




InApp Questions – an Approach for Contextual Evaluation of Applications

Jorge Ferraz de Abreu^(✉) , Pedro Almeida , and Pedro Beça 

University of Aveiro DigiMedia, Aveiro, Portugal
{jfa, almeida, pedrobeca}@ua.pt

Abstract. The growing success of second-screen devices is changing the way users relate and interact with the television bringing a new dimension to the TV ecosystem. Nevertheless, the specificities of mobile devices are creating numerous and significant challenges in the field of usability and user experience (UX). In addition, the evaluation of applications of the TV ecosystem should not be restricted to the measurement of its instrumental qualities typically belonging to the usability dimension, since it is also important to consider dimensions related to the UX. This paper focuses on the evaluation of a second screen application - the TV content discovery App GUIDER. Two data collection strategies for the evaluation of the App were used: a methodology based on InApp questions (that allows to measure momentary UX) and a more conventional evaluation based on an online questionnaire answered at the end of evaluation sessions (episodic UX). The use of both data collection strategies was intended to get comparative insights to validate if the InApp questions approach could get similar results when compared with those from traditional surveys. The results showed that the response deviation was very low reinforcing the suitability of such alternative data gathering method. The results indicate that triggering InApp questions while users are interacting is a promising method for evaluating applications, namely for UX dimensions, either for functional prototypes or final products. Nevertheless, the validation of such hypothesis requires further and dedicated research.

Keywords: User experience · Usability · Evaluation methodologies · Interactive television

1 Introduction

The specificities of mobile devices created numerous and significant challenges in the field of usability and user experience (UX). Mobile context, multimodality, connectivity, small screen size, different display resolutions and power emerged as factors to be taken into consideration when designing interfaces for mobile devices [1]. Interacting with these devices implies a different look at usability and UX. User's hand is no longer over a mouse but directly interacting with the interface through multi-touch gestures such as swipe, pinch, press and hold [2, 3], creating the need for bigger buttons (to solve the “fat-finger” problem), wider distance between icons and new navigation paradigms.

In addition, mobile usability models often focus on the effectiveness, efficiency and satisfaction, disregarding the cognitive load and the emotional impact of interacting with applications in no-longer defined time and place. However, when developing second-screen applications designed to enhance the TV viewing experience, it is even more relevant to take into account the cognitive capacity and attention selectiveness as well as other UX dimensions.

Following the work on TV Discovery & Enjoy (TDE) project that lead to the development of an IPTV application to assist the user in finding the content to watch on his/her TV set [4], this paper addresses the evaluation of a second screen approach, a tablet application as an alternative to traditional EPGs, to support the discovery of content and to assess how second screens, interconnected with an iTV platform, can improve the residential TV ecosystem. The application, implemented for iOS, was branded GUIDER and allows users to discover TV content from unified sources.

While addressing the evaluation of the App, the paper main focus is on the comparative results of the two adopted strategies for data collection: a methodology based on InApp questions triggered upon users' interaction and a more conventional evaluation based on an online questionnaire delivered to participants after a lab session evaluation. Taking in consideration the Roto [5] model of time spans of User Experience, one can say that the InApp questions give us an interesting opportunity to measure the momentary UX, while final questionnaire evaluation allows to measure the episodic UX.

Taking this main objective in consideration, the paper is structured as follows. The next section presents the state of the art addressing the second-screen applications domain and its integration in the TV ecosystem; the specificities of the evaluation processes of applications belonging to the current TV ecosystem, namely when UX dimensions are at stake; and the need for new and dynamic data gathering mechanisms. In order to contextualize the main issue of the paper, Sect. 3 presents the GUIDER Application, opening the floor to Sect. 4 ("Evaluation procedure") where the two adopted strategies for data collection are described. Section 5 presents a comparison of the results based on the average results gathered from the two mechanisms highlighting that the deviation is rather small. Finally, the paper concludes by reporting on the possible advantages of the InApp data collection method.

2 State of the Art

2.1 Second-Screen Applications

The growing success of second-screen devices is changing the way users relate and interact with the television bringing a new dimension to the TV ecosystem. Using second-screen devices while watching TV is an increasingly common activity: according to recent numbers of Nielsen Company [6], 62% of North Americans and 44% of European consumers use second-screen devices while watching TV. In this sense, consumers are increasingly adopting a lean forward approach to the television experience, using connected devices as extensions of the program they are watching [6].

