Integrated Questionnaires: Maintaining Presence in Game Environments for Self-Reported Data Acquisition

Julian Frommel, Katja Rogers, Julia Brich, Daniel Besserer, Leonard Bradatsch, Isabel Ortinau, Ramona Schabenberger, Valentin Riemer, Claudia Schrader, Michael Weber
Ulm University
Ulm, Germany
{firstname.lastname}@uni-ulm.de

ABSTRACT
Research in human-computer interaction often requires the acquisition of self-reported data. Particularly concerning serious games, the interaction between the game and the user still holds many unknown aspects, partly due to the user's double role as player and learner. An easy way of collecting data consists of questionnaires, mostly employed in pen-and-paper or electronic form. In order to gather data points during game play, the player is interrupted, potentially causing unintentional side effects. We suggest an integration of questionnaires into games as game elements, in order to mitigate the effects of interruption. A serious game prototype with an integrated survey was implemented, and evaluated regarding its effects on the players' experience of presence.

Author Keywords
integrated questionnaires; self-reported data; presence; player assessment; serious games; immersion

ACM Classification Keywords
K.8.0 General: Games; H.5.2 User Interfaces: Evaluation/methodology

INTRODUCTION
Data collection is a crucial part of the design, development, and evaluation of games. From the acquired data, developers can extract feedback and guidelines to iteratively better the game prototype, while researchers can validate their hypotheses, for example regarding interaction patterns or affective states, among many other aspects. As the automatic acquisition of user data remains problematic in many aspects (e.g. regarding accuracy, interpretation, and handling contradictory data), many researchers instead focus on the acquisition of self-reported data via questionnaires. The timing of the questionnaires leads to further limitations. The administration of questionnaires after game play may exacerbate the issues of accuracy, as the data is reported retrospectively. The administration of questionnaires during game sessions may increase accuracy, but may also lead to distortion due to the interruption of the game play. At best, game play is briefly halted until the player has entered the required data in an electronic survey; sometimes, however, the player is physically moved to a different location, where they are presented with a pen-and-paper survey.

These interruptions may negatively affect players' immersion, and presence. These factors have been shown to influence learners' execution of tasks and motivation [5]. Serious games in particular seek to maximise these factors; therefore, interruptions due to questionnaires may pose a problem to game research.

To address this issue, we suggest a concept in which questionnaires are integrated into games as game elements, in order to mitigate the effect of interruption. The contributions of this paper are presented as follows:

- A concept to maintain presence while collecting self-reported data in games
- A prototypical serious game for the practise of vocabulary recognition, in which the proposed concept is employed
- An evaluation of the concept with an empirical user study
- A discussion of the concept and future research approaches

BACKGROUND
Like many other computer science areas, the field of games research and development increasingly produces large datasets. This section discusses some issues regarding data acquisition in and for games development, and then outlines related psychological phenomena.

Data Collection In Games
Nowadays, many games collect data about player behaviour and interaction. Pre-release user testing and in-situ monitoring of player behaviour result in extensive amounts of data, which are then analysed, e.g. for churn rate prediction in revenue-based online games [21]. User data is required by both academic and commercial game development, yet the acquisition thereof is still largely non-automated and non-standardised. Furthermore, the commercial datasets usually remain under strict proprietary control. Information on the scope of user data collection in commercial games development is usually either difficult to procure or outdated, and few companies discuss how they utilise the information gained. Some commercial games, however, incorporate explicit user query methods into their games' narrative. Notably, *Silent Hill: Shattered Memories* was openly promoted as game play

CHI PLAY 2015, October 03 - 07, 2015, London, United Kingdom.

ACM 978-1-4503-3466-2/15/10…$15.00
DOI: http://dx.doi.org/10.1145/2793107.2793130

customised for the player's personality. The player is confronted by a psychological profiler in the opening scenes. Based on the player's interaction in the dialogue and tests (as well as subsequent in-game monitoring), the game adapts a number of features (e.g. the storyline and NPC characteristics) to play toward or against the player's personality [20].

For researchers, user data is hard to collect in similar depth and breadth, which limits the amount and variety of adaptation they can implement and research. Instead, some game researchers use web-crawling techniques to gain partial data sets for games with otherwise heavily restricted game metrics (i.e. the more popular large-scale commercial games) [17]. Other research approaches prefer to focus on automatic data collection techniques. The automatic detection of the player's affective states, i.e. emotion recognition, remains fraught with considerable limitations, and requires training, or the amalgamation of multiple potentially contradictory sources. For example, emotion recognition via neural network classifiers may require lengthy training sessions to detect emotions from facial expressions or speech patterns [7]. Visual detection may require additional facial markers, and is frequently hindered by features such as facial hair or glasses [6]. The addition of biofeedback methods (e.g. to measure heart rate, or brain activity) can improve the accuracy measurably [10], but requires the above mentioned merging of the data sources, as well as additional measuring equipment which may face user acceptance issues in the gaming community.

