potential issues in the system, in relation to a set of identified principles that reflect usability characteristics.

2 Utilising Heuristic Evaluation as an approach to evaluation pedagogic software: an empirical application

Heuristic Evaluation is an informal method of usability analysis that lends itself to domains like serious games and MLearning. Whilst we strongly support the use of more traditional, empirical methods there are circumstances when this is not always possible. The use of traditional evaluation methodologies presupposes that we have access to the end users. In an education context, this could be in a direct classroom scenario or by bringing learners into a laboratory under controlled conditions. However, with modern computer connectivity our “end users” can literally be anywhere. Indeed the underlying presupposition of MLearning is that they are anywhere and mobile to boot! Therefore, we have to develop ways of evaluating our solutions before they go out. Heuristic evaluation allows pedagogical solutions to be tested at source – before they are shipped – by getting experts to evaluate the solution ahead of time.

2.1 Benefits of Heuristic Evaluation

Four major advantages of Heuristic Evaluation are

- Heuristic Evaluation, as opposed to a traditional evaluation, is cheap and providing the relevant expertise is readily to hand, easy to perform.
- The task is itself easy to grasp and once explained the required expertise from the identified cognoscente is usually happily given.
- The planning and control of conventional evaluation is not needed, something we will exploit in the work outlined here.
- It is useable through the project development lifecycle. It thus does not need a completed system. From first design, to iterative prototyping stages, the final deliverables it is a usable evaluation methodology.

2.2 Heuristics for software design

The most received version of Heuristic Evaluation is by Neilson [15, 16, 17]. It consists of a series of heuristics or rules of thumb that advise on design. They are as follows

- Transparency of system status – can the user see what state they are in?
- Correlation between the system and the real world – the desktop metaphor is frequently used here as a good example. It gives a good mapping between the computer and what they are trying to achieve in the real world.
- User Control and Freedom

- Consistency and Standards. Consistency in a user interface is clearly vital as are the application of relevant standards.
- Error Prevention. Can errors be anticipated and designed out of the equation.
- Recognition rather than recall. Having to remember or recall is problematic. If users can work things out live - this is better.
- Flexibility and Efficiency of Use
- Aesthetic and Minimalist Design
- Help Users Recognise, diagnose, and recover from errors
- Help and Documentation

2.3 Heuristics for educational applications

In the context of educational evaluation, [18 19] added the following:

- Feedback and designer/ learner models (Squires and Preece):- In an education context, feedback is incredibly important. It needs to be cogent and timely.
- Cosmetic authenticity (Squires and Preece):- avoid interface components that the learner could misinterpret.
- Representational forms (Squires and Preece):- The interface should place a “low cognitive demand on the learner and functionality should be obvious” (Squires and Preece). The system should be transparent and encourage learning.
- Multiple views/representations (Squires and Preece):- does the learning software support different forms of learning? Is there one content model or can others be supported?
- Interaction flow (Squires and Preece):- extrinsic feedback e.g. error messages can cause distractions. Is there a consistent and uninterrupted flow to learning?
- Navigation (Squires and Preece):- can the learner easily navigate through their learning episodes with appropriate feedback given at critical points on this journey?
- Learners Control and Self- Directed learning (Squires and Preece):- can learners express their autonomy and ownership of their journey?
- Subject Content [19]:- The preamble and context setting should be relevant to the questions and tutorials and to the appropriate skill level. Is choice of media delivery right and addresses targeted learning outcomes?
- Assessment (Benson et al):- is self-assessment available and the feedback to that assessment at the correct level. In what terms can we look at the quality and content of the assessment and feedback?

These heuristics are the starting point to evaluation. Based up these we are in a position to reflect critically upon what would otherwise be hard to evaluate empirically software solutions.