Secondary Screen applications are, according to the description of Red Bee Media [7], those that provide a companion experience, aiming to increase and synchronously improve the viewer experience with content related to what is being displayed in the TV.

When considering TV recommendation systems (with a scope in line with the goals of this research), many research projects have been developed over the past few years. In 2007 the “AIMED” system [8], which uses as indicators for recommendations information based in the activities, interests, mood, experience and demographics, was created. It is a hybrid recommendation system as it combines two different techniques: content-based recommendation (based on the user’s television footprint) and collaborative recommendation (based on the tastes of users with similar profile). In addition, the AIMED system stores the viewing behaviours of each user (e.g. channel, time and length of viewing session), comparing it with the different moods of users (manually inputted via a remote control which includes 3 coloured buttons: “Happy”, “Bored” and “Unhappy”) [8]. The “Sybil Recommender System”, an experimental prototype for the web, funded by BBC Research & Development, introduces a model for the evaluation and recommendation of TV programs. The system displays a set of recommendations that can be filtered by gender (e.g. children, comedy or drama). For that, users drag and drop the recommendations to “Like” or “Dislike” boxes. Each time a content is drawn into one of the boxes, the list of is updated with new content, increasingly closer to the tastes of the user [9]. The Fraunhofer FOKUS has also introduced an application for the recommendation of TV content called “TV Predictor”. This system analyses the viewing habits and the evaluations of users to determine, among other things, channels, genres, directors or favourite actors, as well as preferred viewing times [10].

2.2 TV Apps Evaluation Process

The evaluation processes of applications belonging to the current TV ecosystem are shaped by its specific characteristics, which derive from several contextual factors (spatial, temporal, social and personal) as well as technological [11]. In addition to the challenges brought by these factors, the evaluation of this kind of applications (namely when UX dimensions are at stake) presents an additional complexity due to the multiple devices present in the living room: the TV set (often used as the main display for the most varied audio-visual contents); the common Set-Top Box (STB); the likely sound system; the remote controls; and ultimately the omnipresent mobile devices (which often play the role of secondary screens or even primary screens - depending on the generation who uses it).

There is also another factor that adds to this already complex equation - the purpose (based on the uses and gratifications theory [12]) with which the user uses the TV ecosystem: informational clarification; strengthening of personal identity; integration and social interaction; or simple entertainment.

2.3 The User Experience

According to what was previously mentioned, the evaluation of applications of the TV ecosystem should not be restricted to the measurement of its instrumental qualities (such as control, effectiveness and ease of learning) typical belonging to the usability dimension, since it is also important to consider dimensions related to the UX. This issue is of decisive importance for the TV industry and operators, since the overall UX is one of the central aspects that influence the choice of customers by one or another pay-TV solution.

UX Definition Currently, it is not possible to find a unique understanding of the concept of UX and, consequently, of the methods that can be applied in the evaluation of applications belonging to the TV ecosystem. The difficulty in finding a single, global definition of UX derives from its self-nature being associated with different concepts and meanings:

“User Experience (UX) is a catchy as well as tricky research topic, given its broad applications in a diversity of interactive systems and its deep root in various conceptual frameworks, for instance, psychological theories of emotion.” Law et al. [13, p. 1].

There is an ISO (9241-110:2010 - clause 2.15) definition for UX: “A person’s perceptions and responses that result from the use and/or anticipated use of a product, system or service” [14], but even this approach allows for distinct interpretations, namely others less formal like: “User experience explores how a person feels about using a product, i.e., the experiential, affective, meaningful and valuable aspects of product use” [15].

In this framework, one can observe that the interest on the UX topic by the industry and the scientific community is reflected in a profusion of methods, for the conception and global appreciation of the UX, whose systematization and organization in application domains has been attempted by some authors [16, 17].

UX Dimensions Taken in consideration the reported (in)definition, a possible approach to better analyse the UX concept is to consider its various dimensions (namely the ones more relevant for the applications of the TV ecosystem) as proposed by Bernhaupt and Pirker [18]:

- Stimulation – which “describes to what extent a product can support the human need for innovative and interesting functions, interactions and contents.”
- Identification – that “indicates to what extent a certain product allows the user to identify with it.”
- Emotional – which includes the feelings and emotions brought about during the course of the experience, such as satisfaction, motivation and control;
- Visual/aesthetics – which is related to the pleasure and attractiveness translated by sensory perceptions.