Frequently, these issues necessitate the acquisition of self-reported data, for example via questionnaires. The use of self-reported questionnaire data in research is accompanied by some inherent limitations, for example, issues of retrospective accuracy, trustworthiness of the user (particularly for remote questionnaires), and randomness (i.e. guessing behaviour) [28]. Nevertheless, questionnaires are often the only – or most efficient – option of collecting user data; the resulting self-reported data then commonly consists of Likert scale items, association pairs, and a number of other survey methods. The acquisition of accurate user data often requires timely assessment; i.e. the game play is interrupted for the questionnaire [3]. Sometimes this process even quite literally physically removes the user from the game medium (e.g. the PC) to a different location in order to fill out a pen-and-paper questionnaire. Such interruptions are known to negatively affect the learning process and may complicate the assessment of user performance [4]. Shute et al. thus suggest using less obtrusive methods of assessment [24], in order to support learning and maintain *flow*. The psychological phenomenon of flow has been studied in great detail, but for our purposes can be defined as an optimal state of performance [8].

Immersion and Presence

At this point, researchers must often differentiate between two other concepts that are closely related thereto: immersion, and presence. The term "immersion" originated in the virtual environments research area, and was defined by Witmer and Singer as "perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences" [30]. It has since often been described in terms of technology [25], requiring extensive and surrounding visual displays [18], and/or physical sensory information [15]. Today's definition seems to no longer include a PC, whereas Witmer and Singer thought that even "a book, movie, or video game may permit some immersion". They also singled out natural interaction, and decreasing outside distractions, as factors that may increase immersion. Presence, however, has largely been defined as a cognitive phenomenon: "the subjective experience of being in one place or environment, even when one is physically situated in another" [30], determined by both media and user characteristics [2]. As it is largely a construct of the mind, researchers have come to the conclusion that presence is independent of the technology, and can be "present even in the impoverished environments [. . .] that some VR currently provides" [2]. The emotional involvement and engagement provided by the environment's narrative can "elicit a high sense of presence" even in non-immersive systems [2, 16, 22]. Weibel and Wissmath examined the relationship between presence and flow in the context of three different game genres, and concluded that both depend on motivation and immersive tendency of the user, and both enhance performance and enjoyment [27].

Despite much research, the interconnection between immersion, presence, and flow remains unsettled. Nevertheless previous research has indicated that interruptions are not conductive to flow, immersion, or presence, which in turn seem correlated with learning and performance [30]; players' virtual presence has also been connected with more concentrated execution of tasks, and higher motivation in virtual environments [5]. Motivation and positive emotions have been connected with a facilitating effect on learning, and are thus particularly important for serious games [26]. Given that the above mentioned previous research indicates that natural interaction with the game's virtual environment should increase immersion and presence, we hypothesise that integrating questionnaires as game elements (operated in the same natural manner as the game itself) should better maintain presence in comparison with an overlay questionnaire.

CONCEPT

We argue that when players experience presence in a game, they experience the virtual world as a natural environment. To maintain that sense of presence, the questionnaire must fit into that world as naturally as possible. Thus, the factors that contribute to presence (e.g. see [30]) will also be crucial to maintain presence in our questionnaire scenario. Further, we expect interruptions of the game play to negatively influence presence and immersion, as they do not fit the natural environment. Therefore, we hypothesise that suddenly encountering a questionnaire while immersed in the game play will lower or even break presence.

As previously discussed, the collection of self-reported data is essential for research on games (and serious games in particular), for example to assess learner performance or current status. As presence has been linked with more concentrated execution of tasks and motivation [5], questionnaires that interrupt the game play, and thus break presence, may then distort subsequent measurements. The detrimental effect on presence should therefore be minimised for self-reported data acquisition.

Mapping Questionnaire Elements to Game Elements

As we intend to maintain presence, attention should first be paid to the factors that have been shown to contribute to presence. Witmer and Singer [30] point out *realism factors* and *control factors*, among others, as contributing to presence. These can easily be adopted for our purposes.

The *realism factors* imply that the information has to be consistent with the virtual world. A part of the *realism factors* is *scene realism*, which "refers to the connectedness and continuity of the stimuli being experienced". Accordingly, we suggest that the questionnaire has to be presented in a way that is consistent with the game's context in order to maintain presence.

Concerning *control factors*, Witmer and Singer argue that "[p]resence in a situation may be enhanced if the manner in which one interacts with the environment is a natural or well-practiced method for that environment". Therefore, we argue that the player should be familiar with the input method used for answering the questionnaire integrated within the game play. As there is usually no direct match to answering questionnaires in the game's context, the mapping will likely be on a metaphorical level.

In summary, the main idea of integrating questionnaires is to keep the player in the context of the game, by maintaining the game's metaphor while displaying the questionnaire. This forms the foundation of providing consistent information and interaction with the virtual world. Therefore, it is essential to map questionnaire elements to game elements. This implies that it is necessary to find equivalencies in the game's context for the presentation of a question and the method of answering it. For example, mapping a Likert scale to a game element requires the search for the game's equivalence of a scale item. Furthermore, it is crucial to implement the selection of scale items without changing the player's input scheme. Although in general our goal is to integrate all types of questionnaires in a fitting manner, in this paper, we focus on embedding Likert scales, as they are widely used and a mapping can be found quite easily.

