EvscApp: Evaluating the Pedagogical Relevance of Educational Escape Games for Computer Science

Rudy Kabimbi Ngoy, Gonzague Yernaux[a] and Wim Vanhoof[b]
Faculty of Computer Science, University of Namur, Namur, Belgium

Keywords: Educational Escape Games, Evaluation Framework, Computer Science Learning, Tool Presentation.

Abstract: While there is consensus that educational escape games have a beneficial impact on student learning in computer science, this hypothesis is not empirically demonstrated because the evaluation methods used by researchers in the field are carried out in an ad hoc manner, lack reproducibility and often rely on confidential samples. We introduce EvscApp, a standard methodology for evaluating educational escape games intended for the learning of computer science at the undergraduate level. Based on a state of the art in the realm of educational escape games and on the different associated pedagogical approaches existing in the literature, we arrive at a general-purpose experimental process divided in fifteen steps. The evaluation criteria used for assessing an escape game's efficiency concern the aspects of motivation, user experience and learning. The EvscApp methodology has been implemented as an open source Web dashboard that helps researchers to carry out structured experimentations of educational escape games designed to teach computer science. The tool allows designers of educational computer escape games to escape the ad hoc construction of evaluation methods while gaining in methodological rigor and comparability. All the results collected through the experiments carried out with EvscApp are scheduled to be compiled in order to be able to rule empirically as to the pedagogical effectiveness of pedagogical escape games for computer science in general. A few preliminary experiments indicate positive early results of the method.

1 INTRODUCTION

In 2001, it was said that there was a lack of more than 800,000 qualified IT workers around the world (Pawlowski and Datta, 2001). This phenomenon seems to have expanded since then. Indeed, more recently, in 2022, more than one news report exposed that an estimated lack of tens of millions of tech workers was to be expected by 2030 (Armstrong, 2022). This includes computer scientists as well as technicians. Both categories are going through a period of extreme lack of talent that is partially explained by the global digitalization process, substantially accelerated by the COVID-19 pandemic (Coquard, 2021).

Another factor contributing to this shortage of IT professionals is the general public's apprehension towards the difficulty and inaccessibility associated with computer science, which constitutes an important obstacle to its popularity (Marín et al., 2018).

Games form an effective learning vector (Clarke et al., 2017) and millennials have been observed to be more sensitive to it than to theoretical concepts. These are the reasons why we have witnessed the rise of a field of research in its own right dedicated to reconciling games and learning, namely *game-based learning* (Queiruga-Dios et al., 2020).

For the first time in 2007, a novel type of game appeared in Japan (López-Pernas et al., 2019a) that would experience a phenomenal success around the world: *escape games* (Gordillo et al., 2020). This success, coupled with the growing research interest in the benefits of so-called gamification in the learning process (López-Pernas et al., 2019a), has led researchers to consider the transposition of game mechanisms for educational purposes. This is how we eventually witnessed the advent of *educational escape games* (Veldkamp et al., 2020).

All the studies carried out thus far conclude with a neutral or positive evaluation as to the benefits in terms of user experience and learning of educational escape games intended for computer science teaching. However, these studies lack reproducibility (Petri

[a] https://orcid.org/0000-0001-6430-8168
[b] https://orcid.org/0000-0003-3769-6294

Kabimbi Ngoy, R., Yernaux, G. and Vanhoof, W.
EvscApp: Evaluating the Pedagogical Relevance of Educational Escape Games for Computer Science.
DOI: 10.5220/0011715100003470
In *Proceedings of the 15th International Conference on Computer Supported Education (CSEDU 2023) - Volume 2*, pages 241-251
ISBN: 978-989-758-641-5; ISSN: 2184-5026

et al., 2016), were carried out on confidential samples (Deeb and Hickey, 2019) or on an ad hoc basis. It is therefore difficult to make any attempt at an objective general conclusion as to their educational effectiveness based on their results (Veldkamp et al., 2020). This issue is all the more important as the design of such educational games is time-consuming and can be expensive (Taladriz, 2021).

