Engaging Children
in Interactive Application Evaluation

Lynne HALL[1], Colette HUME[1] and Sarah TAZZYMAN[1,2]

ABSTRACT

Interactive applications designed specifically for children offer great potential for education and play. However, to ascertain that the aims of applications are achieved, child-centred evaluations must be conducted. The design of any evaluation with children requires significant consideration of potential problems with comprehension, cognitive ability, response biases and study attrition. Multidisciplinary R&D project evaluation requirements are often extensive, requiring an all-encompassing and prolonged evaluation design. Discontinuity between the highly engaging interaction experience and the multitude of measures that form the evaluation poses a major issue for the evaluation of interactive applications. In response, we have developed Transmedia Evaluation, a method that aims to maintain engagement throughout the evaluation process. In this paper, the Transmedia Evaluation process is explained and applied to evaluate a learning application for children, MIXER (Moderating Interactions for Cross Cultural Empathic Relationships). Children aged 9-11 (N = 117) used the MIXER application and completed an evaluation battery including pre- and post-test questionnaires, immediate learning assessment and qualitative evaluation. Using Transmedia Evaluation to develop the MIXER evaluation resulted in complete data-sets (100%) for quantitative data (by self-regulated completion) along with rich, high quality qualitative responses. Transmedia Evaluation transformed the evaluation, with children fully engaging in and enjoying their experience.

KEY-WORDS: EVALUATION, CHILD-CENTRED DESIGN, USER EXPERIENCE, LEARNING TECHNOLOGY

Correspondence concerning this article should be addressed to Lynne Hall, Department of Computing, Engineering and Technology, David Goldman Informatics Centre, University of Sunderland, St Michael's Way, Sunderland, SR6 0DD. *Email:* lynne.hall@sunderland.ac.uk.

1. Lynne Hall and Colette Hume, Department of Computing, Engineering and Technology, David Goldman Informatics Centre, University of Sunderland, St Michael's Way, Sunderland, SR6 0DD., UK *Email:* lynne.hall@sunderland.ac.uk, me@colettehume.com

2. Sarah Tazzyman, Henry Wellcome Building, The School of Psychology, University of Leicester, Leicester, LE1 9HN, UK. *Email:* snt5@le.ac.uk.

RÉSUMÉ

**L'engagement des enfants
dans l'évaluation d'applications interactives**

Les applications interactives conçues spécifiquement pour les enfants offrent un grand potentiel pour l'éducation et le jeu. Cependant, pour s'assurer que les objectifs des applications sont atteints, des évaluations centrées sur les enfants doivent être menées. La conception de toute évaluation avec des enfants nécessite un examen minutieux des problèmes potentiels liés à la compréhension, à la capacité cognitive, au biais des réponses et à l'étude de l'attrition. Les exigences des évaluations des projets R & D multidisciplinaires sont souvent considérables, nécessitant une conception de leur évaluation globale et prolongée. La discontinuité entre l'expérience de l'interaction très attrayante et la multitude de mesures dont est composée l'évaluation pose un problème majeur pour l'évaluation des applications interactives. Pour répondre à ce problème, nous avons développé l'évaluation Transmedia, une méthode qui vise à maintenir l'engagement tout au long du processus d'évaluation. Dans cet article, le processus d'évaluation Transmedia est expliqué et appliqué pour évaluer une application d'apprentissage pour les enfants, MIXER (Modération des Interactions pour des Relations empathiques cross-culturelles). Des enfants âgés de 9 à 11 (N = 117) ont utilisé l'application MIXER et ont complété un grand nombre d'évaluation, y compris des questionnaires de pré- et post-test, une évaluation immédiate de l'apprentissage et une évaluation qualitative. Utiliser l'évaluation Transmedia pour développer l'évaluation de l'application MIXER a abouti à des ensembles de données complets (100 %) pour les données quantitatives (par une méthode auto-réglementée), ainsi que des réponses qualitatives riches. L'évaluation Transmedia a transformé l'évaluation des enfants en leur permettant de s'engager pleinement et de profiter de leur expérience.

MOTS-CLÉS : APPLICATIONS INTERACTIVES, ÉVALUATION PAR LES ENFANTS, ÉDUCATION, JEU

INTRODUCTION

In evaluating children's experience of interactive applications we, as researchers and evaluators, are aiming to provide further evidence for or against specific issues, expectations and concerns related to the impact of the interaction on the child. Whilst innovations and experiments across the reality spectrum have produced a myriad of engaging applications for children, this trend has not been followed in their evaluation. Although there has been a significant increase in studies about children's use of interactive technologies, this has not resulted in a significant diversity of methods used to gather evaluation data. With rare exceptions, the evaluation of even the most radical system has relied on surveillance techniques (e.g. video observation, logging, usage data, etc.) and/or explicit evaluation activities (e.g. paper/pencil questionnaires, interviews, panels, etc.). In Ólafsson, Livingstone, & Haddon's (2013), review of studies of children's use of the internet, over two thirds of studies only collected quantitative data and few studies used mixed methods.

Interactive applications developed for children often intend to immerse and engage them within a self-created and maintained experience. Yet, when the focus turns from interaction to evaluation, this immersion is frequently fractured. The focus, design, specific tasks and overall image of evaluations are often significantly different, and at odds with the interactive experience. Whether for games, recreation, learning or social environments, evaluation is often disruptive, provided as a separate, dislocated activity, see figure 1, with little consideration of the user's experience. For children, this can result in being taken from being engaged and having fun in roles such as "virtual pet owner" "secret friend" or "space cadet" to instead being placed into the role of "subject" in an evaluation procedure. A standard-format questionnaire can be viewed as a disengaging follow-up activity, especially if it follows a novel and immersive technological experience.
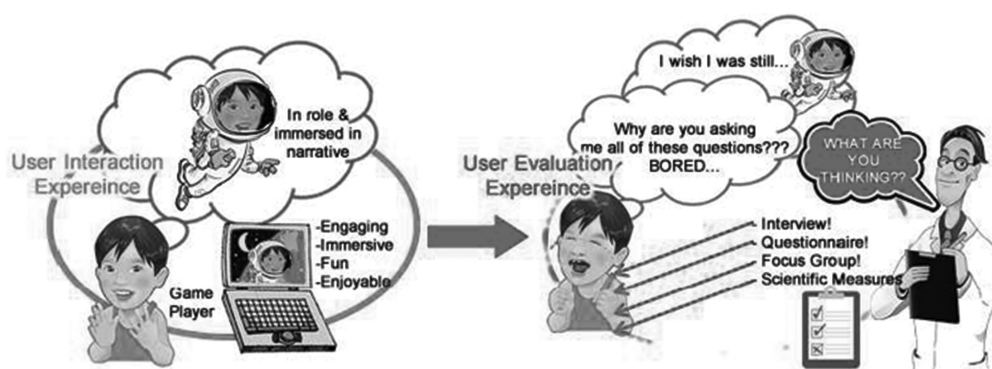
Figure 1.
Standard "disruptive" evaluation approach

In R&D evaluations of children's use of technology the primary instrument is questionnaires. Administration is typically straightforward and data analysis from structured questionnaires provides a well understood and accepted evaluation

methodology throughout the research, public sector and business communities. However, child-centred factors that can impact on question answering, such as developmental effects including language ability, reading age, and motor skills, as well as temperamental effects such as confidence, self-belief and the desire to please (Read & MacFarlane, 2006) are rarely dealt with in the evaluation design. Many evaluations involve children filling in instruments that use adult language and formats, continuing the trend noted in Jensen and Skov (2005). Although some evaluations do attempt to create appropriate methods, in general, most evaluations for children are very similar to adult evaluations, where interaction is surrounded by arduous, possibly unappealing and frequently inappropriate evaluation instruments. This can all result in study attrition and incomplete data sets, which can greatly impact on the overall results and conclusions drawn from the evaluation.

Using traditional evaluation approaches with children can have serious implications, both for the child's experience and the quality of data collected. A lack of engagement typically results in providing sub-optimal responses in questionnaires, with a high chance of satisficing (Krosnick, Narayan, & Smith, 1996) and acquiescence bias (Babbitt & Nystrom, 1989). Usability and user experience satisfaction studies tend to demonstrate extremely positive results, with child respondents agreeing or strongly agreeing to scaled questions. Throughout the literature these results are interpreted as showing that the interactive system is engaging, easy to use and entertaining. Few really ask the question of whether the data was high quality or sub-optimal. This can have important implications for conclusions drawn and future development. For instance, Buckleitner, (1999) noted "As we move into the 21st century, our children deserve rigorous, well constructed evaluation methods applied to the products they use that are subject to public criticism and evaluation." However, even though researchers are evaluating with children more than ever before, and have increased public availability of results through a significant increase in dissemination and publications there are continuing doubts about the validity of many evaluation results (Zaman, Vanden Abeele, Markopoulos, & Marshall, 2012).

Child representation and respect are further issues raised in the evaluation of interactive applications for children, highlighted by Read et al., (2008) who note that "A core value for the field of Child–Computer Interaction is that the interests of children are represented and respected in the research and design processes." However, in many evaluations there appears to be very little representation of, or respect for children's interests. Nor do studies typically report on children's response to evaluation, although tworks and Sapouna et al., (2010) note that the additional activity required by evaluation can diminish the child's enjoyment of the experience. With the focus of evaluation on the capture of valid and reliable data to substantiate hypotheses, the centre of an evaluation design is not the child, but rather the R&D motivation. Appropriately designed evaluations need to place children at the centre of the evaluation experience, just as we recognize that we should place them at the centre of the interaction design.

This paper discusses Transmedia Evaluation, a methodology for creating evaluation experiences that places users at the centre of the design. The approach aims to seamlessly embed evaluation into the user experience, providing valid and reliable data and adding value to the user. Transmedia Evaluation was developed and trialed with 9-11 year old children as the primary users and critical participants in the evaluation. We focus on the Transmedia Evaluation of MIXER, a technology enhanced learning application targeting intercultural conflict developed for 9-11 year olds (eCute, 2012), that provides users with immersive virtual role-play with intelligent interactive graphical characters.

