

Incorporating Multi-Modal Evaluation into a Technology Enhanced Learning Experience

Lynne Hall¹, Ruth Aylett², Colette Hume¹, Eva Krumhuber³, & Nick Degens⁴

¹ University of Sunderland, Sunderland, SR6 0DD, UK

² Heriot-Watt University, Edinburgh, EH14 4AS, UK

³ Jacobs University Bremen, Bremen, 28759, Germany

⁴ Wageningen University, Leeuwenborch, Netherlands

lynne.hall@sunderland.ac.uk, ruth@macs.hw.ac.uk, colette.hume@sunderland.ac.uk,
e.krumhuber@jacobs-university.de, nick.degens@wur.nl

Abstract. This paper outlines Transmedia Evaluation and its application to the evaluation of TRAVELLER, a Technology Enhanced Learning application that aims to provide 18-25 year olds with improved cultural understanding and sensitivity. TRAVELLER aims to enhance the users' intercultural communication through engaging users in multi-modal interaction with intelligent characters deployed in emergent narratives. Transmedia Evaluation will enable us to gather and analyse a wide range of complimentary user data in response to diverse R&D questions, whilst embedding the evaluation within the user experience, so that the user is unaware of evaluation instruments and measures. The described approach will enable us to investigate subjective views and physiological responses to the narrative and characters, reflective and immediate reactions to innovative interaction. To achieve this we will use data collection instruments ranging from questionnaires to observation to physiological measurement, all embedded within the context of a Live (Inter)Action Role Play.

Keywords: Evaluation, Transmedia, Technology Enhanced Learning, Role Play

Introduction

For applications involving multi-modal interaction, analyzing the user's data and then interpreting that in a manner that is appropriate to the user's context and expectations remains a critical issue. But how do we evaluate if this has been achieved? Evaluation is crucial to determining if the multi-modal design decisions that we are making actually achieve the system's intended impact, and further, if, as is always implicitly assumed, that it will improve, enhance or add value to the interaction.

Evaluation is a challenging and difficult issue for designers and developers of innovative systems and interaction approaches. From the issue of what are we evaluating to how that evaluation impacts on the user's experience, the evaluation landscape is relatively uncharted and sufficiently difficult to navigate that even where evaluation does occur, it often appears somewhat ad hoc. For example, (Van Velsen et al, 2008) in a review of 55 approaches to evaluating user experience in Augmented Reality,

noted that the “*reports encountered lacked empirical value,*” with a dominance of questionable quality questionnaires, shallow reporting from interviews and think aloud protocols, and limited triangulation with logged data. There is a recognized need for innovation, evolution and improvement in the evaluation of contemporary systems (Wiberg, Jegers & Desurvire, 2009) and in the work we report here, we focus on multi-modal interaction between users and intelligent graphical characters within the context of an embracing narrative framework. Here, social immersion and affective engagement become key issues and further complicate evaluation approaches.

The 7th Framework EU eCUTE (education in Cultural Understanding, Technology Enhanced) project aims to develop cultural understanding by providing immersive virtual role-play with intelligent interactive graphical characters embodying models of culturally specific behaviour and interaction. TRAVELLER (TRAINing for Virtually Every Location for Learning Empathic Relationships) is an application being developed by eCUTE. This is targeted at young adults aged 18-25 and on culturally specific communication and interaction behaviour. TRAVELLER employs multiple modalities for input and output where the user experience involves full-body interaction, with speech, gesture and mixed reality interfaces providing the tools of interaction.

A key issue for eCUTE is to increase quality and innovation in evaluation approaches for measuring affective responses, empathy and engagement. To achieve this we will create a multi-level multi-method dependent variable matrix involving physiological and expressive responses as well as subjective experience. This matrix will be informed through multiple user and stakeholder evaluation techniques, instruments and approaches. In this paper, we outline our intended approach to gather the information to construct our matrix within the context of the TRAVELLER application.

The Approach: Transmedia Evaluation

The evaluation approach developed in eCUTE, Transmedia Evaluation is detailed in figure 1.