Incorporating the Questionnaire Game Elements

Once the mapping from questionnaire elements to game elements has been found on a conceptual level, it is still necessary to determine when and how to present the questionnaire. While a lot of the "how" is already determined by the conceptual design, the "when" is basically independent thereof. The interruption of the game play may effectively be alleviated by employing further strategies.

Ideally, a questionnaire should be hidden in a way that the player does not perceive an incisive break of the game play, or does not even realise a break occurred at all. As this is a lofty goal, we instead propose an approach that softens the experienced break. The actual game play is effectively halted and the game guides the player through a transitional phase designed to gradually prepare the player for the break. Then the questionnaire is presented in the chosen design, while still upholding the characteristics of the game's virtual environment as naturally as possible. Afterwards, the game play is continued by transitioning back to the game play effectively softening a break, because the player still remains in a natural environment. The purpose of this manipulation is to make the interruption seem less sudden than it would otherwise appear. Thus, we expect the presence level to remain at a more balanced level compared with overlay questionnaires (see Figure 1).

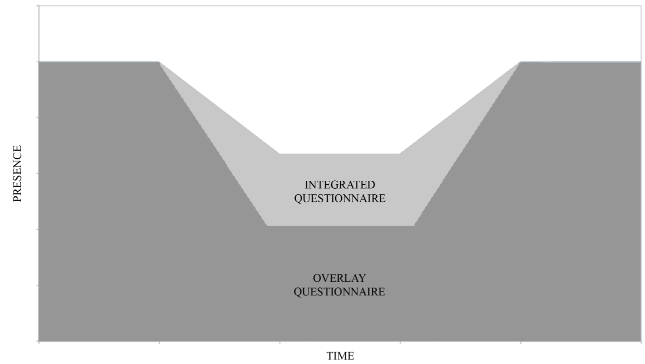


Figure 1: Expected development of presence for overlay questionnaires and integrated questionnaires.

To summarise, we expect that integrating questionnaires as game elements leads to less disturbance of the experience of presence when designed in a way that is consistent with the game's context.

CASE STUDY: LENGUADRIVE

We tested our hypotheses with the implementation of a prototypical serious game called *LenguaDrive*. LenguaDrive was developed with Unity, and employs a racing game paradigm to help players practise Spanish vocabulary recognition. The principle of racing along an endless track while having to collect items classifies the game as a high-speed action game that repeatedly poses small reactive challenges to the player. This is said to facilitate tactical immersion, as classified by Adams [1]. This is particularly suited to engage the player on an instinctive level; conscious strategic thinking is superseded by automated reflexive behaviour. This type of game mechanic is therefore quite appropriate to train a player's recollection of previously learned vocabulary.

Game Concept

The game design of our prototype provides a single car as the playing piece that races along a three-lane endless road (similar to rhythm game AudioSurf [9], which served as the inspiration for this approach). Three lanes were deemed optimal to provide a challenge for lane changes, while also still feasible for automated reactions at high speed. Periodically, the car displays a word in the player's native language (here: German), and a potential foreign translation (here: Spanish) appears a short distance ahead on each of the lanes (see Figure 2). The player has to move to the appropriate lane (via

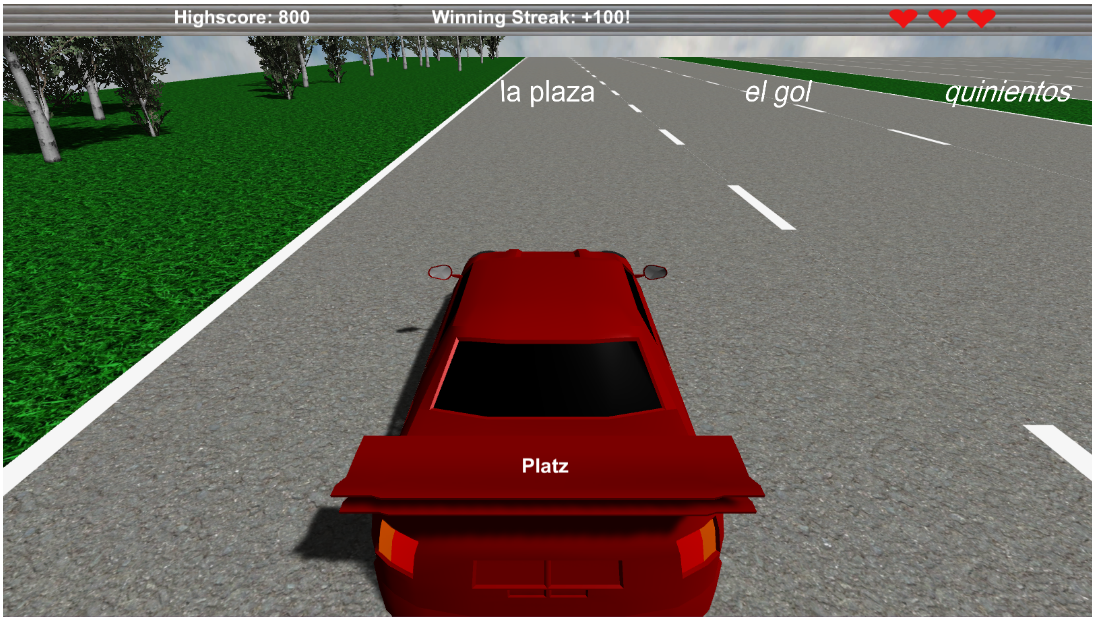


Figure 2: A LenguaDrive screenshot with a vocabulary prompt on the car, and three potential translations ahead.