2.4 Existing Work Utilising Heuristic Evaluation

As reported in [1] we have used Heuristic Evaluation extensively and successfully in the past. In particular, we have used it to evaluate a Semantic Web Based Personalised VLE [20] that looked to semantically synthesise multiple sources of media to produce a personalized learning experience; for evaluating software for those with disabilities [21], and as a design tool in an evolving iterative prototyping tutoring system for teach computer programming [22]. Each time it provided a flexible tool to evaluate and reflect upon the work undertaken. It is in this context that we sought to again use this technique and apply it in contexts where simple end user evaluation is less straightforward.

3 Technology Enhanced Learning through...

3.1 Serious Immersive Games

Introduction to serious immersive games.

Serious games are games intended to do more than simply entertain; they have a serious use to educate or promote other types of change [23]. This approach is known as edutainment – and offers the potential to motivate learners by making learning a fun experience [24]. Whilst the games may be based on human-to-human interaction, card or other activity based, the arrival of computer video games in the 1960's onwards has enabled the richer and more varied set of interaction and automatic gameplay to enable different approach. A variant of utilising games to teach is to utilise game mechanics in other areas – such gamification [25] can offer benefits in designing learning material. In this paper, serious games will focus on computer-based games for teaching.

One particular approach, as computer graphics and sound have evolved, has been the rise of immersive games, that can use 2D or 3D graphics and stereo sound, frequently involving many players interacting in a shared rich and complex (often web-based) mixed reality world, where the player circumstances will be multi and varied. The player reality may be augmented and often self-composed, as in a user-defined avatar in a virtual world. The technology for this can range from the 2D representation on a traditional monitor -where the immersion is more limited though with modern large curved screens can still effective – through to a CAVE (Cave Automatic Virtual Environment) where the player is surrounded on multiple sides giving a more complete illusion.

In this context, education and training can overlap – the value of many immersive environments is in the ability to simulate a real world scenario so that the player can learn from the experience, whether the focus is on learning knowledge, developing physical skills, or a mix of the two.

Challenges in evaluating immersive games.

Evaluating serious games can be considered in several dimensions. As a game, there is the question of how fun and playable the game is. As an educational platform, the attention is on how effectively the learner engages and learns. Finally, as a piece of

software, the focus is on how well the software achieves the functional requirements – for both gameplay and learning. Assessing the learning functionality is beyond the scope of the current work; instead, we focus on the engagement as we can measure it through the usability of the software. There are numerous HCI approaches, but as noted above, typical usability testing depends on being able to create typical user scenarios. The key benefit of serious immersive games is that the user – the player – may be in a wide variety of contexts. Moreover, in the case of multi-player games, the gameplay and experience will vary depending on a wide set of variables. Where serious games are used outside of a controlled educational setting, the variables will include not knowing the nature or profile of the user, nor their own motivation. Where players are joining and leaving the game environment because of their other commitments, it is difficult to monitor their experience of using the system. Monitoring of their activity may be useful in indicating engagement and the apparent success of the system – but too late to improve it.

3.2 M-instruction

Introduction to M-instruction.

M-instruction, also known as mobile learning is about being able to learn on the move [26]. An increasing amount of modern computer use is via smart devices, pads, and laptops – indeed, in some contexts this is the preferred and main route for access to computing. People use these devices in a wide variety of places and contexts, giving them flexibility in when and where they access content; thus it is a natural extension to want to use these devices to learn: M-instruction encompasses the concept of utilising mobile computer devices to support teaching and learning anywhere.

Mobile technologies allow for a wide variety of learning support, from static content (web pages, course notes through PDF etc.), with interactive content, and apps and internet based material.

M-instruction is of particular relevance to lifelong learners who are not situated in a traditional learning environment [28]. However, it can also allow more choice for students on traditional courses, enabling blended learning, with some traditional on-site provision, supplemented and complemented by mobile.

What challenges are there when evaluating M-instruction.