THE EVALUAND: MIXER

MIXER (see figures 2, 3 and 4), is a Virtual Learning Environment populated by intelligent, affective and interactive characters targeted at 9-11 year old children, highlighting strategies and supporting the development of intercultural skills and competences. The Summative Evaluation of MIXER aimed to provide demonstrable evidence that experiential intercultural learning could be provided to children through the innovative technology, further detailed in (Aylett et al., 2014; Endrass, Hall, Hume, Tazzyman, & Andre, 2014), developed in the eCute project (www.ecute.eu).

MIXER engages users in an interactive narrative set in a virtual summer camp, where two groups of school children (intelligent characters) play Werewolves, a popular intergenerational game widely known in many cultures. As is common in summer camps, the children were dressed in team T-shirts, see figure 2, representing the two teams: the Reds and the Yellows.
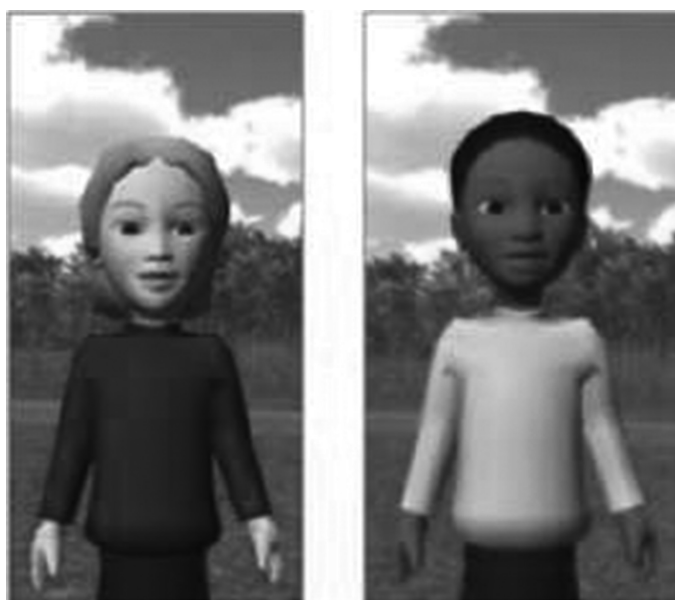
Figure 2.
Alex and Lisa, characters from MIXER
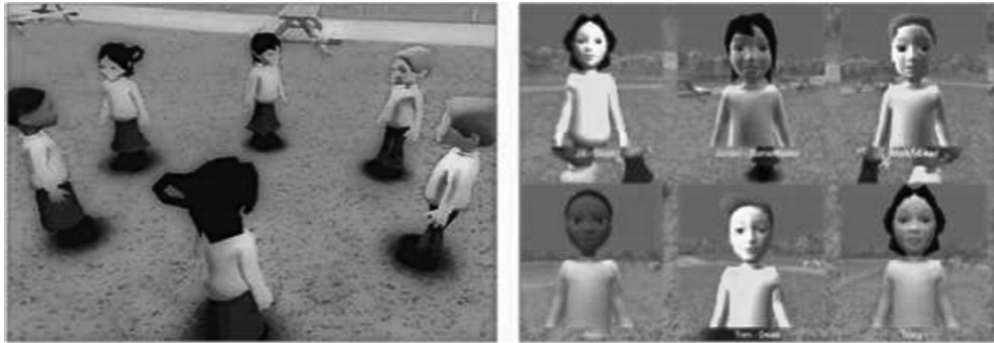
Figure 3.
Scenes from MIXER

MIXER depicts a peer conflict scenario, occurring when Tom (protagonist) plays the Werewolves game with two different teams of children, the Yellows and the Reds at a summer camp (see figure 3). Each of the teams plays by a different rule set. Within the Yellow team, the rules are that each player takes turns to say who they think the werewolf is and why, the player with the most votes is then killed off and is out of the game. In the red team, one player states who they believe the werewolf to be, if they do not have majority agreement from the other players then they themselves are killed and they are out of the game. As the scenario unfolds, it becomes evident to Tom and the child user that the groups adhere to different rules (reflecting that they belong to two different cultures or moral circles). Tom accuses the red team of cheating, because he does not understand the rule change, and Tom tells the child user that he thinks they just don't want him to play. MIXER ends by Tom resolving the conflict with the Red team by discussing the differences in the two versions of Werewolves.

In MIXER, the child does not directly appear in the virtual world. Instead their role is to interact with Tom, as an invisible friend and to support his play by responding to Tom's requests for advice on how to react and what to do at different stages of the game. The child interacts with Tom through a tablet using a Pictorial Interaction Language (Endrass, Hall, Hume, Tazzyman, Andre, et al., 2014), (see figure 4), providing children with access to over 70 graphics structured for use in sentences, enabling them to interact with Tom.

TRANSMEDIA EVALUATION: BACKGROUND & BASIS

Talking to children about evaluation quickly identifies that their expectations are constructive and optimistic. Children expect to have an interesting, entertaining, and engaging experience. Placing this expectation of enjoyment and engagement on evaluation, quickly changes the nature of the activity, away from the traditional approach of "doing something to someone to gather data for R&D purposes" instead "to designing an engaging experience for the user enabling them to provide quality data."
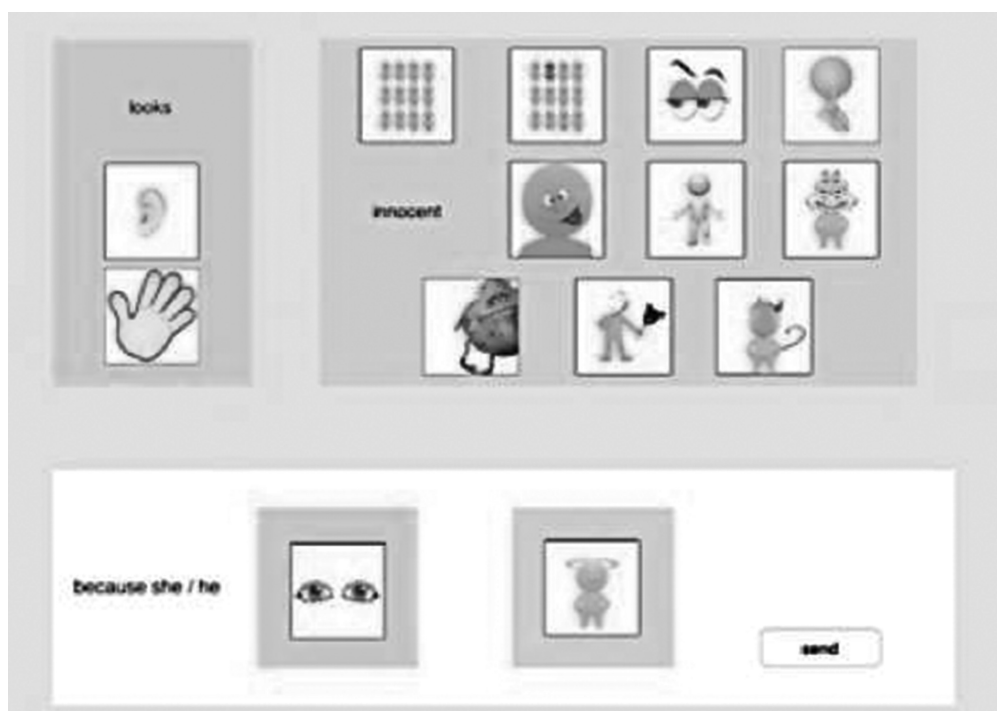
Figure 4.
Fragment of Pictorial Interaction Language

Research has rarely considered creating engaging evaluation experiences of interactive applications, whilst there has been considerable focus on enhancing engagement. Engagement is viewed as a quality of user experience that facilitates more enriching interactions with interactive applications (O'Brien & MacLean, 2009; O'Brien & Toms, 2010). Further, it can be defined by a core set of attributes: aesthetic appeal, novelty, involvement, focused attention, perceived usability and endurability. Designing and implementing these attributes into evaluation experiences would clearly create more engaging and enriched experiences. Whilst it is relatively straightforward to create usable (e.g. sensible number of age appropriate questions) and appealing (e.g. age appropriate graphics) materials, incorporating attributes such as involvement and focused attention is more challenging.

Engaging users requires a dramatic rethink of how we present the experience to the user. Our approach has been inspired by transmedia: "...a process where integral elements of a fiction get dispersed systematically across multiple delivery channels for the purpose of creating a unified and coordinated entertainment experience" (Jenkins, 2011). The most successful transmedia extends the primary user experience (e.g. viewing a movie or programme or in our case, engaging with an interactive application), taking the narrative from a TV show or movie to create a nucleus that is surrounded by supplemental story lines and activities. Transmedia is "a user-focused experience that is collaborative, immersive, and

interactive" (Parker & McDonald, 2014). In contrast to evaluation, the additional experience and activity offered by transmedia adds considerable value to the user experience and users wish to engage with it.

Creating transmedia content requires that the themes, tone and message are compatible and consistent with the narrative of the film (Gomez & Pulman, 2012), authentically extending the story world in which the experience unfolds (Weiler, 2012). Thus, transmedia has provided the underpinning philosophy for our approach, with Transmedia Evaluation aiming to provide users with a unified, themed and coordinated experience. This experience should provide consistent, integrated content through appropriate channels, platforms, devices and activities designed to meet user expectations and to reinforce engagement with the experience.

As figure 5 depicts, Transmedia Evaluation aims to seamlessly embed evaluation into the user experience by creating evaluation materials and activities that are both appropriate and engaging for the target user group, connect to the evaluand and result in high quality data for the research team. In large R&D projects, these goals cannot be met by simply embedding the evaluation within the evaluand. Rather Transmedia Evaluation aims to integrate interaction with an innovative, interactive system and the related evaluation battery into a consistent, coherent user experience.