2.4 A UX Evaluation Approach for Second-Screen Applications

Despite the existence of methods valid to specific application domains, the authors of this paper noticed a considerable lack of free access contributions specifically adapted to

the evaluation of applications of the TV ecosystem [18] [19], allowing to quantify variations of UX between the original version and the corrected version of a same prototype/product. Taking this in consideration and in order to contribute with an open methodology, the authors present in [20] a UX evaluation approach for second-screen applications (also adaptable to other applications of the TV ecosystem). This approach was specified to assess users' perspectives on the instrumental and non-instrumental qualities of the application, as well as the emotional reactions triggered by episodic UX. The proposed methodology is based on validated questionnaires articulated as follows.

In order to evaluate the instrumental qualities of the application (which is useful for analysing how the perception of usability of the application relates to its UX), the SUS - System Usability Scale [21] and the Pragmatic Dimension of the questionnaire AttrakDiff ("PQ") [22] are used.

In order to evaluate the non-instrumental qualities of the application (stimulation and identification), the AttrakDiff Hedonic Quality component is used (sub dimensions "HQ-S" and "HQ-I").

To evaluate the emotional reactions (satisfaction, motivation and control) the approach make use of the SAM questionnaire [9] and the attractiveness value obtained from Attrakdiff ("ATT") that indicates a global value of the product based on the quality perception.

All these questionnaires were designed to be applied immediately at the end of an experience cycle (guided or free) in the following order: SAM, SUS and, finally, Attrakdiff.

2.5 Dynamic Data Gathering Processes

In the development of software applications, the User Interface (UI) is a fundamental element. To ensure that the UI is easy to use and meets the users' expectations usability and UX evaluations during and even after de development cycle are needed [23].

In this context, laboratory usability evaluation tests are the most common [23] method. However, despite being useful for the UI design phase, this type of evaluation procedures loose its relevance when trying to measure the satisfaction in using a product [24] since the use of such technique is unlikely to reproduce the natural usage environment or context.

On the other end, field trials that allow researchers to test their prototypes or products in real conditions appear as an ideal alternative. However, these methods are very demanding in what relates to the needed resources for its implementation and techniques for data gathering operations. It is however important to notice, that the use of different evaluation techniques allows to identify different problems. For example, the use of questionnaires to report critical incidents or other relevant information allows the users to provide different feedback about the product or application being evaluated [25].

One of the most relevant aspects in the usability of mobile devices is the ubiquitous context in which they are used. Since these devices are idealized to be used in movement, some of their characteristics, such as small screen sizes, limited connectivity, high power consumption rates and limited input modalities, may condition their use in different contexts. Another factor that has great relevance in usability and UX is the cognitive overload, mainly due to multitasking use and device size limitations [26].

To overcome these issues, it is possible to make use of automated usability evaluation processes, as for example in-app user feedback tools that allows gathering users' opinions in a contextualized way straight from the App being evaluated. This is by far a much more promising method than gathering opinions from external surveys or even reviews carried at App stores. A recent research shows that the users' app reviews allow analysts and app designers to know about the users' requirements, to detect bugs, get requests for new features, and the overall UX with a specific app [27]. However, the analysis of such (qualitative) information (user reviews) is highly demanding, and may suffer from bias resulting from the diversity of the quality and relevance of reviews. On the other hand, the utility of star ratings (quantitative and therefore easy to process data) is very limited [27]. Another limitation of this information, is that it is generated after the use of the App, not allowing to gather the UX right while using it. Thus, in mobile usability studies the hands-on measurements are one of the most used evaluation methodologies [28]. The use of these methodologies makes it possible to test the applications in real-life contexts without the need for observer/researcher intervention (as is typical of the final questionnaire evaluations), allowing to segment the evaluation by very specific functionalities, synchronizing the "question/answer" with the moment of its use.