arrow keys) in time to collect the correct translation. Each correct translation earns the player points, but also increases the speed, and thus, the challenge. A winning streak of 20 correct translations yields bonus points, and instead of a further increase, the speed is decreased by half. Incorrect translations also decrease the speed, however this also leads to the loss of one of the three initial life points. The game ends after the third life point has been lost. Conceptually, there is no victory condition; the game could technically continue indefinitely. However, even with continuously correct translations, the speed eventually increases beyond reasonable capacities for reaction, and thus leads to a loss of all three life points. The player's goal is to achieve a maximum high score before this occurs.

To be able to use LenguaDrive to test our hypotheses, we needed different versions of the game that varied in their display of questionnaires. The developed variants are described in the following paragraph.

Questionnaire Integration

At specific measuring points, LenguaDrive can pause the game play, and display a questionnaire with Likert scale items. The questionnaire presentation was embedded into the game in two different variants to support our subsequent comparative evaluation: a) an overlay questionnaire screen, and b) an *emotion road* game element.

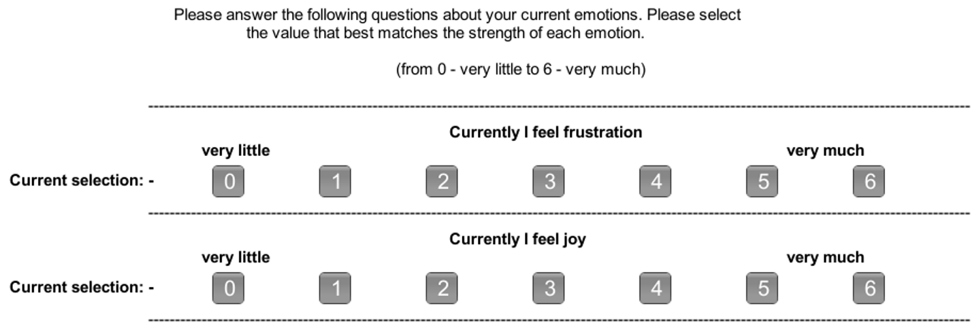
The overlay screen variant represents a digital version of the usual pen-and-paper questionnaire; the game play is paused in order to display the questions and resumed upon submission of the player's responses.

The *emotion road* variant transforms the questionnaire in line with the game metaphor; the questionnaire thus becomes a game element. At the time of the measuring point, the game takes control over the player's car, and drives it to a motorway exit leading to a second, parallel road representing the questionnaire. On this road, the vocabulary prompt of the main game play is replaced by a questionnaire prompt, and the response options to the questionnaire replace the potential translations.

The three-lane layout is changed to one corresponding to the Likert scale employed in the questionnaire, e.g. a seven-lane road where each lane displays an answer on the Likert scale. To support orientation, the values are also indicated as markings on the road's tarmac surface. The player is then given control over the car again, in order to select the appropriate lane. After driving through each of the questionnaire items, the player's car is guided back to the main road via motorway access, and the game continues. With this variant, the game concept and mechanics stay in place. The car's speed was drastically reduced for the *emotion road* in comparison with the regular game play in order to ensure that players had enough time to appropriately answer the questions. The speed of the car, and thus the time that players had to answer, was determined in a pre-study during the iterative game design process. Further, the players could manually accelerate the car to confirm the selection when they had chosen an answer. A comparison of the two questionnaire variants in the prototype can be seen in Figure 3.

Application of the Concept for Integrated Questionnaires

Regarding the realisation of the concept proposed in this paper, the presentation of the questionnaire in the *emotion road* variant is basically the same as compared to the regular game play (see Figures 2 and 3b). The player's avatar is still a car that drives along a road with multiple lanes. On these lanes the player encounters virtual objects that can be collected by driving through them, i.e. a translation for a given word in the



(a)

(b)

Figure 3: Integration of a Likert scale questionnaire (a) as an overlay screen and (b) as an *emotion road* game element; the design of the questionnaire is almost identical to the regular game play and thus resembles the natural game environment.

regular game play, versus an answer to a questionnaire item on the *emotion road*.

The player uses the same control method for the regular game play and the *emotion road*. The player changes between the lanes with the arrow keys. This corresponds to changing the current selection, while driving through an item corresponds to confirming the selection. The player should therefore immediately be able to select the answers to questions, despite the small change in the environment. In contrast, the player encounters a completely new presentational style with the overlay screen and also has to use a different control method that is not used in the regular game play, i.e. selection via mouse.

As mentioned, the interruption is alleviated by the integration of transition phases. This is realised in the form of motorway exits and access roads. By taking control over the player's car, the game provides a transition phase before another task (i.e. the questionnaire) has to be executed. Thereafter, the game guides the player back to the main road, to resume regular game play. In this manner, we hope to also re-assert the baseline level of presence to that associated with the regular game play.