Early piloting of Transmedia Evaluation focused on ensuring that R&D requirements were met even if instruments had been transformed to provide children with a single, coherent, transmedia inspired experience. Using a low fidelity evaluand of MIXER (a comic strip), three variants of the same instruments were provided (see figure 6):

- **Basic**, traditional evaluation approach (A4, black and white, numbered quantitative and qualitative questions – age appropriate language and format);
- **Better**, more hybrid approach, providing cosmetically improved instruments, for example, appealing colour graphics, interactive activities and some variety in question and response formats, but without a clear connection to the evaluand;
- **Best**, as an integrated comic book incorporating the MIXER comic strip with evaluation materials, based on the cosmetically improved instruments but designed to reinforce a connection with the evaluand (e.g. using figures from the comic strip in evaluation activities).

The results were startling. Not only was appropriate data provided even in the most transformed of the instruments, but further this data was more complete, of better quality, showed richer qualitative responses and improved user engagement (Hall & Hume, 2011). The more the evaluation materials met the children's expectations (e.g. the better they looked), the more interactive they were (e.g. in the sense of requiring children to actively interact with the comic book by
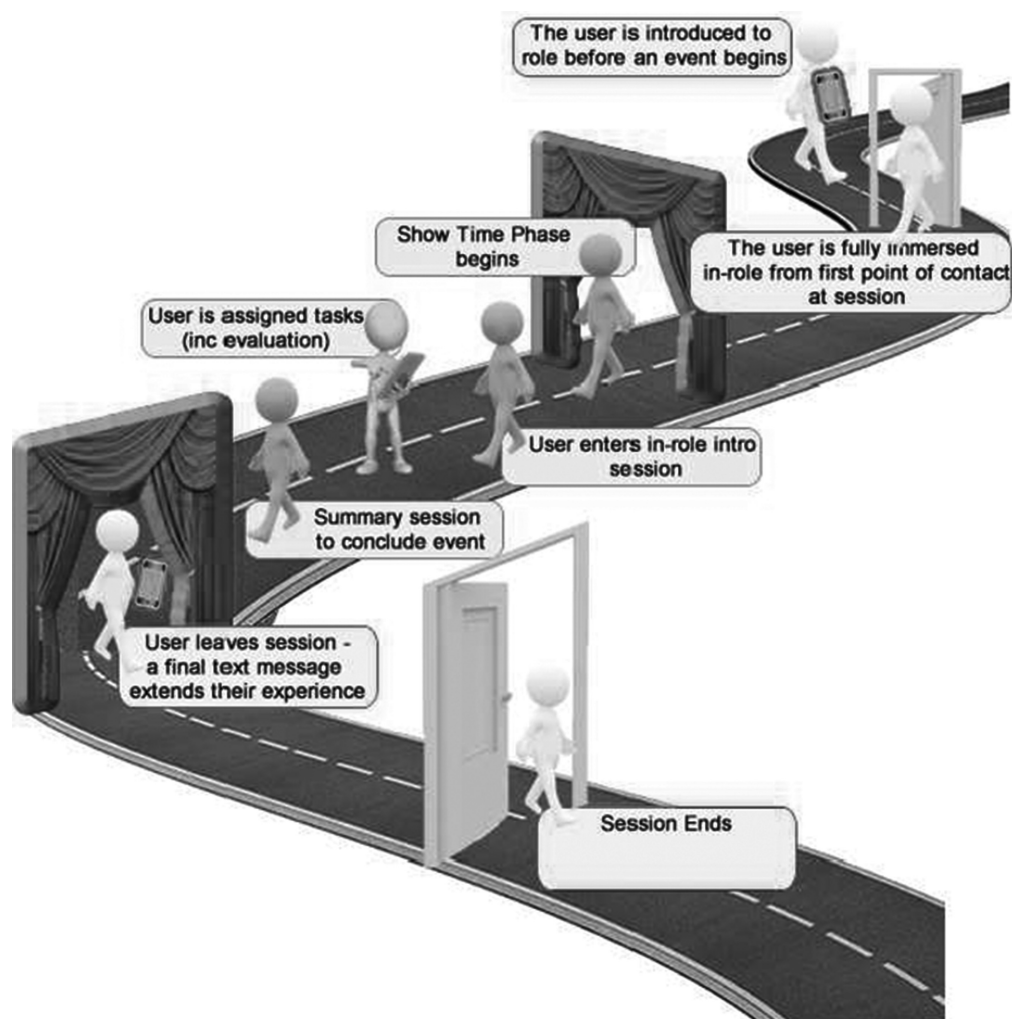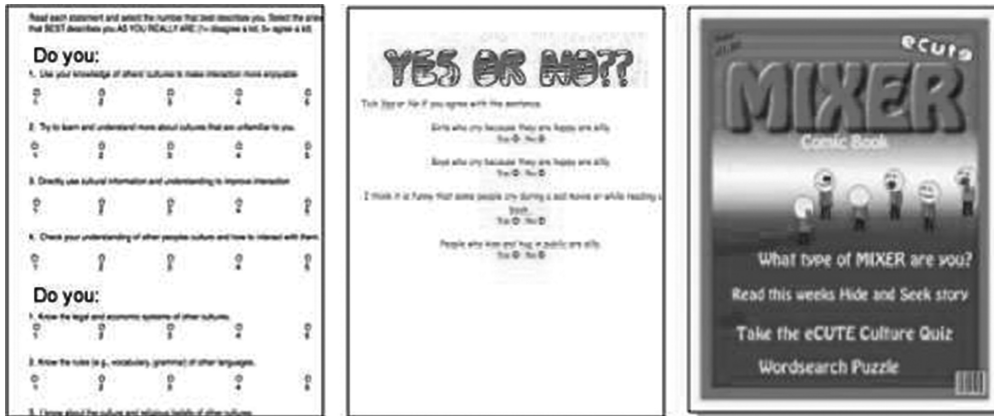
Figure 5.
A Transmedia Evaluation Event

drawing, ticking, circling, writing and using stickers) and the more they connected to the evaluand (e.g. relating to the narrative in the comic book), the more the children engaged and the higher the quality of data (e.g. completion rates and optimization). These initial positive results inspired the development of an engaging methodology designed to be value laden for the user.

THE TRANSMEDIA EVALUATION METHODOLOGY AND ITS APPLICATION TO THE MIXER SUMMATIVE EVALUATION

Figure 7 provides an outline of the Transmedia Evaluation methodology. This supports the development of an evaluation providing the plot (R&D perspective); role (intended user experience); props (evaluation battery and evaluand); and the script (experience protocol) required for a Transmedia Evaluation event. The event, and the elements within it, are rehearsed and refined (piloted,

Figure 6.

Basic, Better and Best materials from early stage evaluation

incrementally iterated), with the performance of the event (all aspects of user experience including evaluation, training (if required) and interaction) followed by a review phase (evaluation of event and data), which then feeds back into subsequent evaluations. The following sections further detail each of these stages, outlining how the approach was applied to the Summative Evaluation of MIXER, providing the final evaluation against the R&D objectives of the eCute project.
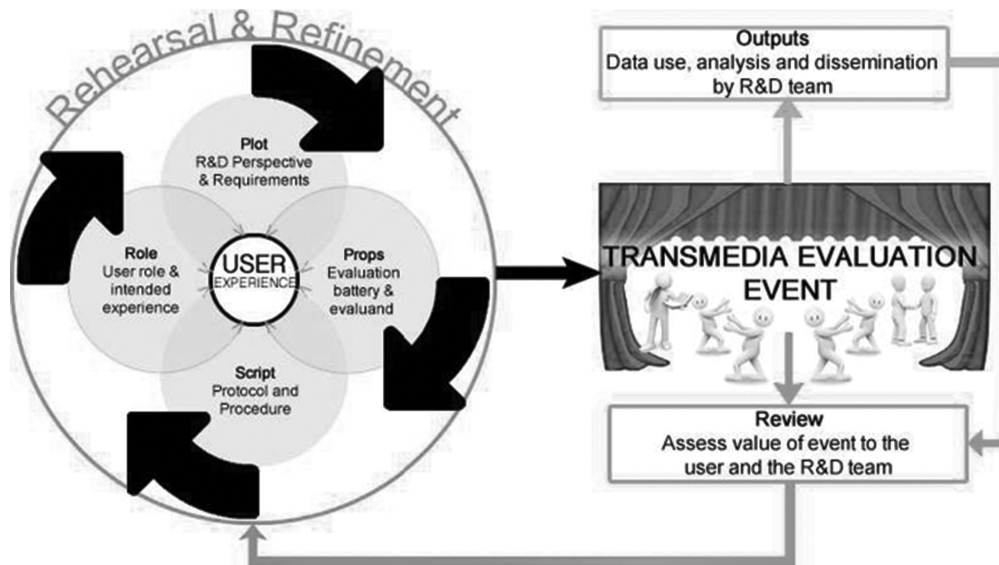
Figure 7.

Transmedia Evaluation Framework

The Plot—R&D Perspective & Requirements

Transmedia Evaluation begins with an outline "plot" providing specific R&D hypotheses, constraints (e.g. setting, participant numbers, interactions with

evaluand, training requirements) and empirical parameters (e.g. within group, between group; qualitative, quantitative). The R&D requirements provide the key elements that must be incorporated into the user experience to achieve an effective evaluation from the perspective of the R&D team.