3 The GUIDER Application

The current television ecosystem has been going through very significant changes, one being the huge increase in the existing TV content, either live or on demand, which is available to users with access to pay-TV solutions. In several countries, this offer has been increased because, in addition to over a hundred of linear-TV channels and thousands of contents available on the Video On Demand (VoD) service provided by most of the TV operators, they also offer Catch-up TV services (based in automatic cloud recordings) of the TV programs aired in the last 7 days. In this context, the viewer has constant access to a huge TV offer that may exceed 20,000 different TV programs.

Considering that many times when viewers sit in front of the TV they don't know in advance what to see, this huge offer of content means that they may feel "lost" with such offering. This may lead to a typical "mindless zapping" behaviour to find something matching their preferences; possibly missing a specific TV content that would be of their real interest. It is precisely to assist the user in this context that the authors had been focused on developing technological solutions to assist the user in discovering the TV content most suitable for his/her actual situation.

As referred, the first steps of development of the GUIDER iOS App started in the TDE research project [29], whose main objectives were to understand and identify the cognitive processes associated with deciding what to watch and, based on this study, to develop a prototype of an IPTV application that supports viewers in the search and selection of a TV program to watch. At the time, the research team also wanted to identify implementation strategies and opportunities for having a similar approach based on TV companion devices. Therefore, a medium fidelity prototype was developed, being the genesis for the development of the fully functional GUIDER application now presented.

3.1 System Architecture

Considering the system architecture, the GUIDER App is supported by a client and server approach. The client side includes, besides an iOS device (typically an iPad), a set-top-box (for a fully functional usage, namely allowing to start playing the program that was discovered/selected in the GUIDER app on the TV set).

The application connects to the home network via Wi-Fi and finds existing set-top boxes on the same network to have access to. If the App is used in a different TV operator from the one it was developed to, this control of the set-top-box is not available (it is worth to say that to maintain the same evaluation scenario between Brazil and Portugal this feature was disabled). At server side, the SearchEngine is responsible for the classification of TV contents and the search and discover features. In parallel, a module ©Parse is used to store all the information related with the use of the application, namely the answers to in App evaluation forms (later described).

3.2 Main Features

GUIDER includes some social features (namely profiling and social networks interaction) that require users to login in the app, which can be made through Facebook or Twitter. To achieve a complete integration with the TV ecosystem, the user has the ability to connect, if he/she is a client from the partner operator, to his/her set-top box. This association allows starting TV programs from the App. The main GUIDER area is the discovery screen (Fig. 1).

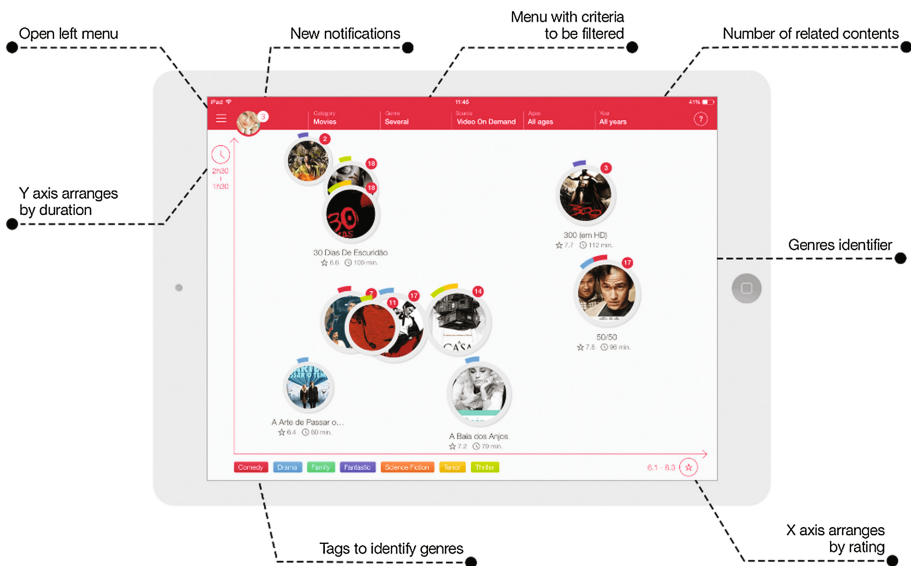


Fig. 1. The main screen

This area is structured as a multidimensional graphic presenting TV programs according to the criteria defined by the user. Programs are represented as circles and spatially distributed according to its classification (IMDB) (x axis) and duration (y axis). The user can easily change these values. The size of the circles varies accordingly to the popularity of the programs, which is determined by its reputation in the Catch-up TV of the operator (when possible) and the VOD viewers' evaluations.