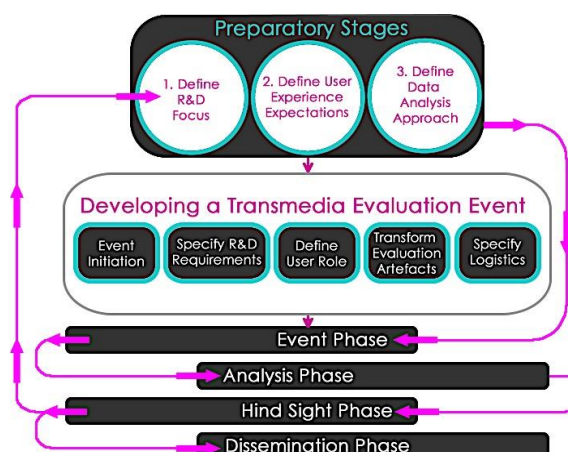


Fig. 1. Transmedia Evaluation

Transmedia Evaluation has two main aims:

- To provide the data to answer challenging, multiple, interdisciplinary research questions from the R&D team involved in creating the applications.
- To reduce the disruption of evaluation on the user experience and ensure that evaluation does not diminish the user experience.

Transmedia Evaluation is an effective approach, seamlessly embedding R&D evaluation into the user experience. We have conducted extensive evaluations where users were unable to identify which elements of their experience had involved evaluation (Hall & Hume, 2011). Further, in all of the studies conducted using Transmedia Evaluation, the required data and more has been provided and disseminated to the R&D team (Hall et al, 2012).

Transmedia Evaluation events are more than simply embedding questionnaires and activities. Effectively the user experiences a Live Action Role Play, where part of that action involves interaction. We place the user into role as soon as possible and retain them in that role until the end of the experience, see Figure 2. This role is determined prior to any evaluation development through assessing the potential narrative that can scaffold the evaluation and interaction, identifying user characteristics and expectations and establishing the critical R&D issues that the evaluation must cover.

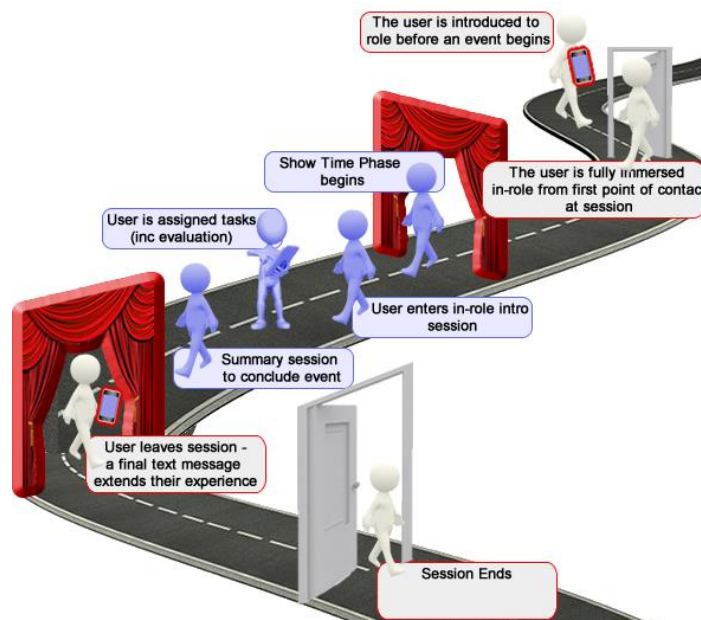


Fig. 2. User Experience of a Transmedia Evaluation Event

TRAVELLER

The specific learning aim of engaging with TRAVELLER is to improve intercultural communication through increasing cross-cultural sensitivity. Cross-cultural sensitivity generally refers to the capability of people to deal with negative emotions that can arise from interacting with people from different cultures (i.e. frustrations and anger due to misunderstandings). Using TRAVELLER, learners will interact with culturally configured agents in emergent critical incidents related to intercultural communication. By progressing through these critical incidents, learners will progress through the learning goals and increase their cultural sensitivity, aiming to remove stereotypes and increase empathy for people from different groups. The eCUTE learning goals, as outlined in table 1 (see also Swiderska, Krumhuber, Kappas, Degens, & Hofstede, 2011). A focus on important elements of intercultural training (Cushner & Brislin, 1997; Deardorff, 2006; Stephan & Stephan, 2006).