The plot for the MIXER Summative Evaluation was to identify if the learning goals of the interaction as specified in the eCute Intercultural Competence Learning Framework (Swiderska, Krumhuber, Kappas, Degens, & Hofstede, 2011) were met:

- Emotional: MIXER supports children to recognise emotions (for example fear and anxiety) when dealing with the strange behaviours of another group
- Cognitive: MIXER supports children to start learning the specific practices and values of another group
- Behavioural: MIXER supports children in being fully present in attending to others verbal and non-verbal messages

A further goal was to determine whether the MIXER technology (e.g. intelligent agents, interaction modality, emergent narrative) was an effective approach for technology enhanced learning:

- Experience: MIXER engages children in the narrative and with the characters, supporting the children's understanding and learning of strategies for coping with intercultural conflict

The evaluation of MIXER's impact on children's learning provided the R&D requirements of a controlled randomized pre- post- design, collecting quantitative data to enable the assessment of far transfer (e.g. sustained learning). Within the test, R&D requirements identified the need for evaluation to include qualitative and quantitative measures to assess near transfer (e.g. immediate learning); and the user's response to the underpinning technology as provided by the characters and the interaction modality. Subsequent to the interaction, a reflective session to reinforce children's learning of intercultural conflict had to be incorporated into the evaluation design. The evaluation was designed to be classroom-based, involving 100+ children.

The plot for a Transmedia Evaluation Event is developed as a series of nodes or acts, within which users have to perform certain activities (such as interacting with the evaluand) or certain elements of the evaluation (e.g. participant information questionnaire). Plot development requires the evaluators to identify established instruments, data capture approaches and activities that can be used to assess and meet R&D hypotheses. Transmedia Evaluation advocates the use and/or adaption of existing measures and techniques wherever possible as this improves the reliability and validity of the data. Obviously,

there are contexts where no measure or activity exists, for example, in assessing evaluand specific hypotheses (e.g. assessing a user's comprehension of specific story elements in a storytelling application).

In the MIXER Summative Evaluation plot, the three measures aimed at assessing far transfer according to the specified learning goals were taken from the CQS—Cultural Quotient Scale (Ang et al., 2007); the MESSY Scale—Matson Evaluation of Social Skills (Matson et al., 2010) and Bryant's Empathy Index (Bryant, 1982). The behavioural subscale of the CQS was used as a pre- and post- measure of a child's capability to adapt verbal and nonverbal behaviour in different situations and cultures. Factor One from the Bryant Empathy Index focuses on understanding feelings and was used as a measure of children's empathic behaviour. Factor 2 "Social Skills/Assertiveness" of the MESSY questionnaire was used to assess the child's self-perception of their own social skills and competences.

Quantitative measures to assess the user's engagement, interaction and immediate learning were based on questionnaires developed for assessing the user experience in VLEs populated by embodied characters, based on Hall et al., (2013) and Hall, Woods and Aylett, (2006). Theory of Mind questions required by the R&D team used to assess children's advice to Tom are embedded in the conversation the child has with Tom, following the approach in (Hall, Woods, & Hall, 2009). For example, in advising Tom during the conflict incident, Tom asks the child what he should do, why, what makes the child think that will work, etc.

As the evaluation event is piloted (Rehearsal) and further developed, the plot is extended to incorporate specific instruments, activities and data capture approaches. Finalised instruments are supported with relevant protocols, merging them into the script. Coding frames and datasets are provided ensuring that analysis can begin rapidly after the evaluation event has finished.

The plot aims to provide a structure that is infused by the user role, by creating coherent and engaging props and scripts (see figure 8). The interaction with the technology and evaluation becomes one part of a connected, coherent experience for the user who is having a great time using new technology and participating in engaging activities.

Role—User Role & Intended Experience

The user is at the centre of a Transmedia Evaluation event, with the experience designed to meet the most basic user expectation of having an enjoyable time. Initial considerations of the user typically involve a review of current literature, applications, media and on-/off-line activities and experiences, using techniques frequently seen in persona creation. This exploration of the user's world aims to immerse the evaluators, inspiring and informing them about what interests and engages the target users.

The user role must be sympathetic to the evaluand, connecting with this in a way that is consistent, comprehensible and credible for the user. The user role may be an in-application role such as playing a character in a specified storyworld
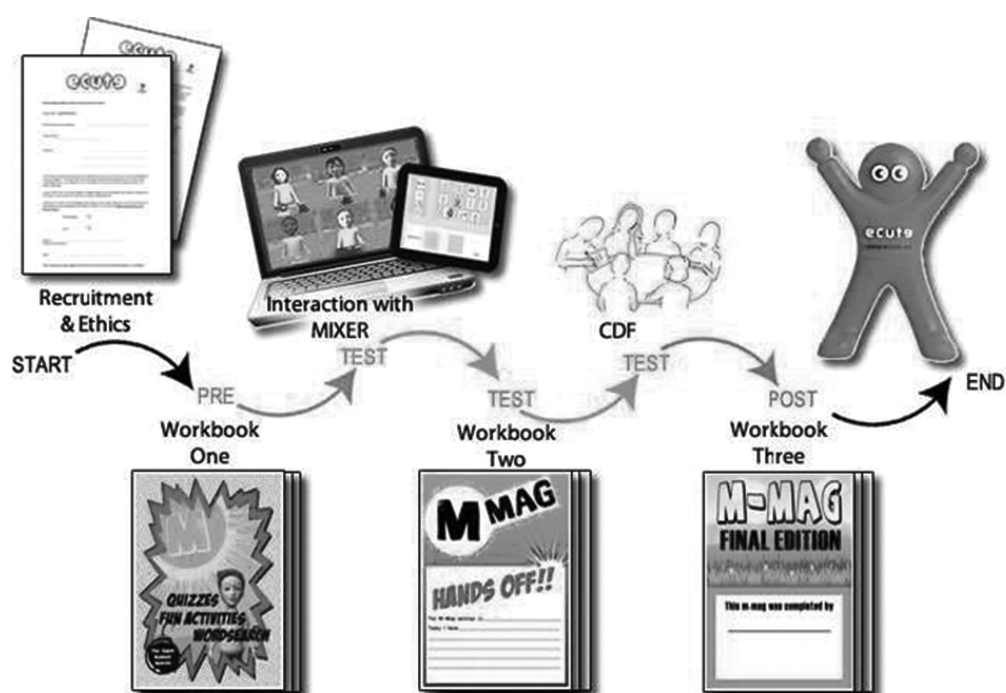
Figure 8.
MIXER Plot for Summative Evaluation

setting. It can also be in-task roles of learner, storyteller, player, etc.; traditional evaluation roles such as subject, designer or critic; and even that of the user's everyday self, effectively not changing role at all. Transmedia Evaluation places the user quickly in role for their experience, with recruitment reinforcing the user expectation of having a good time both by introducing their in-experience role and by highlighting their value to us. Recruitment must not only achieve informed consent, but must also reinforce the sense that the children are going to participate in something interesting, novel, important and relevant to them.

Typically for any evaluation, only a limited number of roles are possible. For instance, with a game if the user role was as player, then the evaluation experience and artefacts must become part of the game, by expanding the experience of the game world such as completing a self-rating scale as part of the entry requirements to a guild. If the user role was as critic, then the experience and artefacts must support the user in critical activity in a way that meets their expectations, for example completing rating scales (e.g. how many stars the game merits) or posting reviews to a Critic's Website.

The user role unifies the various elements of the experience, just as transmedia is unified by the overarching theme of the film or programme that it encircles. Ideas for user role, along with initial props, such as instruments and early versions of the evaluand, are piloted with the target user group gaining their input.

A range of user roles were considered for the users of MIXER, including:

- Related to the user's role in MIXER (invisible friend) with the child being Tom's friend in the evaluation experience
- Related to MIXER's storyworld (but not in the interaction) with the child's role being as camp counselor, for example.
- Related to MIXER's aim with the child's role being as a learner

Maintaining in-application user roles throughout the pre- and post- test phases, and particularly in incorporating the repeated 3 measures for far transfer and the learning reinforcement experience highlighted that fracturing of in-application roles was likely, thus rejecting placing the user in the role of Tom's friend. Although placing the user in a role such as camp counselor was considered, we decided against this as it implied that the user was operating at an expert level (e.g. already able to help and advise others) rather than as a novice learner. With the need to fit within the school day and to engage with children over multiple, separate sessions, we refined the user role as "learner" to the children's everyday role as a school-based learner, with the evaluation event being one of the children's lessons.

A review of information about children's interests, expectations and activities engaged in, informed the user role, with the aim not just for the child to have an average lesson, but rather a user role where the child is having an excellent experience using state-of-the-art technology to learn something different. Our interpretation of this role can be seen in the props detailed below.

Props (Evaluation battery & evaluand)

All evaluation instruments and approaches that are visible and require active participation by the user (as opposed to surveillance, covert data collection) are viewed as props, integrated into the plot and user role. Transforming the instruments into props is an incremental, iterative process and users are involved in the design and piloting of all evaluation props.

Each of the instruments and evaluation points identified in the plot is initially provided in a basic form. For example, with questionnaires the usual approach to administering the instrument is provided, this is often black and white, with numbered questions often with Likert rating scales or categories. Qualitative issues and questions (e.g. for interviews and focus groups) are listed. Techniques are provided as brief outlines, indicating required activities or outputs. At the start of the transformation process, a key issue to be addressed is "how appropriate is the intended instrument and battery for the intended user group?" With many questionnaires incorporating multiple sub-scales or factors and possible duplication between proposed instruments, the initial transformation ensures only necessary data is collected.

Once the instruments are provided, instrument refinement then permits further assessment and improvements if deemed necessary. Irrespective of aesthetic appeal, if questions do not make sense, seem repetitive or burdensome to answer, then user responses will be less optimal. As only representative users can tell you if the questions are appropriate, this transformation requires piloting with users.

An immediate issue in using the identified far transfer measures for the MIXER Summative Evaluation was the number of questions (with 104 questions across all three questionnaires) and the adult focus. Incrementally, with the R&D team, the instruments were refined, for example only using the behavioural subscale of the CQS. Sessions with users were held to improve the language and comprehensibility of the measures.