EVALUATION

Our concept suggests a way to integrate assessment (e.g. of emotions) into a game's progression with minimal interruption to the experience of presence. In order to investigate the validity of our concept, we performed a comparative evaluation of the questionnaire integration in our prototype. Previous research has indicated that presence is correlated with positive emotions, and positive emotions, in turn, have been shown to benefit task execution and learning [5]. The in-game questionnaire was therefore used to assess players' emotions during game play, i.e. enjoyment and frustration. The aim of the study was to explore whether the players' experience of presence during game play differed depending on the type of assessment: with an explicit questionnaire screen, or with an implicit questionnaire that is presented as a game element. Our hypothesis consisted of the following:

The integration of the emotional assessment into the game play will coincide with a higher experience of presence than interrupting the game play with the overlay of a traditional questionnaire screen.

Participants

The sample of this study consists of sixty-one university students (40 female, 21 male) in Germany, who participated in a beginner course for learning the Spanish language. The students were recruited by contacting their Spanish teachers. The age range varied between 18 and 38 with the mean age equal to 23.20 years (SD = 3.12). The participants formed a homogenous group with respect to their previous knowledge of the Spanish language and regarding their frequency of playing games in their leisure time.

Measures

Immersive Tendencies

As one variable that might influence presence during game play, learners' tendency to invest in virtual presence was measured through Witmer and Singers' *immersive tendencies questionnaire* (ITQ) [30]. The ITQ consists of 23 items measured on a 7-point Likert rating scale ranging from 0 (strongly disagree) to 6 (strongly agree). The ITQ asks how frequently learners play games, their previous experiences of involvement in immersive activities, and their ability of selective attention. For our study we used a translated version of the ITQ (see [11]). The reliability of the overall questionnaire was .70 (Cronbach's alpha).

Presence

The participants' subjective experience of virtual presence was measured by a post-test questionnaire immediately after they had finished working with one of the two versions of the game. The questionnaire is based on Witmer and Singers' *Presence Questionnaire* (PQ) [30] that is often used for studies that attempt to determine the sense of presence in virtual environments (e.g. [11, 19]). This original questionnaire was tested in diverse studies that showed that the responses to the questions were relatively consistent [13]. Thus, the subjective evaluation of virtual presence seems to be an adequate means of assessment. We adopted the translated questionnaire of Heers ([11]), albeit with minor changes in order to refer to our game environment. It consists of 16 items to be answered on a 7-point Likert scale ranging from 0 (strongly disagree) to 6 (strongly agree). The items address the players' subjective feeling of being in control of the game, whether they perceived the interactions and movements as natural, and if they were involved so deeply that they could fade out external distractions that interfered with performance. The reliability of the virtual presence questionnaire was .94 (Cronbach's alpha).

Emotions

As previous works have indicated that presence increases positive emotions, two emotions were also measured, i.e. enjoyment and frustration. Responses on how much of each of the two emotions were experienced by the participants during game play were assessed on a 7-point Likert scale ranging from 0 (=very little) to 6 (=very much).

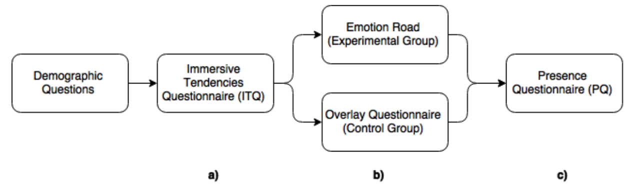


Figure 4: Relevant measures during the study: a) immersive tendencies, b) emotions and c) presence.

Procedure

The experiment was organised in a university computer lab. All participants were randomly assigned to one of the two game versions, with each version representing one of the manipulation conditions of assessment of emotions during game play in the experiment: the implicit assessment condition with the integrated questionnaire (n = 30), or the explicit assessment condition with the electronic overlay questionnaire (n = 31). At the start, all participants were introduced to the game and its rules by means of the presentation of a video tutorial. Then, participants completed the questionnaire with respect to their immersive tendencies (ITQ). Afterwards, participants were invited to play. During game play the participants worked on an individual basis on a computer. They could interact through the keyboard (for the game) and mouse (for the overlay questionnaire). Their emotional state during game play was measured once during game play, as soon as one of the following conditions was met: a) the player achieved five correct translations in a row, or b) the player lost their last life. After finishing the game, participants had to rate their experience of virtual presence during game play (PQ). It should be noted that the measurement of presence was a post-test after game play, while the two variants of the in-game questionnaire measured emotion (see Figure 4).

Results

First, it is noted that there was no difference between the two conditions according to gender (male = 10, female = 20 for the implicit assessment condition; male = 11, and female = 20 for the explicit assessment condition), or their immersive tendencies ($t(59) = 1.18$, $p = .24$). Hence, results are not likely to be differences in artefacts of gender or immersive tendencies between both conditions.