Table 1. eCUTE Learning Goals

Stage of learner	Emotional goals	Cognitive goals	Behavioural goals
Beginner (conscious incompetence) Observation and acquisition	Be able to recognize your emotions (for example fear and anxiety) when dealing with strange behaviours of another group	Start learning the specific practices and values of another group	Be fully present in attending to the other's verbal and non-verbal messages
Journeyman (conscious competence) Relating and Experimenting	Be able to observe the behaviour of another group without feeling prejudice	Understand on a basic level the differences and similarities between another group and your own.	Practice skills learned in the previous stage and experiment with different forms of behaviour
Expert (unconscious competence) Adapting and belonging	Be able to share the emotions (such as sadness or happiness) of a member of another group and other's experiences through empathy	Players should be able to discriminate and select appropriate strategies in cultural contexts	Be able to unconsciously participate in a group as a native

TRAVELLER is being developed using the FATiMA architecture (Dias & Paiva, 2005). This provides an implementation of the OCC model (Ortony et al, 1988) and PSI (Doerner, 2003) enabling agents to appraise emotional and socially challenging situations. With TRAVELLER this architecture has been extended, with agent appraisal refined and impacted by culture as represented by Hofstede's cultural dimensions (Hofstede, 2001). Detailed information on this architecture and its extensions can be found in (Lim et al, 2012).

TRAVELLER provides users with an adventure style game, benefiting from the global appeal of films such as Indiana Jones with treasure seeking underpinning the plot the user is pursuing. Within the application, users travel to different countries to find the missing pages of their dead grandfather's journal. Within these episodes, users engage with agents from different cultures experiencing critical incidents related to cultural communication (e.g. greetings). The users interact with TRAVELLER through Microsoft's Kinect, with the modalities of speech and gesture enabling them to communicate and engage with the characters in the various TRAVELLER episodes. Figure 3 depicts the scene where the character is in the 'hub', in which reflection will occur following critical incidents.



Fig. 3. A Scene from TRAVELLER

Evaluating TRAVELLER

With TRAVELLER, a number of our R&D questions relate to how the user responds to the application's narrative, mechanics, characters and learning. We want to know this from a range of perspectives using vastly different measurement approaches. In earlier Transmedia Evaluations we have focused on embedding relatively traditional instruments and approaches, such as questionnaires, focus groups, interviews and educational assessment tools. However, with the evaluation of TRAVELLER, we will extend the evaluation to incorporate expressivity and physiological measures.

With traditional measures, such as questionnaires, interviews and focus groups, it is extremely easy to re-brand and re-contextualise instruments to fit with almost any scenario. Even very light Transmedia Evaluations using only simple aesthetic improvements have an impact on user engagement and are clearly not perceived by users to be evaluation (Hall & Hume, 2011). However, if part of the evaluation experience involves the application of electrodes and your mind appearing on a screen somewhere, as you use some piece of software, it is much harder to avoid it becoming an intrusive experience.

The aim of Transmedia Evaluation is to remove any sense that the user is participating in an evaluation. However, expressivity and physiological measures are often necessarily intrusive, so how then can they be incorporated in such a way as to be part of the user experience?

4.1 Transmedia Evaluation of TRAVELLER

The TRAVELLER age group, 18-25, is particularly open to innovative approaches to learning and readily participates in wide-ranging narratives across multiple media, modalities and platforms. So, we are provided with an utterly flexible user group who will probably buy-in to anything we do as long as it engages them and makes sense within their user role and context. In creating the Transmedia Evaluation Event for TRAVELLER, the following design choices have been made:

1. User role: learner (e.g. not gamer / companion / friend / interactor / designer, etc.)
2. Plot: Users experiencing a half-day workshop, *Diplomat Training*, targeted at those wanting to work in an international setting.
3. Transmedia: Props and media to be created for the event will seek to provide the user with the sense that they are participating in an interesting way to learn something that might help them if they decided to follow international careers. Using the context of a learning workshop provides massive scope, as almost any learning approach and activity will be considered acceptable by the users as long as it is engaging and appears to relate to the overall focus, that of diplomatic training.
4. Set: Sunderland University and Jacobs University Bremen, including their X-labs equipped with experimental software and monitoring equipment.
5. User experience: innovative learning experience branded as Diplomat Training with a range of experiences including interacting with TRAVELLER and participating in an evaluation of TRAVELLER
6. Users: students at Sunderland University and Jacobs University with a wide range of student nationalities
7. Cast: Trainers of Diplomats
8. Evaluation Battery: including a range of cultural, social and interaction related questionnaires and activities, physiological and expressivity measures, user generation and consumption, discussion and in-role tasks and experiences.
9. Key Question: TRAVELLER is being developed as a serious game. Thus, the users appear to be engaging in a game, whilst actually they are learning about intercultural communication, but are they actually learning or are they just playing?

Figure 4 provides the evaluation and interaction experience that the user will engage in with TRAVELLER. With Traveller, the user engages in a series of nodes, each of these is placed within the workshop context, with a clear in-role purpose given for user activities. This involves a range of activities that provide context setting for the role play as well as for gathering data for evaluation purposes.

The TRAVELLER experience invisibly collects and logs user data, for example, the user is interacting with the Kinect, they are unaware of the logging of the depth and colour video stream, skeleton tracking data, gesture recognition, etc. Where intrusive data collection occurs this is embedded, contextualized and justified linking the use of video, physiological measures, instruments and devices to the learner experience.

Constant videoing is justified through the use of Teaching Process Recall (Claydon & McDowell, 1993), with learners learning this approach, then applying it to view and reflect on their activities in the debrief. Physiological measures are contextualized through learners being told of their growing importance of physiological testing for

issues such as national security and the need to be aware of the potential of such technology. Questionnaires (e.g. the Cultural Intelligence Scale (Ang et al, 2007), General Evaluation Scale (Wright et al, 1997)) are incorporated as quiz style activities in workbooks and/or on iPads. Qualitative data is gathered through the “Football Manager’s Booth,” with the learners responding in a very short structured interview with a pre-given set of questions, a reporter and camera, to get their overall perspective of “the play” and the key moments / critical elements of their experience.

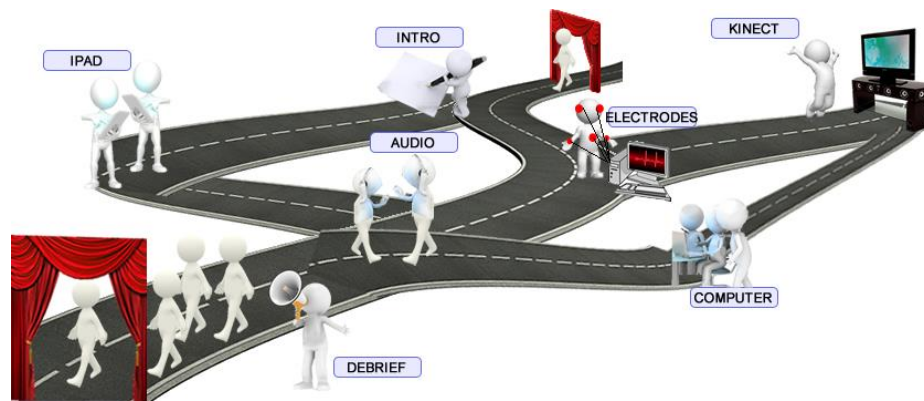


Fig. 4. User's Experience with TRAVELLER

Discussion

The main purpose of evaluating TRAVELLER is to contribute to technology enhanced learning understanding, thus a key part of our evaluation is related to whether TRAVELLER impacts on learning about intercultural communication. However, in addition we aim to increase our understanding of the value of incorporating multi-modal interaction and of the use of multi-modal data for evaluation.

Through creating an evaluation approach where different plot nodes of the same application can allow testing across different modalities whilst the user is unaware of the evaluation intent, provides data from users concentrating and engaging with the serious game, rather than focusing on the technologies they are surrounded by. Thus, the technology should move into the background and allow us to evaluate the whole system in a relatively natural, if highly contrived setting.