With the basic prop confirmed, the second level of transformation aims to reinforce the user role and to connect the prop to the evaluand. Qualitative data collection readily lends itself to the reinforcement of user role and integration into the plot enabling the collection of required data. In many studies, qualitative data is collected as written or spoken answers to open questions, with considerable flexibility as to how these questions are asked. There are many natural ways to incorporate such data collection into almost any user role and age group, using text (e.g. postcards, notebooks, posters), verbal (e.g. focus groups, interviews) and digital (e.g. texts (SMS), selfies, user-generated media) approaches. For example, if we put the user into the role of a 1900's news reporter with a history focused evaluand and then ask them to provide short, qualitative data about their learning (e.g. story comprehension, fact identification) via mobile phone to call a friend, the user's immersion with this out of place prop and reference to modern TV games shows would be ruptured. A more fitting prop would be a notebook into which the user could make notes on the events around them and then post these to an editor. Although the user sees nothing more than stage props in the story world experience, these items are actually the transformed evaluation materials that collect qualitative data and reinforce the user's role as in experiencing the 19th century context of the evaluand.

In the MIXER Summative Evaluation qualitative data was collected both from elements in the workbook (see figure 9) and also as part of a Classroom Discussion Forum (CDF) (Hall, Woods, & Dautenhahn, 2004) session about MIXER, held after the child had interacted with MIXER and completed workbook 2. The CDF session encouraged reflection and learning reinforcement on the children's experience with MIXER. This activity included typical in-role classroom activities with a Q&A session, table discussions (small groups based on typical classroom seating plan) and general discussion about MIXER and the experience.

Questionnaire transformation is guided by user expectations, for example using images, colour, layout, interaction modality and style to transform instruments. For instance, if the evaluand is a space-based game and the user-role as space cadet, then questionnaires can be given a space age look and feel, incorporated into the experience as part of the information needed to play
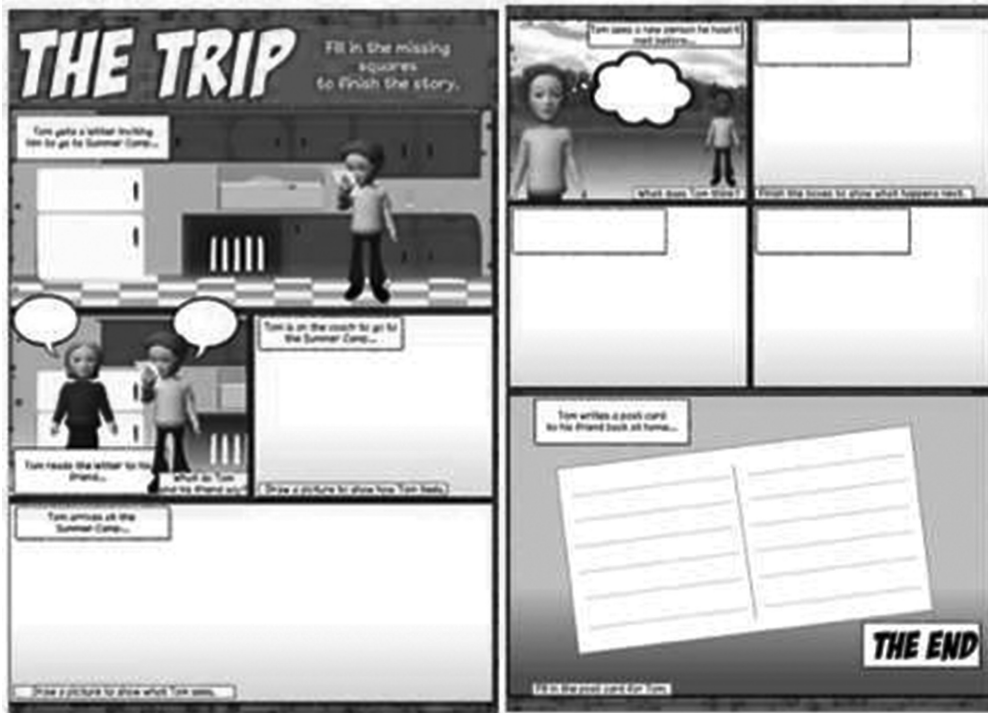
Figure 9.
Qualitative Data Collection in the MIXER Workbooks

the game. If our evaluand, was a child-focused tourist app providing facts and information about a stately home, we could ask the child to answer a quiz about their experience (e.g. showing retained learning), automatically receiving a badge on completion, thus resonating more clearly with user role as tourist.

Focusing on the user having an excellent experience both with MIXER and the evaluation, we found children appeared to enjoy responding to questions using a rating scale Rubie-Davies & Hattie, (2012). Inspired by child-focused hard-copy media aimed at recreational activity, such as comics, annuals and summer specials, we identified that children enjoyed: quizzes where they "discover" something about themselves; activities with interactive elements, such as colouring-in, using stickers and limited text entry (e.g. completing empty speech bubbles); and questions incorporating puzzles, such as wordsearches, mazes, spot-the-difference. With comic and activity books children expected a range of short, typically unrelated, complimentary, engaging and fun activities.

Although the media we sampled presented questions and activities with very different aesthetics, most comic books have the same elements, interspersed with, and themed by, the selling point of the comic, whether that is articles for pre-teen girls or more intrepid adventures for fans of Dr. Who. The techniques used to engage the user in comic books are relatively simple. Many activities incorporate vaguely relevant, but attractive, archetypal images (e.g. flowers and hearts; Dr. Who's sonic screwdriver); others use colour blocking to link facts or present

a group of related questions; motivators are also included such as directional arrows to move through an activity.

In the development of the props, we held questionnaire design workshops with children, both considering instrument design and to investigate whether providing the evaluation instruments in a comic book format was perceived as appropriate and engaging. However, the user role of school-based learner meant that the term comic book seemed inappropriate, conflicting and confused the role of learner with that of the role of comic book reader and fun-haver. Children instead suggested to us that we should call them workbooks, so that it was obvious that they were doing schoolwork. Using the term workbook also met with parent and teacher expectations, with many schools already using workbooks in the classroom.

Three workbooks were created for the pre-test (workbook 1), evaluation of immediate learning and experience of using MIXER (workbook 2) and the post-test. In many pre- post- tests, identical instruments (in content and format) are used. Instead, we wanted users to continue to engage with the questions rather than to feel a sense of *déjà vu* of having done all this before in Workbook 1. Thus, Workbook 3 presented a different appearance to incorporate the same questions and instruments, providing children with an engaging experience. Tables 1 and 2 provide the content of the workbooks with some sample pages in figures 10, 11 and 12.

In the pre-test (Workbook 1), children were provided with some additional activities, including a maze to help Tom get to the summer camp (preparing the children for their interaction with Tom) and The Trip (see figure 9), a comic strip activity in which the children are given half of the story of Tom being invited to go to camp. Children are asked to complete the empty thought and speech bubbles and comic book squares. The children also write out a postcard for Tom to send home. The trip provides qualitative data on the children's perceptions of going to new places and meeting new people along with how they think another child may feel when away from home. A wordsearch was included as the final activity for workbook 1 and a colouring activity in workbook 3, so that any children who finished ahead of the other children would have something to do whilst the rest of the class finished.

Workbook 2, (see table 2), collects data related to children's immediate learning (near transfer); their narrative comprehension, empathic engagement; and their perspectives and views of the MIXER characters and experience. Workbook 2 addresses all four of the MIXER goals identified in the plot.

Our approach to data collection transformation has had a significant impact on the user's perception of what they are doing. Users are usually unaware that they are completing questionnaires, being assessed on their learning or participating in a focus group for example, as the props that they are engaging with are embedded and just part of their in-role experience (Hall et al., 2013). We recognize that to any experienced evaluator they are clearly questionnaires and focus groups, however, this is not the user perception, with the instruments

Tableau 1.
Pre & Post Tests in Workbooks 1 and 3

Measures	Workbook 1 (Pre)	Workbook 3 (Post)
CQS	Which woodland animal are you? Designed as a quiz, with children rating which statements are like them and which not. Children are then identified as being a Badger, Fox or Deer, where all of the possible outcomes are constructively phrased and desirable for the children.	New People, New Places Children are given a series of images of mobile phones and asked to text Tom a number, 1 to 5, to tell him what they would do when making new friends
MESSY	New Friends The 20-item MESSY is designed to look like a puzzle, with children asked to help guide Ben to Barney. The cartoon bees are linked along a dotted line, interspersed with questions. The children move along this line "helping" to get Ben back to Barney.	The MESSY was divided into three separate sets of questions:The Epic Quiz - children identify on a scale how similar/dissimilar they are to the items. Friends: a series of questions providing learning about yourself. Maze Days: children make their way through a maze answering the questions as they go.
Empathy Index	Yes or NO Presented as a comic strip, with each frame offering yes/no responses and the children following the arrow to the next box. Although this was the same box, whether yes or no was selected, no child has ever mentioned this.	Think Fast Think fast is a sticker activity where children are provided with YES and NO stickers to use to answer the questions.

masked through adhering to user role and meeting the user expectations of that role. This was achieved with MIXER, with all props reinforcing the user role of school-based learner and ensuring that the children were having an excellent experience in that role. Children eagerly engaged with the workbooks, with 100% self-regulated completion. Children were very positive about all elements of the MIXER Summative Evaluation, with some children saying that they enjoyed the workbooks more than interacting with MIXER. From observation and discussion, throughout the MIXER Summative Evaluation children appeared to be as engaged with the evaluation battery as with the evaluand.