Differences in Emotions During Game Play

We investigated how participants in the different game conditions differed in their experienced enjoyment and frustration during game play. For both emotions, an independent sample t-test with enjoyment and frustration as dependent variables and the game condition as the independent variable showed that the assessment condition had a significant effect on participants' enjoyment ($t(59) = 3.91$, $p < .05$, $d = 1.00$). The means indicate that participants in the implicit assessment condition experienced a higher degree of enjoyment

(M = 3.87, SE = 1.54) compared to participants in the explicit assessment condition (M = 2.58, SE = .95). Regarding the negative emotion, participants in the implicit assessment condition experienced lower frustration on average (M = 1.87, SE = 1.52), than those in the explicit assessment condition (M = 2.32, SE = 1.97). However, this difference was non-significant ($t(59) = -1.01$, $p = .31$).

Differences in the Experience of Presence

Next, we investigated how participants in both game conditions differed in their experienced presence. An independent sample t-test with the virtual presence as the dependent variable and the game condition as the independent variable showed a significant effect on participants' experience of virtual presence ($t(59) = 3.96$, $p < .05$, $d = 1.27$). Descriptive results confirm the expected experience of higher virtual presence for the implicit assessment condition (M = 3.62, SE = 1.04) compared to the explicit assessment condition (M = 2.53, SE = .62).

The Relation Between Emotions and Presence

Correlation analyses were used to test the relationship between emotions during game play and the experience of virtual presence. The data shows that all variables were related in the expected directions. Enjoyment and frustration were both significantly correlated to virtual presence. As enjoyment increased, virtual presence also increased, $r = .53$, $p < .05$. The analysis revealed an inverse correlation for the relation between frustration and virtual presence, i.e. as frustration increased, virtual presence decreased, $r = -.35$, $p < .05$.

Summary

The results of our evaluation indicate that the core concept presented in this paper is well-founded. As hypothesised, the experience of presence was significantly higher for the game play with the integrated questionnaire than with the overlay questionnaire. The connection between positive emotions and presence was also substantiated.

DISCUSSION

There are limitations that arise either from the concept itself or from the implementation of the prototype. Nevertheless we argue that the concept is suitable to maintain presence for self-reported data acquisition, when applied in a fitting manner. Further, we propose possible factors contributing to the maintenance of presence despite questionnaire administration, based on earlier work on possible factors governing presence [30], and the results of the conducted user study.

Limitations

Limitations of the Concept

The biggest limitation of the concept is likely that integrating questionnaires as game elements within game environments effectively still interrupts the actual game play. The measurement of physiological data and the automatic analysis thereof will always prove less of a disruption to the sense of presence, as it can be performed unobtrusively during game play. Therefore, questionnaire integration is a method that should be used when other kinds of assessment prove inept; for instance, because emotion recognition has failed or supplied insufficient data, or subjective variables are being measured that cannot be assessed without self-reported data. The presented study offers a first validation of the concept by integrating questionnaires in one game concept. In general however, it can be quite difficult to find a required mapping from questionnaire elements to game elements for different game genres or game concepts. But as it is crucial that the player stays in the game's context and the metaphor is upheld, such mapping is required. Finding this mapping is a problem that may have to be solved individually for each game; this may be a significant limitation of the concept and has to be examined in future work. In some games, this might translate into wrapping the questionnaire in a dialogue with a non-player character (cf. the previously mentioned example in *Silent Hill: Shattered Memories* [20]); for many other games, such interaction would seem out of context.

As the questionnaire in the *emotion road* is presented within the game context, players might "play around" with the questionnaire. However, we neither expected players to nor observed them engaging in such exploratory behaviour, likely because they were familiar with the presentation and input scheme, and regular game play already limits exploration. In more open-ended games, however, this might be a limitation of the proposed concept.

Further, while the integration of short questionnaires can more easily maintain the experience of presence, this effect could potentially be reversed when the questionnaire's length reaches a size at which the user grows annoyed. This issue must be addressed in future work.

Limitations of the Prototype

Regarding the limitations of the prototype, it should be noted that the players' play time was longer for the *emotion road* condition than in the control group. While the time spent with regular game play was, on average, equal for both conditions, each of the items of the questionnaire was presented sequentially, while the items were displayed concurrently in the control group.

The implemented prototype was used to test the concept of integrating questionnaires, but also to verify whether a racing game paradigm can help players practise Spanish vocabulary recognition. However, it should be noted that the learning performance was not evaluated in this user study. Therefore, we cannot at this point confirm that this game concept is indeed suited to this use case. Lastly, it may be necessary to compare alternative designs of the overlay questionnaire. The difference in the experienced presence may decrease when the survey is designed in a more appealing way. Similarly, some details regarding the integration of the questionnaire could also be implemented differently, and consequently lead to a different effect on presence.

Contributing Factors to the Maintenance of Presence

From the comparison of the design of the *emotion road* and the overlay questionnaire, we conclude that two factors are particularly important for the integration of questionnaires: *control consistency* and *presentational consistency*.