According to what could be found on the IPTV application of the TDE project users may define other filters like program categories; content source (e.g. TV, VOD); age rating, and; starring decade.

The details panel about a program consists of general information of the program, such as genre, year, length and a brief synopsis and score (the IMDB rating and GUIDER users rating). Through this panel, users can also add programs to their favourites, watch a trailer, share or recommend it to friends in social networks, evaluate the program and, if in a compatible TV ecosystem, play it in the TV set. For a full list of features check [29].

4 Evaluation Procedure

Although the GUIDER App was designed to be used by clients of a Portuguese IPTV provider the research team decided to open the scope of the evaluation to try to predict if it had the potential to be adopted in other markets, namely by Brazilian users. To accomplish this, it was decided to evaluate the application in both countries with a sample of 20 participants in Brazil (herein referred as BR) and other 20 in Portugal (herein referred as PT). The number of evaluators in each country was decided in order to make the best usage of the Attrackdiff tool [30] that complemented the evaluation of the UX. Along with UX objectives, the evaluation goals were related to: determining the level of interest in the several features of the application, filtering criteria and future features; identifying usability issues; and predicting the future uses of GUIDER in domestic scenarios.

The research team chose to carry an evaluation in both countries following the same methodology, which allowed assessing the application features and particularities. It was also envisioned to compare the two adopted strategies for data collection regarding the usage experience of application (explained below).

The lab sessions took place at three universities (in Brazil at UFPE and UFPB and in Portugal at UA) in rooms prepared to recreate a friendly atmosphere, in some way similar to what one can find in a living room with some sofas and a TV. Participants of both groups were characterized in several dimensions by a first online questionnaire (Q1). In most of the cases they answered this questionnaire some days before the evaluation session. Just before the evaluation session, a brief overview of the application was made based on a promotional video (available at <http://tinyurl.com/ns5dbdp>). Then they were asked to freely explore and use the application, which took an average of 20 min per session.

In order to ensure a complete data collection during the evaluation sessions (and during a future regular use of the application) to allow future improvement of GUIDER, the team chose to use two complementary data gathering methods:

QInApp (InApp questions): The first method consists of an internal rating system, based on closed queries that appear contextualized while users interact with the app. These InApp questions are triggered when a specific GUIDER feature is used a certain number of times – being the related threshold (that was configured in an associated platform - at the time the ©Parse platform was used) dependent of the probability of use of each functionality at stake (i.e. for the features with a high probability of use the InApp question appeared after a higher number of interactions than the ones needed to trigger an InApp question for a feature with an expected lower probability of use).

For each functionality two questions are triggered (see Fig. 2), one concerning the level of interest in that specific feature and another about usability related issues. Participants were not obliged to answer the InApp questions.

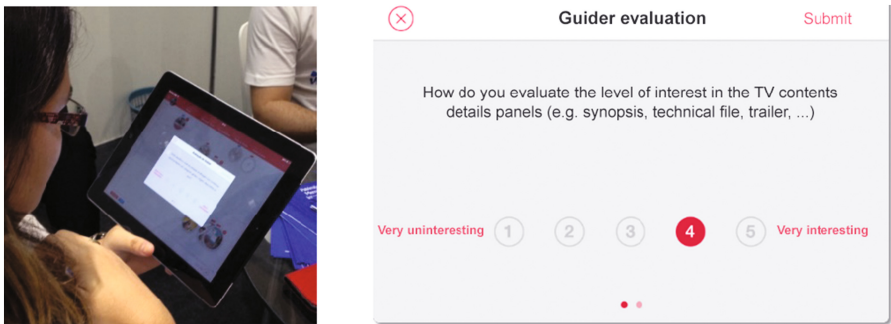


Fig. 2. A user getting an InApp question (left) and the details of the InApp interface.