Script — Protocol & Procedure

The Transmedia script provides the experimental protocol and procedure for the evaluation event. The script implements the plot, ensuring that each plot node can be achieved, whether that be to engage in training, interaction or evaluation, whilst the user role (in-role expectations and user experience expectations) can be maintained and R&D requirements met through appropriate props and activities.
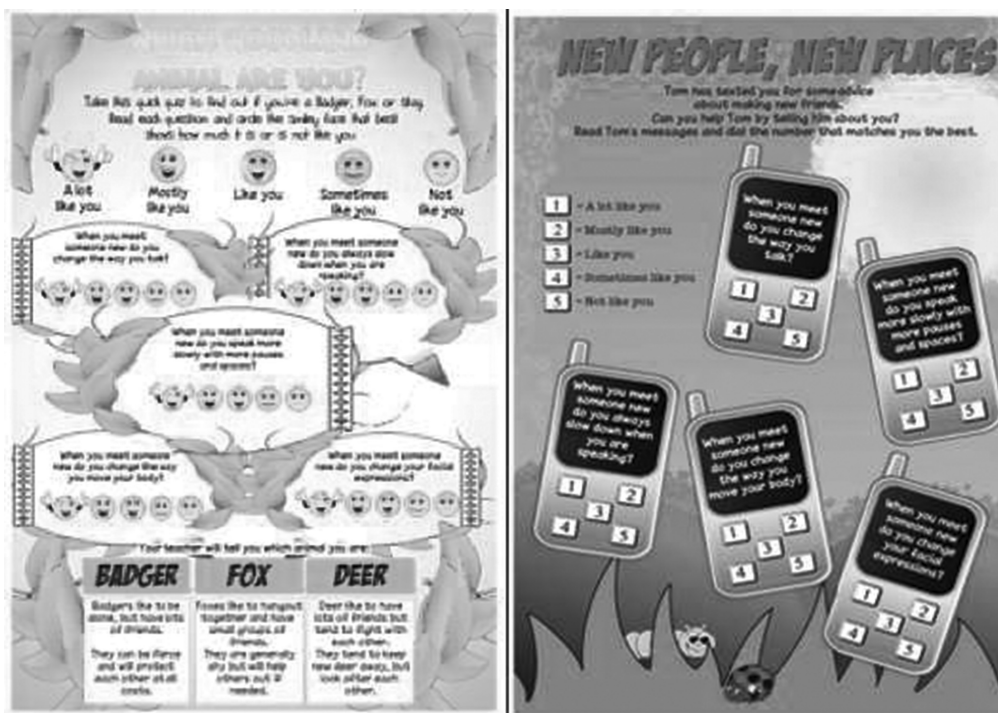
Figure 10.
Workbooks 1 & 3—MESSY

Figure 11.
Workbook 1 & 3 - Bryant's Empathy Index

Tableau 2.

Engagement Experience Questionnaire

Instrument	Outline	Rating Approach
Who Wins?	Having used MIXER, children should have engaged with and have a deeper relationship with Tom than any of the other characters. It was expected that the majority of children would choose to put Tom in first place. This relates to the emotional and behavioural learning objectives.	Children place stickers of their 3 favourite characters onto a picture of a winner's podium.
Roving Reporter	Comprehension/opinion exercise to assess children's narrative comprehension and engagement with Tom. Higher scores for narrative show that children listened and paid attention to the story line. Positive responses equating to cognitive comprehension and deeper engagement with Tom.	Varied ratings from yes/no responses, and circling correct answer
True or False?	Features 8 questions. 6 questions address engagement and comprehension, i.e. they have a correct true/false answer, equating to emotional, behavioural and cognitive learning. 2 questions gather children's opinions of the rule conflict reflecting the cognitive and behavioural learning outcomes.	The children use "True" or "false" stickers to answer the questions.
MIXER views	Features questions on user experience with MIXER (e.g. appropriateness of duration, desire to use MIXER again, etc.), equating to experiential learning.	Children circle one of the given responses.
What do you think?	Evaluates usability (e.g. voices, text, etc.) and experience (e.g. who explained the rules the best) of the MIXER application, relating to experiential learning.	Selections and Yes/No responses
iPad Page	Provides an evaluation of the interaction approach. e.g. "Do you think the game on the iPad was easy to use/not easy to use exciting/dull"	5-point Likert scale represented as faces.

Figure 12.
Workbook 2 - Engagement with Tom; immediate Learning Assessment; Interaction Modality Evaluation

The script unites the various elements of the evaluation into a single coherent narrative. The script of the event incorporates all of the user experience, including the initiation of an event, recruitment for the event and the completion of the event, typically a final engagement with the evaluators or the information relating to the next event. The finish point of a Transmedia Evaluation reinforces the user role and the expectation that their experience has been of value to the researchers.

Depending on user role, the script may have a theatrical focus, placing evaluators and researchers into in-context roles, for example as non-player characters in game evaluations with specific utterances. Or it may leave evaluators in a primarily researcher role, to cope with software failings for example. The script typically requires the evaluation team to explain certain issues or to say specific texts (particularly if the evaluation team take in-evaluation experience roles) and assumes a positive, constructive and upbeat approach from the evaluators. This upbeat approach is a vital part of evaluation, especially for children. The assumption and basis of the script in Transmedia Evaluation is that whatever role the evaluator is in, they will improve the experience for the user.

The MIXER Summative Evaluation script placed users firmly in their role as school based learners having a great time using educational technology. Initial recruitment of children involves explaining the evaluand, evaluation battery and researcher role through a script that highlights that in big technology enhanced learning projects we need to get users (e.g. them) to try out the learning materials. More detail explains that we are University researchers working on Personal, Social and Emotional (PSE), and we want the children to try out our technology and see if it works. Telling the children that the experience will include using an iPad gives it considerable appeal for the target age group. In the recruitment phase, the script clearly identified that the children were engaging

in an evaluation, with the ethics forms and the information accompanying them clearly stating that the purpose of the experience was a user evaluation.

Workbook 1 was completed during the pre-test. At the end of the session, the researchers explained that children would interact with some new learning technology using an iPad in their next session. In the script, is the instruction for the evaluators to 'prime' users to expect a good experience, to look forward to their next encounter, and to excite them about what will happen next.

Children interact with MIXER during the test phase followed by completion of, workbook 2. They also engage in a learning reinforcement session and qualitative data collection related to children's immediate learning, their engagement with MIXER and their satisfaction (enjoyment) with the interaction. The event concludes with the evaluators explaining the next meeting and priming children's anticipation.

Although the script initially incorporated a finish point where we returned to the school and provided results, the school requirements (related to Christmas Plays and seasonal events) meant that we could only realistically have 3 sessions, requiring the post-test to also provide the completion point.

Rehearsal & Refinement

Transmedia Evaluation requires an iterative, incremental method, with all elements of the user experience, such as the role, instruments, approaches and activities developed with design input from users and then piloted with representative users. With the focus of the evaluation being the provision of data to the R&D team we also pilot the data capture and analysis approaches, aiming to ensure that R&D expectations and requirements are met. Rehearsal is used to develop the evaluation instruments and experience in parallel to the development of the evaluand. As the evaluand develops from low-fi (pen and paper) versions to hi-tech (implemented system) so to does the evaluation.

The user is required to suspend disbelief, interpreting and achieving all aspects of their experience whist immersed and engaged in their role. Rehearsal identifies points where immersion may fracture identifying aspects of the experience that need improvement. All this extensive piloting identifies problems that can be resolved or reduced through appropriate experience design. Although rehearsal happens throughout the design of the experience and may frequently be targeted at specific elements, such as the instrument to capture the data, it is critical to regularly have rehearsals of the entire experience to ensure that the event works as a whole performance and not just in parts.

The MIXER Summative Evaluation was the culmination of 3 years of work for the R&D team. The 3 workbooks and the CDF had been extensively piloted with users, with an initial workbook design piloted in the first year of the project. The Summative Evaluation had a large-scale pilot as the final rehearsal with results highlighting a significant flaw in our experience design through placing the discursive and qualitative activities at the end of the experience, rather than at the end of the interaction (Aylett et al., 2014). R&D input and discussions

with teachers highlighted the need to change the experience design to reinforce intercultural learning soon after the MIXER interaction, rather than after the entire experience. Children's response to MIXER was the expectation that they would get a chance to talk to each other and us about MIXER straight away rather than a week later. This also met with R&D team expectations, as immediate user response to the interaction was more valuable than their memories of the experience.

Performance (Transmedia Evaluation Event)

A Transmedia Evaluation event incorporates all of the user experience, from recruitment to completion. Singular one-off experiences (e.g. interaction followed by user satisfaction questionnaire or learning assessment) or longitudinal designs are supported (e.g. as with multiple episodes as required by a pre- post-test design).

All of the elements (plot, role, props and script) feed into the event phase during which the participants and evaluators are in role and the evaluation occurs. The event is the shortest phase of an evaluation, with the procedure, instruments, data capture and evaluand all prepared, rehearsed and refined. After the event, the data is prepared for analysis following the specified protocols and analysis begins.

The script provides both the protocol and instructions to the evaluators for how the evaluation is to occur, providing the detail underpinning the plot. With MIXER (see figure 8), the script has the following nodes:

- **Start**: With the MIXER summative evaluation we met with children prior to the Pre-test, for a brief 10-minute session at the beginning of the school day. Our instructions were to introduce ourselves, the project and the experience, with the bottom line being to enthuse the users about their experience. We briefly explained that they would be completing some workbooks, we showed them these from the front of the class, and that they would get to use MIXER where they would meet Tom, some images shown. We told them we'd be coming three times (pre, interaction, post). The ethics documentation and experience information was provided to children and the school provided us with the completed ethics forms prior to the pre-test.
- **Pre**: The pre-test involved the children being given the workbook and asked to complete it. Children worked individually on their group tables. At the end of the pre-test the children were briefly told what the next session would include. 100% self-regulated completion was achieved. The final activity of workbook one was a time filler, a word search used for those children who completed the questionnaire quickly.
- **Test**: The limited equipment and school requirements resulted in children interacting with MIXER (individually) in small groups of 4 in the library near to the classroom. Children not using MIXER were engaged in class-based activities with the teacher. Once all children had interacted with MIXER and

completed workbook 2, the learning reinforcement session stimulated debate and discussion, naturally moving through the various qualitative questions, relating to the children's experience of MIXER.