The players in the *emotion road* group used the same control method to select an answer to the questionnaire as they used to control the car during regular game play, i.e. arrow keys to change the selection, and driving through the virtual object to confirm the selection. The players in the control group also used the arrow keys control method during game play sessions, but they answered the questionnaire by selecting an answer via mouse click. This control method was not used during game play, and even required the players to change the input device. We therefore suspect that *control consistency* is important to maintain presence, as players should instantly recognise how they can respond to the questionnaire prompt. The interaction should be realised in a manner that is already familiar to the player from the game play. Input with other devices than those used in the game, or alternative interaction mechanisms should therefore be avoided.

The second possible factor is *presentational consistency*, that is, the presentation of the questionnaire should be appropriate in the game's context. The questionnaire should not be displayed in a way that clashes with the game's presentation. The players in the integrated questionnaire condition always stayed in the same context, i.e. steering a car, while the user interface only gave them feedback about their status (compare Figures 2 and 3b). The players in the control group, on the other hand, encountered a user interface that matched neither the game's context nor its metaphor.

Considering Csikszentmihalyi's theory of flow [8], which has also been shown to be related to presence [14], *timing* appears to be an important factor. While it may seem obvious, it is necessary to keep in mind that the moment at which the questionnaire is displayed needs to be carefully selected. When the player is experiencing a state of flow, the sudden display of a questionnaire would appear jarring, and decrease or even break the flow. Additionally, the time spent answering the questionnaire as a whole might eventually extend beyond what can be considered a sustainable duration for flow, or presence.

As previously discussed, we based our concept on earlier work concerning factors that have been shown to govern the experience of presence [30]. With the results of our evaluation, we propose that the two factors of *control* and *presentational consistency* contribute to the maintenance of presence with regard to integrated questionnaires. These suggested contributing factors of course also inherit significantly from the work by Witmer and Signer on the original presence factors.

FUTURE WORK

While the case study provided evidence that the concept does work in general, it suffers from the mentioned limitations. Therefore, it is necessary to conduct future work that explores the concept's general applicability, and its performance in comparison with other forms of assessment and implementation. The factors contributing to the maintenance of presence need also be examined further.

General Applicability

While the results of the evaluation were promising, it is necessary to investigate whether the concept is also more generally applicable. We intend to examine in what ways the concept can be applied to other game concepts and questionnaire variants. This issue has to be considered from two perspectives. First, it is necessary to examine if a metaphorical mapping for Likert scales can be found for different genres. Second, we have to investigate if it is possible to map other kinds of user assessments onto game elements. For example, the integration of other kinds of user assessment (e.g. open questions) may not be feasible in the current prototype's game paradigm, but there may be a different genre well-suited to them. The presented case study focused on the integration of Likert scales; in future work we intend to integrate other questionnaire methods in a variety of game environments. Ideally, this could lead to guidelines regarding which game genres or scenarios are most appropriate for different questionnaire methods, and suggestions on how to map these to game elements.

Performance

The presented case study only compared two different types of assessment: self-reported data via an integrated questionnaire, as opposed to an overlay questionnaire. We intend to study how the employed concept performs in comparison with other forms of data collection, for example, measurement of variables with physiological sensors. This is of particular interest as it coincides with the potential trade-offs between self-reported subjective data, and unobtrusively detected but potentially inaccurate objective data.

Implementation

The strategy that is applied for alleviating the breaks in game play is based on transition phases, i.e. guiding the player's car towards the motorway exit. Whether this regulation does indeed work as intended remains uncertain, and has to be validated in further studies.

Furthermore, alternative strategies could be implemented to hide the break in game play. For example, if the timing and duration of questionnaires was adjusted to display only singular questions before game play is resumed, the short game play breaks might be less intrusive than one long break with multiple questions. This approach is not applicable for many questionnaires, as the split into single items would distribute the measurement over multiple points in time, due to the phases of game play between each question. Nevertheless we would like to explore if this strategy could be applied in some cases, and its effect on presence.

Contributing Factors

Finally, we intend to further study the factors involved in maintaining presence during the integration of questionnaires in game environments. The list of factors suggested in the discussion is by no means certain, or complete. Therefore, the factors identified in earlier work as contributing to presence in general (see [12, 23, 29, 30]) need further examination regarding their application to the integration of questionnaires.

Based on these factors, we plan to develop foundational design guidelines for the integration of questionnaires in game environments.

CONCLUSION

This paper presents a concept for maintaining presence in game environments by integrating questionnaires as game elements. The proposed design of the integration is based on factors shown to contribute towards presence in virtual environments [30]. We propose the mapping of questionnaire elements to game elements, while maintaining consistency regarding both presentation and input. The concept was investigated with a prototypical serious game that employs a racing game paradigm to help players practise Spanish vocabulary recognition. A user study was conducted with 61 participants comparing two variants of questionnaires during game play. The results show that participants experienced significantly higher presence in the variant in which the questionnaire was integrated as a game element, when compared to the variant realising the questionnaire as an overlay screen. Despite the discussed limitations of the prototype, the results of the evaluation are promising, and indicate that the concept is sound. Finally, we identified two potential contributing factors (*control* and *presentational consistency*) on which to base future work towards the development of design recommendations on the integration of questionnaires into games.