Both TRAVELLER and its Transmedia Evaluation are under development. We are creating the “Evaluation Pack” that will surround the evaluation, providing plot, scripts, protocol, instruments, approaches, coding frames and so on. As in the diagram in figure 2, the TRAVELLER Transmedia Evaluation Event, will require a “prequel”, with event marketing and sign-up, along with pre-event transmedia preparing the user for their role and gathering initial data. We intend to evaluate TRAVELLER in May 2013, aiming to collect and collate a diverse range of data covering expressivity, physiological measures, speech, gesture, learning and subjective measures.

Acknowledgments. This work was partially supported by European Community (EC) and is currently funded by the ECUTE project (ICT-5-4.2 257666). The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein. Further information on eCUTE can be found at www.ecute.eu

References

- Ang, S., Van Dyne, L., Koh, C., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural Intelligence: Its Measurement and Effects on Cultural Judgment and Decision Making, Cultural Adaptation and Task Performance. *Management and Organization Review*, 3(3), 335-371.
- Claydon, T. & McDowell, L. (1993) 'Watching yourself teach and learning from it' in Sally Brown et al, *Observing teaching*, SEDA Paper 79.
- Cushner, K., & Brislin, R. W. (1997). *Improving intercultural interactions: modules for cross-cultural training programs* (Vol. 2). Thousand Oaks, CA: Sage Publications.
- Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internalization. *Journal of Studies in Intercultural Education*, 10, 241-266.
- Dias, J., Paiva, A. (2005) Feeling and reasoning: A computational model for emotional agents. In: 12th Portuguese Conference on Artificial Intelligence (EPIA 2005), Portugal, Springer (2005) 127-14
- Doerner, D.: The mathematics of emotions. In Frank Detje, D.D., Schaub, H., eds.: *Proceedings of the Fifth International Conference on Cognitive Modeling*, Bamberg, Germany (2003) 75-79
- Hall, L., & Hume, C. (2011). Why Numbers , Invites and Visits are not Enough : Evaluating the User Experience in Social Eco-Systems. *SOTICS*, (pp. 8-13) Barcelona, Spain.
- Hall, M., Hall, L., Hodgson, J., Hume, C., & Humphries, L. (2012). Scaffolding the Story Creation Process, *Proceedings of the 4th Conference on Computer Supported Education*, Oporto, Portugal, April 2012.
- Hofstede, G. (2001) *Culture Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*, SAGE Publications, 2001.
- Lim, M,Y; Dias, J; Aylett, R.S. & Paiva, A. (2012) Creating Adaptive Affective Autonomous NPCs. *Journal of Autonomous Agents and Multi-agent Systems*, v24 n2 pp287-311
- Ortony, A., Clore, G., Collins, A.: *The cognitive structure of emotions*. Cambridge University Press, Cambridge, UK (1988)
- Stephan, C. W., & Stephan, W. G. (2002). Cognition and affect in cross-cultural relations. In W. B. Gudykunst & B. Mody (Eds.), *Handbook of international and intercultural communication* (2nd ed) (pp. 127-142). Thousand Oaks: Sage Publications.
- Swiderska, A., Krumhuber, E., Kappas, A., Degens, N., & Hofstede, G. J. (2011). Preliminary cultural learning interdisciplinary framework. Technical report of eCUTE, Education in Cultural Understanding, Technologically-Enhanced, FP7-ICT-2009-5, Project No: 257666. February, 2011.
- Van Velsen, L., Van Der Geest, T., Klaasen, R., & Steehouder, M. (2008). User-centered evaluation of adaptive and adaptable systems: a literature review. *The Knowledge Engineering Review*, 23(03), 261-281.
- Wiberg, C., Jegers, K., & Desurvire, H. (2009). How Applicable is Your Evaluation Methods Really? Analysis and Re-design of Evaluation Methods for Fun and Entertainment. 2009 Second International Conferences on Advances in Computer-Human Interactions, 324-328.

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology*, 73, 73-90.