- **Post-Test**: At the beginning of the school day the workbooks were distributed and completed. After completion, the evaluators then explained what would happen with the results, highlighting the value of the children's input for understanding technology enhanced learning, hoped that they had had a good time and thanked the children, providing each child with an eCute mascot.

Outputs & Review

The review phase assesses the event in relation to the outputs, that is the results achieved and their use by the R&D team. The review phase of the Transmedia Evaluation also provides the evaluation team with the opportunity to reflect upon the evaluation, considering what aspects of the event went well and what could be improved. This then feeds into the design of subsequent evaluations, identifying successful activities.

The MIXER summative evaluation was a very successful experience for all concerned. From an R&D perspective, the data was complete, of high quality and from engaged participants. An overview of the results is presented in table 3.

Qualitative data collected in the learning reinforcement classroom session highlighted children had really engaged in the experience provided to them via the Transmedia Evaluation. Children were enthusiastic about all elements of the MIXER Summative Evaluation, including the interaction, the workbooks and the class-based discussion. All the instruments incorporated into the workbooks were 100% completed. At no stage during the completion of the workbooks did any child ask for help or support in completing the activities. Not only did the evaluation identify that children enjoyed using MIXER, but additionally that children successfully engaged in experiential learning, empathically engaged with the characters and enjoyed their experience of the evaluation.

The MIXER Summative Evaluation resulted in publishable outcomes. Results enabled the R&D team to highlight that children exhibited both near and far transfer, meeting the learning goals of the eCute Intercultural Learning Framework and contributing to the Excellent rating achieved by the eCute project in its final review with the European Commission.

DISCUSSION

Transmedia Evaluation provides a different approach to evaluation than that commonly seen in the design of evaluation experiences for children. Instead of evaluation being a discrete task performed with traditional approaches, Transmedia Evaluation creates an engaging, coherent, integrated experience. All

Tableau 3.

Summary of the main results from the MIXER Transmedia Evaluation

| Learning Goal	Learning Objective	Near Transfer Learning Goal Achieved? (EEQ & CDF measurements)	Far Transfer Learning Goal Achieved? (CQS, Bryant's Empathy, Messy)
Emotion Goals	Be able to recognise emotions (for example fear and anxiety) when dealing with the strange behaviours of another group.	**YES** - Children showed a preference for the characters that they interacted the most with, and those that displayed the most narrative. **YES** - Children wanted to be friends with "Tom" and felt that he had listened to them, which demonstrates the ability of children to recognise their different emotions during MIXER towards Tom (in-group) compared to their emotions towards the "yellow" team (out-group). **YES** - Children demonstrated high levels of engagement with the MIXER software. They thought MIXER was fun, and many children wanted the interaction to be longer. **YES** - Nearly 90% of children expressed a desire to use the MIXER software again.	**NO** – Children's empathy levels remained constant between pre-post-test after interacting with MIXER. **NO** – Messy (social interaction ability) scores were unchanged between pre-post-test after the MIXER interaction.
Cognitive Goal	Start learning the specific practices and values of that group.	**Some evidence** – Ability to comprehend the "rule-change" between the red and yellow teams demonstrated by some children. But, nearly 50% of children did not appreciate the "cultural differences" of the "yellow" team rules just being different and not cheating.	

Tableau 3.

(Contiuned)

Learning Goal	Learning Objective	Near Transfer Learning Goal Achieved? (EEQ & CDF measurements)	Far Transfer Learning Goal Achieved? (CQS, Bryant's Empathy, Messy)
		YES - High levels of comprehension as children understood the events and progression through the game of Werewolves. **YES** - Children understood the MIXER application and its narrative.	**YES** – Children's CQS scores were higher at post-test after the MIXER interaction. Provides some evidence that children had started to learn conceptually about the values and attitudes of the MIXER characters.
Behaviour Goal	Being fully present in attending to others verbal and non-verbal messages.	**YES** – Children who wanted to be friends with Tom also felt that they had helped Tom and that Tom had listened to them. **YES** – Children who believed they had helped Tom believed Tom knew what he was doing and felt that Tom was good at Werewolves. **YES** – Positive association between children believing Tom had listened to them and Tom having fun in Werewolves.	**YES** – children's CQS and MESSY scores were **higher** after interacting with the MIXER software at post-test (T2). This suggests children started to adapt and modify their behaviour/facial expressions/vocalisations to the novel MIXER scenarios.

elements of the event are visible to the user as part of a consistent experience, facilitating the user in adopting and maintaining their assigned role in the event, (see figure 13). Transmedia Evaluation aims to provide a methodology that represents and respects children's interests.

Transmedia Evaluation enables the generation of appropriate results for the R&D teams and research requirements to be met. The plot of a Transmedia Evaluation Event provides the purpose of the evaluation, defining the constraints and requirements that ensure that the appropriate data is gathered. In the MIXER Summative Evaluation detailed in this paper, the plot nodes were recruitment, learner baseline measurement, interaction, evaluation of learner's immediate learning and engagement with MIXER, reinforcement and reflection of intercultural learning and post-test measurement. Specific instruments

included in the pre- and post- tests were Bryant's Empathy questionnaire, the MESSY measurement of social interaction skills and Ang's Cultural Intelligence Questionnaire. Immediate learning was assessed by the CEQ and CDF. Ethics requirements were met (requiring children and parents to complete relevant forms) and included the collection of quantitative (using identified measures), qualitative (using open questions in written format and classroom based discussions), observation (recordings of children engaging with MIXER and the evaluation) and interaction (children's interactions with Tom) data. This sounds like an arduous and disengaging plot, however, by placing the plot within a script where the user played an engaged, motivated school based learner having an excellent experience, completely renegotiates the evaluation space. Perceiving an evaluation as adding to the evaluation experience, rather than simply assessing completely changes the dynamic of the evaluation process.

The plot provides the structure and the key elements of the experience. It is all too easy to lazily respond to research questions by providing a traditional, disruptive evaluation experience that meets R&D requirements but shows little thought for either the user or the evaluand. However, by responding to everyday user expectations (which for children are ALWAYS to have an excellent experience) and the expectations that the user role creates (adding context and/or value to the evaluand), as outlined for the MIXER Summative Evaluation, a vastly different experience can be achieved for the user. Rather than the child experiencing the role of subject completing arduous evaluation instruments,
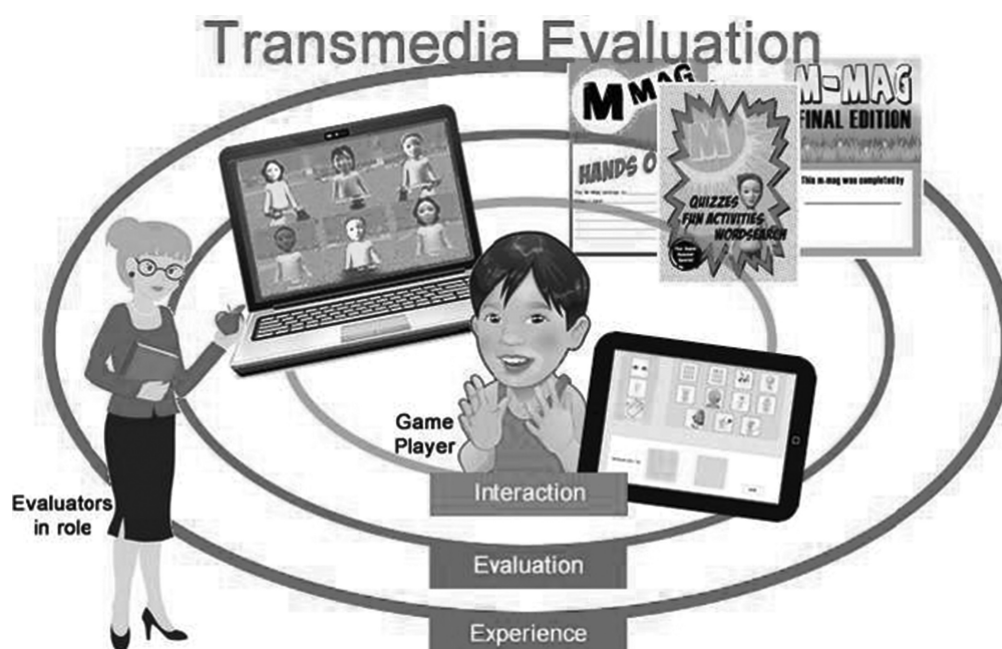
Figure 13.
Child's Experience of the MIXER Summative Evaluation

instead they are experiencing an excellent lesson, designed and crafted around their engagement.

Applying the Transmedia Evaluation methodology to large scale R&D evaluations with children naturally highlights the need for multiple methods to collect data. This not only meets the R&D requirements for a variety of data types, including logged, quantitative and qualitative data. But, further it meets children's expectation of a variety of interesting, diverse activities rather than responding to a group of semi-identical questionnaires for 20 minutes. In the MIXER Summative Evaluation, mixed methods included logged data of the child's interactions with Tom providing responses to Theory of Mind questions; mainly quantitative but also written/ drawn qualitative data in the workbooks; and the CDF verbally assessing immediate learning and the children's experience.

The dominance of questionnaire measures and their blanket acceptance within the research community essentially demands that large R&D evaluations of interactive systems incorporate questionnaire instruments. Repetition in questions and assessment style (in structure/approach rather than content) seems contrary to our vision of evaluation. Yet, as we have detailed in the MIXER Summative Evaluation, initial versions of the questionnaires did present with a similar style, look and feel. Our transformation of these instruments was effective as none of the children who participated in the evaluation were even slightly aware that they were completing questionnaires. Nor could children be prompted to discuss the experience as anything but engaging with a fun workbook. All of the workbooks (357) were 100% complete, with variation in answers, identifying that children were engaged in providing optimal answers, rather than satisficing or adhering to acquiescence bias. Many of the children would have enjoyed further workbooks and mentioned this in the CDF. This is a significant outcome for us as we are unaware of any other evaluations where child users have asked to fill in more questionnaires.