Throughout the paper, we will introduce and present EvscApp, an evaluation framework built upon existing literature results in education, computer science education in particular, and game-based education, that has been implemented as an open source Web application (Kabimbi Ngoy et al., 2022). Thanks to this new tool, it will now be possible to replicate experiments through specification sharing, to conclude on the educational effectiveness of escape games– including games that use interactive IoT-based riddle mechanisms– by aggregating the collected results, and to compare experiments with each other. As a result, researchers in the field will see their research work related to the methodologies for evaluating pedagogical escape games reduced.

The paper is structured as follows. In Section 2, we give a state of the art focused on educational escape games for computer science. Then, in Section 3 we introduce and motivate the fifteen-steps evaluation method that is incorporated in EvscApp. We give an overview of the different features offered in the first version of the EvscApp dashboard. We conclude in Section 5 by discussing future perspectives for our work.

2 STATE OF THE ART

An educational escape game destined for teaching computer science (abbreviated EEGC in what follows) is a team game whose final objective is to discover a secret code or an artifact allowing the opening of a sealed object or door, within a given time frame (Lathwesen and Belova, 2021). The EEGCs belong to the family of *serious games* which are defined as activities that transpose the mechanisms used in games for learning purposes (Lathwesen and Belova, 2021). Serious games can in turn be seen as an instance of the somewhat broader concept of *gamification*. The latter is usually defined as the application of gaming mechanisms in non-gaming environments, including, but not restricted to, educational contexts (Caponetto et al., 2014).

While there is no exclusive definition of the concept, EEGCs are generally played in teams of 2 to 8 players assisted by a game master whose role is to help the players to solve puzzles when the need arises or is expressed explicitely (Nicholson, 2015).

A session typically starts with an introductive scenario immersing the players into the (often fictive) situation that defines the game's starting point. The game can then begin: the players must successively solve a series of puzzles or tasks (Gordillo et al., 2020) which promote the learning (Veldkamp et al., 2020) of one of the 18 areas of knowledge mentioned in the computing curriculum (Sahami et al., 2013). At each stage of the game we observe the same scheme: there is a challenge (riddle/task), followed by a solution and eventually a reward (Wiemker et al., 2015). Solving these puzzles allows the team to evolve towards the final goal (Gordillo et al., 2020). At the end of the game, a debriefing is organized. The players and the game master can then discuss the logic of solving the encountered puzzles (Gordillo et al., 2020).

The interest displayed by the academic world for EEGCs finds its origin on the one hand in the fact that studies have concluded to a greater level of retention than traditional educational activities such as reading (Fu et al., 2009; Gibson and Bell, 2013). On the other hand, the fact that EEGCs natively apply the principles of active learning, collaborative learning and flow experience, which are recognized as promoting learning (Gordillo et al., 2020), is also a reason for this enthusiasm. Active learning is derived from the theory of constructivism, which advocates the construction of knowledge rather than direct transmission from teacher to student (Ben-Ari, 1998). Collaborative learning can also be seen as a constructivist theory. It consists of having students work in groups and discover new ways of understanding concepts (Laal and Ghodsi, 2012). As for the flow experience, it is defined as being a particular type of experience, namely an immersion state which tends to be optimal, or even extreme (Jennett et al., 2008).

Note that the educational effectiveness of an EEGC will depend in particular on the game structure (open, path-based or sequential) (Clarke et al., 2017), the number of participants (López-Pernas et al., 2019b) and the involvement of the game master (Gordillo et al., 2020) .

Existing EEGCs include games dealing with cybersecurity (Seebauer et al., 2020; Oroszi, 2019; Beguin et al., 2019; Taladriz, 2021), cryptography (Deeb and Hickey, 2019; Queiruga-Dios et al., 2020; Ho, 2018), propositional logic and mathematics applied to computer science (Aranda et al., 2021; Towler et al., 2020; Santos et al., 2021), software engineering (Gordillo et al., 2020), programming (López-Pernas et al., 2019a; López-Pernas et al., 2019b;