Whilst M-instruction has its own characteristics, it shares some of the challenges of serious games when considering how to evaluate it; if the aim is to evaluate the effectiveness M-instruction, then the first issue is how to measure that value. The dimensions to consider for M-instruction are around the usefulness of the material for users to access, and to what extent the material is effective in enabling the user to learn. The challenges here

We have no way of knowing their situation, circumstance, education background and motivation, or of the customisation of the final software they are using. Getting to the end user itself can also be problematic as these are learning environments that people will dip into at opportune moments. As with serious games, we will not consider

the effectiveness of the learning itself here, but rather how to attempt to ensure the system is providing suitable functionality, enabling and encouraging engagement and use.

3.3 Ubiquitous or pervasive computing solutions for learning

UbiComp and Pervasive Computing.

The previous two sections have illustrated two areas where computer science has enabled new opportunities for established learning (game based and more general teaching) to take new forms – with computer games and mobile as platforms. As noted, these share characteristics in what they offer – flexible access, the ability to choose if and when to use them, and for how long. Indeed, the two examples overlap where virtual environments are accessed via mobile devices – with technologies such as Google Glass and Oculus Rift showing that the convergence of these is accelerating.

Evaluating UbiComp.

As software and systems become more pervasive, integrated and sometimes hidden, the challenge of how to evaluate them grows. The features already noted in the two examples of this paper show certain commonality – with varied users, in a wide variety of potential use scenarios. Amongst the toolkit of evaluation techniques, user focussed approaches [29] and frameworks [30] rely on being able to identify and observe users. For the serious immersive games and the M-Learning examples considered in this paper, the problem remains that identifying users and being able to monitor and measure their use to evaluate the system. Here we propose a hybrid approach of usability and heuristic evaluation.

4 Heuristic Evaluation as an approach to meet the challenges in evaluating serious immersive games, mLearning and ubiComp

4.1 Heuristic Evaluation

If access to the end user is hard because of location and user self-personalisation, then one solution is to look at the software before it goes out. Heuristic Evaluation allows us to get User Interface (UI) and User eXperience (UX) experts to reflect on the software before it is deployed. As summarised above, we have demonstrated before its use with pedagogical software [1]. In this paper, we propose an extension to existing Heuristics Evaluation Methods that make this technique applicable to Serious Immersive Games and M-instruction. We will also propose how existing Heuristic Methods may be adopted. The result represents a new way of making this methodology applicable to a new developing area of learning technology.

4.2 A hybrid evaluation approach

The system proposed here utilises elements of traditional usability testing - selecting the categories for measurement and evaluation – but then uses a heuristic approach where expert users then evaluate their experience against these measures. This combination approach has been developed for health M-instruction applications [27], which is a more specialised form of M-instruction. Evaluation here – as in the cases described in the case studies considered in [1], is carried out through a process of questionnaire and interview of the experienced and expert users.

This type of approach – with a selective set of usability metrics, evaluated through use by experts – benefits from the utilisation of specialised users, where the general interface and environment can be assessed in the light of longstanding relevant experience.

In an ideal world, it would be desirable to triangulate this approach with experimental based empirical work. This is possible where we can produce an experimental design with clearly identified variables, sufficient balance control and numbers to carry this off in a defensible scientific manner. Running this study alongside a heuristic evaluation would give a way of adding confidence to a heuristic evaluation story. Indeed comparing the two approaches would be insightful. However it is the very nature of the topic of this paper that make this approach very problematic. The end-users of serious immersive games and M-instruction technology are going to be hard to get into the lab. If you did manage to do this, it would still be such an unnatural environment for them to engage in their normal interaction it is not clear what would be actually learnt. If we want to study, serious immersive games and M-instruction in the large, other triangulating techniques need to be investigated. Heuristic Evaluation gives a handle on looking, from a designers and experts perspective, on the software solutions we have made. The actual experience of the users out there in cyberspace is a harder thing to judge.