Transforming both qualitative and quantitative data capture instruments into engaging elements of the user experience is achievable. It requires evaluator investment in understanding user expectations by reviewing literature and experiences and from exploring the appropriateness of the evaluation with the users themselves. Including children in the evaluation design is essential, instruments and techniques must be piloted and children's ideas incorporated ensuring that the expectation of having a great experience is realized. Children's input can be insightful and improve our approaches. For example, with the MIXER Summative Evaluation children's input resulted in the instruments being called workbooks and for the focus group to be re-conceptualised as a Q&A session with raised hands as the best way to gain class-based feedback.

In Transmedia Evaluation we strongly advocate the use of established measures. It is the transformation process that is central to providing an engaging evaluation experience, typically reducing and refining them and almost always changing how they are presented to the user. Clearly, our transformation of

measures may have an impact on the validity and reliability of the chosen instrument. However, without doubt using existing measures and techniques will increase the quality and reliability of the data, rather than a researcher creating a new (usually barely piloted) instrument. In practice, the transformation of instruments is relatively straightforward adhering to the more achievable aspects of engagement such as age appropriateness, length of experience, and aesthetic appeal. Graphical approaches, such as attractive layout and relevant images, significantly change how questionnaires are received. Incorporating 'story' elements from the evaluand into the design and simply improving the appearance to include age-appropriate design improves the evaluation experience for children.

However, with Transmedia Evaluation, the central intention is to engage the user throughout the evaluation experience with equal immersion through all process steps, not just to provide age appropriate instruments nor prettify the evaluation battery. Although well-designed instruments do increase engagement, to ensure immersion and to prevent an experience rupture or dislocation, more is needed than cosmetic improvement. It is essential that the experience is considered at a meta-level, creating an overarching theme, consistent with the child's expectations and the narrative of the evaluand. Only then, do the instruments become analogous to transmedia, encircling the interaction and building the story and experience for the user.

Placing the child at the centre of the evaluation in a clearly defined user role inspires the design of props and scripts. With user role as the guiding principle, instead of designing data collection tools we are designing elements of an experience. The props fit with the role and evaluand, reinforcing the sense of a single, coherent experience for the child. The script seamlessly integrates all elements of the user's experience, including recruitment, training, interaction and evaluation, binding together the various elements into a single coherent narrative. For children, Transmedia Evaluation creates an effective, enjoyable evaluation, where interaction and data collection support the child's immersion into a coherent, engaging and enjoyable experience.

Transmedia Evaluation provides an optimal experience both for children and R&D aims with regards to the quality and richness of data collected. As highlighted above, a potential shortcoming of Transmedia Evaluation may be that issues surrounding instrument validity and reliability are not fully adhered to. Future research should aim to carry out comparative Transmedia Evaluation Studies with traditional pencil-paper approaches to determine whether this is the case. This paper has clearly presented that Transmedia Evaluation provides an engaging and immersive experience for children. Future studies should extend the Transmedia Evaluation approach to include other participant cohorts with adults and other research disciplines.

CONCLUSIONS

This paper clearly demonstrates that by placing children at the centre of evaluation design and meeting their expectation of enjoyment, we can both respect and represent their interests in the evaluation process. Applying the Transmedia Evaluation process to the MIXER Summative Evaluation generated relevant, high quality results for the R&D team, through transforming and integrating the evaluation into an excellent, coherent experience for the child. Transmedia Evaluation provides an innovative, effective evaluation methodology that enriches the evaluation process, generating high quality data whilst providing a different, enhanced experience for children.

ACKNOWLEDGEMENT

RÉFÉRENCES

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. doi:10.1111/j.1740-8784.2007.00082.x

Aylett, R., Lim, M. Y., Hall, L., Endrass, B., Tazzyman, S., Ritter, C., . . . Kappas, A. (2014). *Werewolves, Cheats and Cultural Sensitivity. In 2014 International Conference on Autonomous Agents and Multiagent Systems*. Paris, France: International Foundation for Autonomous Agents and Multiagent Systems, 1085-1092

Babbitt, B. & Nystrom, C. (1989). *Questionnaire construction manual annex. Questionnaires: Literature survey and bibliography*. U.S. Army Research Institute: Fort Hood, Texas.

Bryant, B. (1982). An index of empathy for children and adolescents. *Child Development*, *53*(2), 413-425.

Buckleitner, W. (1999). The State of Children's Software Evaluation—Yesterday, Today, and in the 21st Century. *Information Technology in Childhood Education Annual*, *1*, 211-220.

Endrass, B., Hall, L., Hume, C., Tazzyman, S., & André, E. (2014). *A Pictorial Interaction Language for Children to Communicate with Cultural Virtual Characters. Human-Computer Interaction - Advanced Interaction Modalities and Techniques*, Proceedings 16th International Conference HCI International 2014, LNCS 8511 (1), Heidelberg: Springer. 532-541.

Endrass, B., Hall, L., Hume, C., Tazzyman, S., Andre, E., & Aylett, R. (2014). Engaging with virtual characters using a pictorial interaction language. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, New York: ACM Press, 531-534.

Gomez, J., & Pulman, S. (2012). In Ridley Scott's "Prometheus," the Advertising Is Part of the Picture.

Hall, L., & Hume, C. (2011). *Why Numbers, Invites and Visits are not Enough: Evaluating the User Experience in Social Eco-Systems.* In SOTICS 2011, The First International Conference on Social Eco-Informatics, 8-13.

Hall, L., Jones, S., Aylett, R., Hall, M., Tazzyman, S., Paiva, A., & Humphries, L. (2013). Serious Game Evaluation as a Metagame. *Journal of Interactive Technology and Smart Education*, *10*(2), pp. 130-146.

Hall, L., Woods, S., & Aylett, R. (2006). FEARNOT! Involving Children In The Design Of A Virtual Learning Environment. *Journal of Artificial Intelligence and Education: Special Issue on Learner Centred Methods for Designing Intelligent Learning Environments*, *16*(4), 237-251.

Hall, L., Woods, S. and Dautenhahn, K. (2004) FearNot! Designing in the Classroom, in Fincher, S., Markopoulos, P., Moore, D., & Ruddle, R. (Eds.) *People and computers XVIII: Design for life*, Vol. *2* (Proceedings of HCI 2004), Heidelberg: Springer.

Hall, L., Woods, S., & Hall, M. (2009). Lessons Learned Using Theory of Mind Methods to Investigate User Social Awareness in Virtual Role-Play. *Journal of Human Technology, Special Issue on The End of Cognition*, *5*(May), 68-89.

Jenkins, H. (2011). *Transmedia 202: Further Reflections*, Blog entry August 1st, 2011. Available at: http://henryjenkins.org/2011/08/defining_transmedia_further_re.html

Jensen, J. J., & Skov, M. B. (2005). A review of research methods in children's technology design. In *Proceeding of the 2005 conference on Interaction design and children - IDC '05*, New York: ACM Press. 80-87. doi:10.1145/1109540.1109551

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New Directions for Evaluation*, 29-44. doi:10.1002/ev.1033

Matson, J. L., Neal, D., Fodstad, J. C., Hess, J. A., Mahan, S., & Rivet, T. T. (2010). Reliability and validity of the Matson Evaluation of Social Skills with Youngsters. *Behavior Modification*, *34*(6), 539-58. doi:10.1177/0145445510384844

O'Brien, H. L. & McLean, K. (2009). *Measuring the User Engagement Process.* Paper presented at Engagement by Design Pre- Conference Workshop, CHI 2009 Digital Life New World, Boston, MA, April 5, 2009.

O'Brien, H., & Toms, E. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, *61*(1), 50-69. doi:10.1002/asi.21229.1

Ólafsson, K., Livingstone, S., & Haddon, L. (2013). *Children's use of online technologies in Europe: a review of the European evidence base.* London: LSE & EU Kids Online.

Parker, J., & McDonald, R. (2014). *Stories are More than Paper: Using Transmedia with Young Adults.* IFLA World Library and Information Congress, WLIC 2014, Lyon, France.

Read, J. C., & MacFarlane, S. (2006). *Using the fun toolkit and other survey methods to gather opinions in child computer interaction. Proceeding of the 2006 Conference on Interaction Design and Children IDC 06*, New York: ACM Press, 81-88. doi:10.1145/1139073.1139096

Read, J. C., Markopoulos, P., Par, Eacute, S, N., Hourcade, J. P., & Antle, A. N. (2008). Child computer interaction. *Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems*, *2*, 2419-2422. doi:10.1145/1358628.1358697

Rubie-Davies, C. M., & Hattie, J. A. C. (2012). The dangers of extreme positive responses in Likert scales administered to young children. *The International Journal of Educational and Psychological Assessment*, *11*, 75-89.

Sapouna, M., Wolke, D., Vannini, N., Watson, S., Woods, S., Schneider, W., . . . Aylett, R. (2010). Virtual learning intervention to reduce bullying victimization in primary school: a controlled trial. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *51*(1), 104-12.

Swiderska, A., Krumhuber, E., Kappas, A., Degens, N., & Hofstede, G. J. (2011). *D2.1 Preliminary cultural learning interdisciplinary framework*. Available at: http://ecute.eu/downloads/deliverables/

Weiler, L. (2012). *Building Storyworlds: The Art, Craft & Biz of Storytelling in 21C*. Available at: http://www.buildingstoryworlds.com

Zaman, B., Vanden Abeele, V., Markopoulos, P., & Marshall, P. (2012). Editorial: The evolving field of tangible interaction for children: The challenge of empirical validation. *Personal and Ubiquitous Computing, 16*(4), 367-378