F.Q. (Final questionnaire): The second data collection method involved an online final questionnaire (F.Q.) that participants answered at the end of the evaluation sessions. Questions of the Attrakdiff tool were included as an optional final section of the questionnaire. All participants were able to answer, allowing the research team to subsequently enter these answers on the online platform of the Attrakdiff tool (Fig. 3).

2.2 Please point out your level of interest in each of the following features that you have experienced: *

Rate from 1 (very uninteresting) to 5 (very interesting)

	1	2	3	4	5
TV contents detailed information (e.g. synopsis, technical file, trailer, etc.)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fig. 3. A question table from the final questionnaire (translated from the original Portuguese version).

5 Results from the InApp Questions Versus Final Questionnaire

Considering the main objective of the paper the results from both data collection mechanisms were treated and the answers (both from Likert scales) were processed.

Despite the difference in the data gathering process, it was attempted that the data collected in QInApp and F.Q. were similar. As the questionnaires were applied in different mediums, the questions that compose them were also different and the F.Q. included other questions not addressed in InApp questions. So, it was necessary to create a matching table between the questions of QInApp and of the F.Q. in 7 different blocks.

After doing this match, the results gathered in the questions from both methods were then analyzed. It is important to refer that both type of questions addressed a five-point Likert scale. Table 1 synthetizes the answers with the presentation of the average results from the F.Q. and the QInApp. The results from the Likert scale are normalized in a five-point scale (1 to 5) being 5 the most positive answer (e.g. “The most interesting”, “Very easy”). The table also includes the deviation percentage between both results (%AVEDEV).

Table 1. The correlation between the questionnaire questions and questions in the app

QUESTIONS		BR	PT	ALL
Question Block 1	Avg. F.Q.	3,65	3,50	3,58
	Avg. Qinapp	3,65	3,50	3,60
	% AVEDEV	0,00%	0,00%	0,35%
Question Block 2	Avg. F.Q.	4,55	4,45	4,50
	Avg. Qinapp	4,40	4,00	4,22
	% AVEDEV	1,68%	5,33%	3,26%
Question Block 3	Avg. F.Q.	4,45	3,90	4,18
	Avg. Qinapp	5,00	4,67	4,90
	% AVEDEV	5,82%	8,95%	7,99%
Question Block 4	Avg. F.Q.	4,35	4,03	4,15
	Avg. Qinapp	4,60	4,75	4,19
	% AVEDEV	2,79%	8,26%	5,41%
Question Block 5	Avg. F.Q.	4,80	4,35	4,58
	Avg. Qinapp	4,73	4,33	4,59
	% AVEDEV	0,76%	0,19%	0,14%
Question Block 6	Avg. F.Q.	4,31	3,95	4,13
	Avg. Qinapp	4,11	4,40	4,21
	% AVEDEV	2,36%	5,39%	1,01%
Question Block 7	Avg. F.Q.	4,03	4,05	4,04
	Avg. Qinapp	4,20	4,33	4,25
	% AVEDEV	2,13%	3,38%	2,56%

As can be noticed, the results show a close correlation between the answers on both instruments. Only Question Block 3 (7,99%) and 4 (5,41%) got an overall percentage of deviation higher than 5%. Question Block 3 asked users about their willingness to be able to share program recommendations on social networks. It is interesting to notice that users graded that question with very high responses in the InApp questions (e.g. 5 in 5 from the BR users) but with lower results on the F.Q. One possible reason could be that this feature refers to a reaction behavior. An action that typically is driven by immediate willing to share. This may reinforce the relevance of getting the answers right after using such a feature, and shows that the momentary UX evaluation can have different results from the episodic UX evaluation. Regarding Question Block 4, that is related with the level of interest to rate the audiovisual content (providing a score and a potential target audience), the reason for the inherent deviation is not clear. It is mainly verified within PT participants (8,26%) - the deviation on BR users was only 2,79%, but despite the potential differences between momentary and episodic UX no clear reason for that could be perceived.

It is important to notice that although the deviations are globally very small (see Fig. 4), the higher numbers are associated with the Portuguese participants, probably due to a lower number of InApp questions answered (as referred, participants had the chance to skip these questions).