5 Conclusions and Future Work

5.1 Outcomes and Conclusions

What have we learned from this effort so far is that the identification of suitable metrics and usability measures is non-trivial, but can lead to a more rapid evaluation in the process of development, and so can aid the software engineer in developing the user side functionality. Of course, the heuristic evaluation is just part of the picture, and that user-based evaluation is still needed as a part of the entire process, especially when it comes to evaluating the learning benefits of the serious games and mobile systems. Some raw data might be gained from usage metrics and performance. The trouble with MLearning is that people often only use an online tutorial to find the information that they need. They do not intend to finish the tutorial – they will quit it when they have found what they were after. Therefore, a metric that looks to completion rates or final marks is going to be wide of the mark on many occasions. The social interactions and social computing aspects also means that in a game scenario what can be learnt from

scores or levels is also not going to tell the full story. They may however give us a limited part of the picture – so as means of triangulating our data there is some potential here.

5.2 Future work

To the future - Serious Games and MLearning are just two instances of Pervasive and Ubiquitous Computing. Further work is needed to evaluate how effective Heuristic Evaluation is as a tool in evaluating these applications where the end users are at a distance and we cannot monitor them closely by traditional usability terms. One area of particular interest here is the potential to gather information on usage patterns – time, place, duration – and utilise big-data to attempt to gauge the effectiveness. The importance of Big Data to this endeavour is that it has the potential to look at users in the large. Larger samples will give a better broad-grained picture of user behaviour from which to judge what typical behavioural patterns are. If we have these more general views, we can then compare and contrast them with the insights that heuristic evaluation have given us and potentially confirm our views.

Another area for further work is that of embedding formative and diagnostic assessment within the game environment (for serious games), and within the learning pathway for M-instruction, to attempt to address the question of how to determine the effectiveness of these systems in teaching. Thus, the act of engaging in the use of the software will give us valuable data. By placing implicit performance, gathering spies within an application we have target the questions we want answered. This way the insights that we need into learning experience can be made to flow from our software as a natural consequence of use.

6 References

1. Brayshaw, M. Gordon, N, Nganji, J.T., Wen, L, and Butterfield, A. M.: Investigating Heuristic Evaluation as a Methodology for Evaluating Pedagogical Software: An Analysis Employing Three Case Studies. P. Zaphiras and A. Ioannou (Eds), *Learning and Collaboration Technologies: 16th International Conference on Human Computer Interaction*, Greece, Lecture Notes in Computer Science, Springer (2014)
2. Zyngier, D.: (Re) conceptualising student engagement: Doing education not doing time. *Teaching and Teacher Education*, 24.7, 1765-1776 (2008)
3. Gordon, N.: "Enabling personalised learning through formative and summative assessment." in *Technology-supported Environments for Personalized Learning: Methods and Case Studies*. Hershey: Information Science Publishing 268-283. (2009)
4. Susi, T., Mikael J., and Per B.: *Serious games: An overview*. (2007)
5. Anderson J R, *Learning to Program in LISP*, *Cognitive Science*, 8, p 87-129 (1984)
6. O'Shea T and Self J.: *Learning and Teaching with Computers: Artificial Intelligence in Education*, Great Britain: The Harvester Press Limited (1983)
7. Papert, S.: *MIND-STORMS: Children, Computers, and Powerful Ideas*, The Harvester Press (Brighton, UK) (1980)
8. Skinner, B.F.: The technology of teaching. *Proceedings of the Royal Society, Series B*, 162, 427-43 (1965)