ACKNOWLEDGEMENTS

This work was conducted as part of the project "Serious Games – Skill Advancement Through Adaptive Systems", funded by the Carl Zeiss Foundation, as well as "EffIS – Efficient and Interactive Studying" (FKZ: 160H21032), funded by the German Federal Ministry of Education and Research (BMBF).

REFERENCES

1. Adams, E. *Fundamentals of game design, third edition*. New Riders, 2014.

2. Baos, R. M., Botella, C., Alcaiz, M., Liao, V., Guerrero, B., and Rey, B. Immersion and emotion: their impact on the sense of presence. *CyberPsychology & Behavior 7*, 6 (2004), 734–741.

3. Becker, K., and Parker, J. R. *The guide to computer simulations and games*. John Wiley & Sons, 2011.

4. Bente, G., and Breuer, J. Making the implicit explicit. *Serious games: Mechanisms and effects* (2009), 322–343.

5. Bertram, J. Virtuelles Training: Erfolgsfaktor Präsenz (german). *wissens.blitz*, 11 (2011).

6. Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Lee, S., Neumann, U., and Narayanan, S. Analysis of Emotion Recognition Using Facial Expressions, Speech and Multimodal Information. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, ICMI '04, ACM (New York, NY, USA, 2004), 205–211.

7. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine 18*, 1 (Jan. 2001), 32–80.

8. Csikszentmihalyi, M. Flow: The psychology of optimal performance. *NY: Cambridge UniversityPress* (1990).

9. Fitterer, D. Audiosurf: Ride your music, 2008. Accessed: March 2015. http://www.audio-surf.com.

10. Haag, A., Goronzy, S., Schaich, P., and Williams, J. Emotion Recognition Using Bio-sensors: First Steps towards an Automatic System. In *Affective Dialogue Systems*, E. André, L. Dybkjær, W. Minker, and P. Heisterkamp, Eds., no. 3068 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, 36–48.

11. Heers, R. *"Being There": Untersuchungen zum Wissenserwerb in virtuellen Umgebungen (German)*. PhD thesis, Universität Tübingen, 2005.

12. Held, R. Telepresence. *The Journal of the Acoustical Society of America 92*, 4 (Oct. 1992), 2458–2458.

13. Hendrix, C., and Barfield, W. Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators and virtual environments 5*, 3 (1996), 274–289.

14. Jin, S.-A. A. "I Feel Present. Therefore, I Experience Flow:" A Structural Equation Modeling Approach to Flow and Presence in Video Games. *Journal of Broadcasting & Electronic Media 55*, 1 (2011), 114–136.

15. Kalawsky, R. S. The validity of presence as a reliable human performance metric in immersive environments. In *3rd International Workshop on Presence, Delft, Netherlands.(http://www. presence-research. org)*, Citeseer (2000).

16. Lessiter, J., Freeman, J., Keogh, E., and Davidoff, J. A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence 10*, 3 (2001), 282–297.

17. Lewis, C., and Wardrip-Fruin, N. Mining Game Statistics from Web Services: A World of Warcraft Armory Case Study. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*, FDG '10, ACM (New York, NY, USA, 2010), 100–107.

18. Mania, K., and Chalmers, A. The Effects of Levels of Immersion on Memory and Presence in Virtual Environments: A Reality Centered Approach. *CyberPsychology & Behavior 4*, 2 (Apr. 2001), 247–264.

19. Moreno, R., and Mayer, R. E. Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology 94*, 3 (2002), 598–610.

20. Mountain, G. Psychology profiling in Silent Hill: Shattered Memories. In *Video Presented at the Paris Game/AI Conference* (2010).

21. Runge, J., Gao, P., Garcin, F., and Faltings, B. Churn prediction for high-value players in casual social games. In *2014 IEEE Conference on Computational Intelligence and Games (CIG)* (Aug. 2014), 1–8.

22. Schubert, T., Friedmann, F., and Regenbrecht, H. The experience of presence: Factor analytic insights. *Presence 10*, 3 (2001), 266–281.

23. Sheridan, T. B. Musings on telepresence and virtual presence. *Presence: Teleoperators and virtual environments 1*, 1 (1992), 120–126.

24. Shute, V. J. Stealth assessment in computer-based games to support learning. *Computer games and instruction 55*, 2 (2011), 503–524.

25. Slater, M., Usoh, M., and Steed, A. Depth of presence in virtual environments. *Presence 3*, 2 (1994), 130–144.

26. Um, E. R., Plass, J. L., Hayward, E. O., and Homer, B. D. Emotional design in multimedia learning. *Journal of Educational Psychology 104*, 2 (2012), 485–498.

27. Weibel, D., and Wissmath, B. Immersion in Computer Games: The Role of Spatial Presence and Flow. *Int. J. Comput. Games Technol. 2011* (Jan. 2011), 6:6–6:6.

28. Westera, W. Performance assessment in serious games: Compensating for the effects of randomness. *Education and Information Technologies* (Aug. 2014), 1–17.

29. Witmer, B., and Singer, M. Measuring immersion in virtual environments. *US Army Res. Inst., Alexandria, VA, Tech. Rep 1014* (1994).

30. Witmer, B. G., and Singer, M. J. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments 7*, 3 (June 1998), 225–240.