Michaeli and Romeike, 2021) and networks (Borrego Iglesias et al., 2017). The related studies seem pretty enthusiastic about the effectiveness of EEGCs in terms of learning computer science. All the cited papers that report an experiment conclude with a neutral or positive effect on learning, be it in terms of learning, motivation or user experience. However, these observations do not allow, to date, to define the real impact of EEGCs on the learning of computer science due to the fact that the evaluation methods differ from one study to another (Veldkamp et al., 2020), that the samples are of confidential size and that there is a lack of reproductibility of the experiments (Petri et al., 2016). Given the significant time required to develop an EEGC (Taladriz, 2021) and the pedagogical risk incurred by students who would be taught through these tools which have not yet empirically demonstrated their effectiveness, we deemed appropriate to develop a standard pedagogical evaluation tool for EEGC designers.

3 THE EVSCAPP EVALUATION FRAMEWORK

3.1 Description and Objectives

EvscApp is a quasi-experimental evaluation framework that aims to measure the educational effectiveness of EEGCs by collecting standardized and rigorous empirical data. Thanks to EvscApp, EEGC designers can avoid the heavy work of producing and justifying their evaluation protocol. It also makes it possible to compare the EEGCs that have been evaluated by EvscApp with one another and to replicate the experiments by sharing their specifications. The whole process has been implemented in a Web application.

3.2 Methodology

The construction of our evaluation method was carried out based on Basili's "Goal Question Metric" (Caldiera and Rombach, 1994). Our methodology is summarized in Figure 1.

First, we set out to define our research objectives and the factors of particular interest in the context of our study. This is called the "framing" phase (Petri et al., 2016). As mentioned above, our goal was to be able to position ourselves regarding the effectiveness of EEGCs in terms of learning rate. The evaluation factors retained in the experimentation of learning games for computer science were: motivation, user experience and learning (Petri et al., 2016). The first factor makes it possible to characterize the interest of the learners in the proposed game, its mechanics and its material. The user experience determines the level of fun, satisfaction and engagement of the players. The learning category reports on the acquired knowledge and the level of retention after the activity (Wang et al., 2011).

Next comes the "planning" phase (Petri and Gresse von Wangenheim, 2016). The idea was to carry out a literature review associated with the different measures and application evaluation protocols (Wang et al., 2011), so as to compile the best practices in the area to create quantitative measures for our three factors (Connolly et al., 2008). For the establishment of our data collection tools, we identified 446 evaluation points from twelve publications (López-Pernas et al., 2019a; Jennett et al., 2008; Gordillo et al., 2020; López-Pernas et al., 2019b; de Carvalho, 2012; Fu et al., 2009; Phan et al., 2016; Rêgo and de Medeiros, 2015; Tan et al., 2010; Petri et al., 2016; Bangor et al., 2009). The evaluation items included in our questionnaire were selected on the basis of what the authors of the related studies claimed to evaluate with them. In order to keep the scope of our study limited, some related criteria such as "I feel cooperative with my classmates" (Phan et al., 2016) were excluded as these were meant to measure social interaction rather than learning, experience or motivation. We discarded redundant elements as well and reformulated the residual elements when relevant. While some of our criteria were fully developed by our peers, some elements of our questionnaires originate from other sources (e.g. video game satisfaction surveys) but all of the selected criteria have already been used in a scientific context in the works cited above. The resulting evaluation factors are based on statistical indicators available as an appendix to this paper[1].

After the planning phase comes the "exploitation" step, which is EvscApp's current step, during which one selects an appropriate experimentation method and implements it (Wang et al., 2011). It is also during this step that the data collection is carried out (Petri and Gresse von Wangenheim, 2016).

3.3 Context of Use and Recommendations

EvscApp is intented to be used to evaluate the effectiveness of an activity under real application con-

[1] See the artifact Web page (Kabimbi Ngoy et al., 2022) for the appendices and all documentation of the Web application.