9. Spector, J. M, and Davidsen, P.I.: Designing technology enhanced learning environments. *Instructional and cognitive impacts of Web-based education* 241-261 (2000)
10. Dede, C.: Immersive interfaces for engagement and learning. *Science* 323.5910, 66-69. (2009)
11. Quinn, C.: mLearning: Mobile, Wireless, in your Pocket Learning. *LineZine*, Fall 2000. Found at: <http://www.linezine.com/2.1/features/cqmmwiyp.htm> (2000)
12. Greenfield, A.: *Everyware: The dawning age of ubiquitous computing*. New Riders (2010)
13. Gordon, N.: Flexible Pedagogies: technology-enhanced learning. Higher Education Academy, NIACE. http://www.heacademy.ac.uk/resources/detail/flexible-learning/flexiblepedagogies/tech_enhanced_learning/main_report (2014)
14. Nielsen, J.: Usability inspection methods. *Conference companion on Human factors in computing systems*. ACM (1994)
15. Nielsen J.: How to conduct a Heuristic Evaluation (Online), Available: http://www.useit.com/papers/heuristic/heuristic_evaluation.html (2015)
16. Nielsen J. and Mack R L.: *Heuristic Evaluation. Usability Inspection Methods*. New York, John Wiley & Sons: 25-62 (1994)
17. Nielsen J and Molich R.: Heuristic evaluation of user interfaces, *Proc. ACM CHI'90 Conf.* (Seattle, WA, 1-5 April), 249-256 (1990)
18. Benson L, Dean E., Grant, M., Holschuh, D., Kim, B., Kim, H., Lauber, E., Loh, S., and Reeves, T.: *Heuristic Evaluation Instrument and Protocol for E-Learning Programs* (Online), Available: <http://it.coe.uga.edu/~treeves/edit8350/HEIPEP.html> (2001)
19. Squires, D. & Preece, J.: Predicting quality in educational software: Evaluating for learning, usability and the synergy between them. *Interacting with Computers*, 11, 467-483. (1999)
20. Wen, L., Brayshaw, M., and Gordon, N.: Personalized Content Provision for Virtual Learning Environments via the Semantic Web, *Innovations in Teaching and Learning in Information and Computer Science*. (2012)
21. Nganji, J.T. and Brayshaw, M.: Designing and Reflecting on Disability-Aware E-learning Systems: The Case of ONTODAPS, *IEEE 14th International Conference on Advanced Learning Technologies (ICALT)*, Athens, Greece, DOI: 10.1109/ICALT.2014.167 http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6901543&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6901543 (2014)
22. Butterfield, A.M., and Brayshaw, M.: A Pedagogically Motivated Guided Inquiry Based Tutor for C#, *Proceedings of the HEA STEM (Computing) Learning Technologies 2014 Workshop*, At University of Hull, Volume: Editor: Neil A Gordon and Deryn Graham https://www.researchgate.net/publication/263213733_Proceedings_of_the_HEA_STEM_%28Computing%29_Learning_Technologies_2014_Workshop (2014)
23. Michael, D. R. and Chen, S.L.: *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade (2005)
24. Egenfeldt-Nielsen, S.: *Beyond edutainment: Exploring the educational potential of computer games*. Lulu.com (2005)
25. Gordon, N., Brayshaw, M. and Grey, S.: Maximising Gain for Minimal Pain: Utilising Natural Game Mechanics. *Innovation in Teaching and Learning in Information and Computer Sciences* 12.1 27-38 (2013)
26. Taylor, J.: Towards a task model for mobile learning: a dialectical approach. *International Journal of Learning Technology* 2.2 (2006): 138-158.
27. Aljaber, T., Gordon, N., Kambhampati, C. and Brayshaw, M.: An Evaluation Framework for Mobile Health Education Software. In *proceedings of the 2015 Science and Information Conference* (2015).

28. Holzinger, A., Nischelwitzer, A., & Meisenberger, M.: Lifelong-learning support by M-instruction: example scenarios. *eLearn*, (11), 2 (2005).
29. Korn, M. and Bødker, S.: Looking ahead: how field trials can work in iterative and exploratory design of ubicomp systems. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM (2012).
30. Silva, J.L., Campos, J. and Harrison, M.: Formal analysis of ubiquitous computing environments through the APEX framework. *Proceedings of the 4th ACM SIGCHI symposium on engineering interactive computing systems*. ACM (2012).