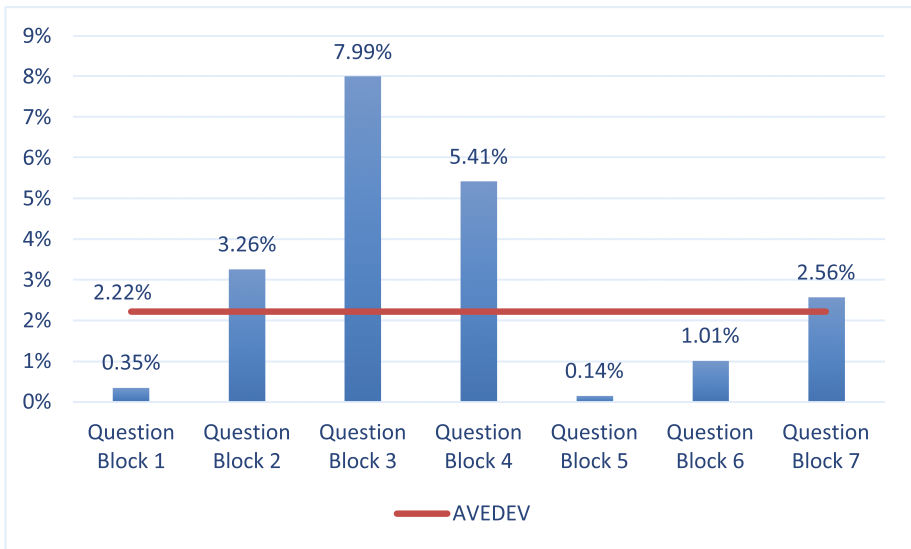


Fig. 4. The average deviations between both type of data collection instruments.

6 Conclusions

The reported evaluation allowed to get complementary information useful to improve the GUIDER application. With the use of both data collection strategies it was also expected to get comparative insights to validate if the InApp questions approach could get similar results when compared with those from traditional surveys. The results showed that the response deviation was very low reinforcing the suitability of such alternative data gathering method. The results from one of the questions can even let us perceive a possible higher accuracy in certain InApp questions, due to its immediate time correlation with the specific feature in analysis. The validation of such hypothesis requires further and dedicated research.

From these results, it was also possible to highlight the particular advantages of the InApp questions as a suitable method for evaluating applications, namely for UX dimensions, either for functional prototypes or final products, because they can be triggered at any time, based or accordingly to the users' activity, without the need for the researchers to be around and allowing to get the results in real time.

Triggering InApp questions also allows to evaluate the addition of new features to already established products, making it easier and faster for companies to decide if such features were appreciated or should be removed. Such methodology will be used in future UX evaluations of other TV related Apps to further confirm the suitability of such approach.

References

1. Zhang, D., Adipat, B.: Challenges, methodologies, and issues in the usability testing of mobile applications. *Int. J. Hum. Comput. Interact.* **18**, 293–308 (2005)
2. Treder, M., Pachucki, A., Zielonko, A., Lukasiewicz, K.: *Mobile Book of Trends* (2014)
3. Bank, C., Zuberi, W.: *Mobile UI Design Patterns* (2014)
4. Abreu, J., Almeida, P., Teles, B.: TV discovery & enjoy: a new approach to help users finding the right TV program to watch. In: *Proceedings of the 2014 ACM International Conference on Interactive Experiences for TV and Online Video*, pp. 63–70 (2014)
5. Roto, V., Law, E., Vermeeren, A., Hoonhout, J.: User experience white paper. In: *Bringing Clarity to Concept User Exp.*, pp. 1–12 (2011)
6. T.N. Company: Screen wars: the battle for eye space in a TV-everywhere world (2015). <http://www.nielsen.com/us/en/insights/reports/2015/screen-wars-the-battle-for-eye-space-in-a-tv-everywhere-world.html>, Accessed: 13 June 2015
7. R. B. Media, *Second Screen Series - Paper 1: Setting The Scene*, vol. 1 (2012)
8. Hsu, S.H., Wen, M.-H., Lin, H.-C., Lee, C.-C., Lee, C.-H.: AIMED-A personalized TV recommendation system. In: *European Conference on Interactive Television*, pp. 166–174 (2007)
9. BBC Research & Development, “Sibyl Recommender System” (2012). <http://sibyl.prototyping.bbc.co.uk/>. Accessed: 26 Dec 2013
10. Krauss, C., George, L., Arbanowski, S.: New Systems Programming. TV predictor: personalized program recommendations to be displayed on SmartTVs. In: *Proceedings of the 2nd Int. Work. Big Data Streams Heterog. Source Min. Algorithms Appl. BigMine 2013*. Vol. 23, pp. 63–70. ACM, New York (2013)