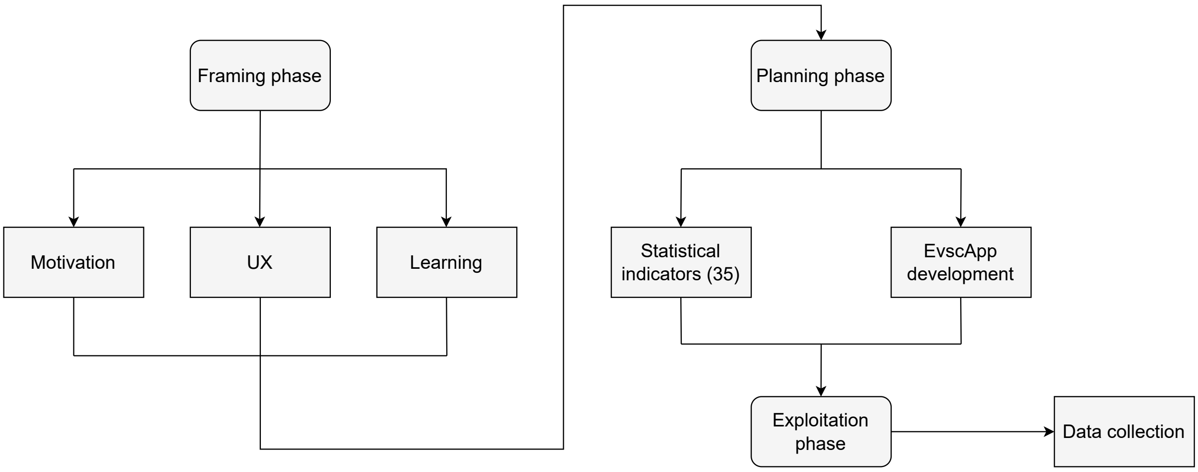

Figure 1: The methodology surrounding the EvscApp framework development.

ditions – a form of summative assessment that implements the quasi-experimental method. A quasi-experiment applied to teaching consists of an experiment in which one constitutes two different groups of students who both carry out a similar task but using a different teaching technique (Marín et al., 2018). We will denote by ”control group” the group of students carrying out the task according to the usual learning techniques, and by ”experimental group” (or ”target group”) the group using the technique being the subject of the study (Deeb and Hickey, 2019), in our case the EEGC submitted to evaluation. Although this experimental technique is not recognized as rigorously scientific, it is nevertheless widely used in fields related to social sciences and, in particular, pedagogy. Its use is justified by the fact that it is extremely difficult in social experiments to apply the one protocol that is specific to the experimental method, as it typically requires varying factors and observing results that are not impacted by external parameters over which the researchers have no control. For a long time discredited by the scientific community, the quasi-experiment has nowadays found some recognition of the academic world and its rigor is no longer questioned. However, it is essential to be aware of the potential influence of external parameters on the results observed and to mention it in the hypotheses and conclusions adopted (Campbell and Stanley, 2015).

For the evaluation to be relevant, the students forming the experimental group may not communicate with those of the control group (Christoph et al., 2006). It is also advisable to limit the introduction of other biases in the study (such as students revising the course to perform well in the experiment) and to collect all the necessary information in a limited time interval (Veldkamp et al., 2020).

An EEGC in the experimental phase involves a pedagogical risk for the students (Christoph et al., 2006). To protect students from negative effects on their learning, we recommend that the EEGC evaluation process, and therefore the use of EvscApp, be carried out outside of mandatory course periods and on a purely voluntary basis. Given this non-binding participation, two risks may arise: insufficient participation or too great a homogeneity in the students typology. To overcome these two potential problems, it will be necessary to find an element of motivation common to all the typological classes of students. Granting an incentive such as bonus grades for their simple participation might contribute to this objective (López-Pernas et al., 2019a). Under no circumstances should the performances achieved during the EEGC itself have an impact on the course grades, so as to limit the desire to cheat and thus introduce bias into the study (Gordillo et al., 2020).

3.4 Evaluation Process

Our evaluation protocol, EvscApp, is broken down into fifteen steps as shown in Figure 2. Documents and questionnaires relative to all of the steps can be found in the documentation folder of the Web application.

The protocol starts with the identification of participants. It is essential that the participants in the experiment belong to the final target population for which the EEGC is intended. To limit bias, one should ensure that the participants have not previously been taught on the topic addressed by the EEGC.