11. Obrist, M., Bernhaupt, R., Tscheligi, M.: Interactive television for the home: An ethnographic study on users' requirements and experiences. *Int. J. HCI* **24**(2) (2008)
12. Gauntlett, D., Hill, A.: *TV living: Television, culture and everyday life*. Routledge (2002)
13. Law, E.: The measurability and predictability of user experience. In: *Proceedings of EICS*, pp. 1–10. ACM, New York (2011)
14. I.S.O. (ISO), *Ergonomics of Human–System Interaction – Part 210: Human-centred Design for Interactive Systems (formerly known as 13407)* (2010)
15. User Experience. https://en.wikipedia.org/wiki/User_experience. Accessed 13 Jan 2017
16. Vermeeren, A.P.O.S., Law, E.L.-C., Roto, V., Obrist, M., Hoonhout, J., Vananen-Vainio-Mattila, K.: User experience evaluation methods: current state and development needs. In: *Proceedings of the 6th Nord. Conf. Human Computer Interact. Extending Boundaries*, pp. 521–530. ACM (2010)
17. All about UX (2017). <http://www.allaboutux.org/>. Accessed 13 Jan 2017
18. Bernhaupt, R., Pirker, M.: Evaluating user experience for interactive television: towards the development of a domain-specific user experience questionnaire. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) *INTERACT 2013*. LNCS, vol. 8118, pp. 642–659. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40480-1_45
19. Drouet, D., Bernhaupt, R.: User experience evaluation methods: lessons learned from an interactive TV case-study. In: Bogdan, C., Gulliksen, J., Sauer, S., Forbrig, P., Winckler, M., Johnson, C., Palanque, P., Bernhaupt, R., Kis, F. (eds.) *HCSE/HESSD -2016*. LNCS, vol. 9856, pp. 351–358. Springer, Cham (2016). doi:10.1007/978-3-319-44902-9_22
20. Abreu, J., Almeida, P., Silva, T.: A UX evaluation approach for second-screen applications. In: Abásolo, M.J., Perales, Francisco J., Bibiloni, A. (eds.) *jAUTI/CTVVDI -2015*. CCIS, vol. 605, pp. 105–120. Springer, Cham (2016). doi:10.1007/978-3-319-38907-3_9
21. Brooke, J., et al.: SUS-A quick and dirty usability scale. *Usability Eval. Ind.* **189**(194), 4–7 (1996)
22. Attrakdiff (2017). <http://www.attrakdiff.de/>, Accessed 13 Jan 2017
23. Jenny Preece, H.S., Rogers, Y.: *Interaction Design: Beyond Human-Computer Interaction* (2015)
24. Benedek, J., Miner, T.: Measuring Desirability: New methods for evaluating desirability in a usability lab setting. *Proc. Usability Prof. Assoc.* **2003**(8–12), 57 (2002)
25. Burzacca, P., Paternò, F.: Remote usability evaluation of mobile web applications. In: Kurosu, M. (ed.) *HCI 2013*. LNCS, vol. 8004, pp. 241–248. Springer, Heidelberg (2013). doi:10.1007/978-3-642-39232-0_27
26. Harrison, R., Flood, D., Duce, D.: Usability of mobile applications: literature review and rationale for a new usability model. *J. Interact. Sci.* **1**(1), 1 (2013)
27. Pagano, D., Maalej, W.: User feedback in the appstore: an empirical study. In: *2013 21st IEEE International Requirements Engineering Conference (RE)*, pp. 125–134 (2013)
28. Nayebi, F., Desharnais, J.-M., Abran, A.: The state of the art of mobile application usability evaluation. *2012 25th IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 1–4, May 2012
29. Abreu, J., Almeida, P., Silva, T., Oliveira, R.: Discovering in a second screen app: perspectives from portuguese and brazilian markets. *Procedia Comput. Sci.* **64**, 1240–1247 (2015)
30. Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products. *Hum.-Comput. Interact.* **19**(4), 319–349 (2004)