Next, the participants are informed of the purpose of the evaluation protocol, the duration of the process, the procedure itself and what they are committing to. They are then submitted a consent form (Chaves et al., 2015). By accepting it, participants agree:

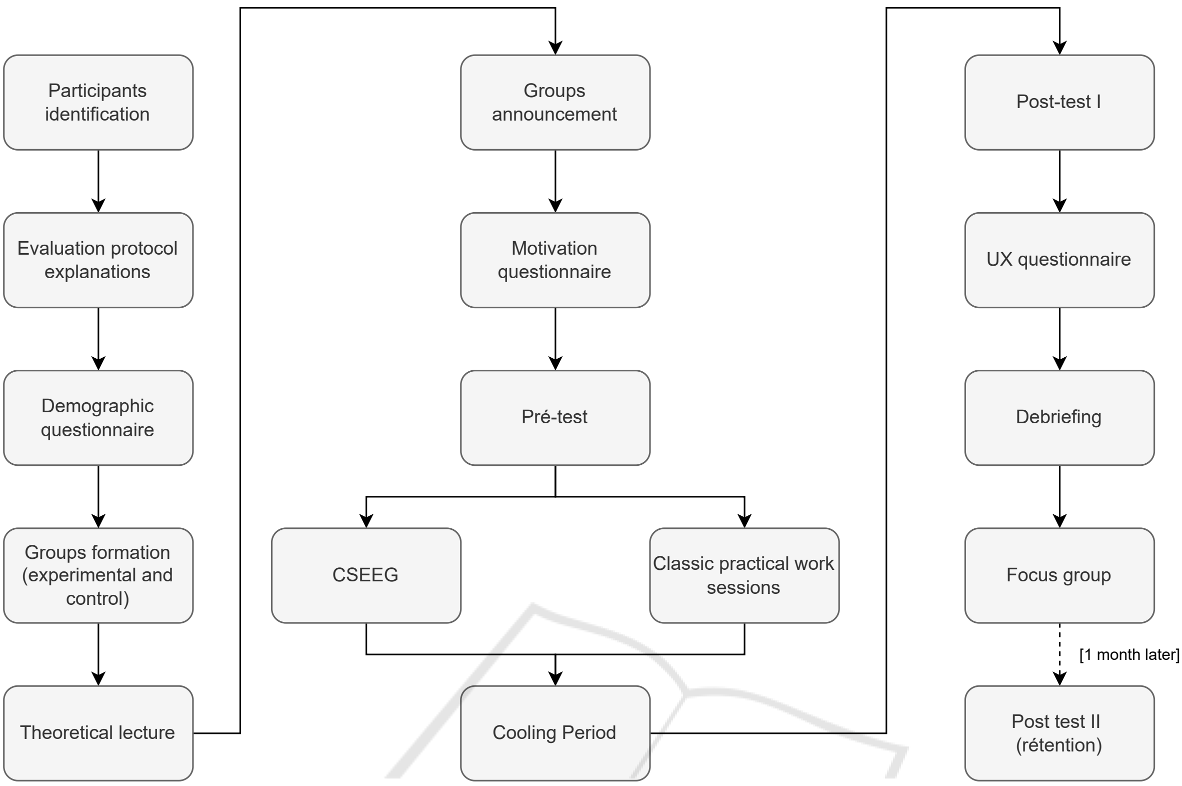

Figure 2: The EvscApp experimental process.

- to participate in the entire experience;
- that their personal data collected be used for research purposes, in compliance with applicable regulations (such as the GDPR[2]);
- to answer questionnaires, surveys and interviews honestly and sincerely;
- to participate in activities honestly and sincerely;
- to participate in the retention test which takes place one month after the activity;
- that they will not disclose any information to any third party that could compromise the results and conclusions of the experiment (e.g. protocol flow or questions asked in the tests).

Then, the participants are subjected to a demographic survey and assigned to a group (experimental or control) through a randomization process. The idea is to guarantee a certain balance between the groups. Doing so will provide a certain degree of confidence in observing any differences between these groups (Chaves et al., 2015). The assignment of each of the participants is not communicated at this stage to ensure that a balanced level of motivation and commitment is maintained for the two groups during the next step.

In said next step, all participants, regardless of the group to which they belong, will participate in a theoretical lecture on the subject to which the EEGC relates, so as to provide the theoretical bases that will enable students to solve the problems encountered during the activities. This prior course finds its justification in the fact that, in accordance with the active learning theory on which the EEGCs are based, we consider that these are a complementary support activity to the theoretical courses. The evaluated EEGC will therefore be opposed to a classic session of practical work.

Afterwards comes the assessment of participant motivation. Prior to the evaluation of this aspect, the assignment of each participant to a group is revealed. Proceeding in this way allows to keep as much enthusiasm as possible for the theoretical activity and to collect information related to the motivation for the activity that they are going to be led into. To do this, all participants, regardless of the group to which they belong, will answer a 5-factor Likert scale-type

[2]Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) - https://eur-lex.europa.eu/eli/reg/2016/679/oj.

questionnaire (from ”strongly disagree” to ”strongly agree”) comprising 3 evaluation points. Likert scale questionnaires are particularly used in the field of learning games for computer science. Such scales are said to allow factor analysis. The method aims to collect information on a complex and nuanced situation by reducing it to a few general elements covering all the possible answers that can be provided. It is therefore easier to analyze the resulting information than in the case of an open questionnaire, because the answers then become countable (Phan et al., 2016). Our Likert scale questionnaire consists in the assessment of three general criteria: ”I am excited about the educational activity I have been assigned to”, ”I am interested in the subject matter taught through the proposed activity” and ”I think that the subject matter taught through the activity is difficult to grasp”. As detailed above, these criteria were selected based on their relevance and their frequent use in the twelve related publications cited in Section 3.2.

During step number eight, all participants, regardless of the group to which they belong, will answer a test to assess their level of knowledge in the subject taught during the theoretical course. The test, unlike the questionnaire, targets all the techniques that extract information from a situation without the protagonists having the possibility of manipulating the vision that the researcher has of it. This type of collection lends itself well to the evaluation of learning, where we lead both a pre-test and a post-test. The pre-test consists of measuring the level of knowledge of the participants on the learning theme developed in the evaluated activity before it has started. The post-test has the same objective but will take place after the activity (de Carvalho, 2012). For ease of processing the information collected, the tests are presented as a multiple-choice questionnaire (López-Pernas et al., 2019a).

Then comes the main stage of our evaluation protocol, namely that during which the groups participate in the activities which are respectively dedicated to them. The students belonging to the control group will follow a traditional session of exercises equivalent in terms of duration to the time planned for the EEGC. By ”traditional session”, we mean that the students are given a series of exercises over a limited period of time. During such a session, individual questions can be asked. A collective correction takes place at the end of the session. This session should not propose the application of the principles of collaborative learning to students, without however prohibiting them if they arise naturally. This session will contain as many exercises as there are puzzles provided in the EEGC, in order to guarantee a certain equity between the students of the control and experimental groups. Meanwhile, a member of the pedagogical team will assume the role of game master for the EEGC played by the experimental group. Although this is a game practice envisioned in some EEGC, no penalty will be imposed on groups asking for clues, for the sake of fairness with the control group whose members can ask as many questions as they want. The only aspect that should drive the experimental students is to get to the end of the puzzles in the allotted time.

A ”cooling period” session will be scheduled after each activity, under the supervision of the teacher in charge of the experiment. The teacher should allow the students to relax while ensuring that complementary elements of understanding cannot be exchanged between members of different groups, which could result in biasing the results.

Next, in order to assess the progress in terms of students’ learning between ”before” and ”after” the activity, the students will again be subjected to a test. To avoid the bias relating to the difference in difficulty between pre-test and post-test, we opt for a test similar to the pre-test (Gordillo et al., 2020).

The twelfth step consists of a survey assessing the experimental students’ user experience. We again used a Likert scale, this time comprising 25 evaluation criteria, which we have selected by aggregating and adapting the criteria used in the twelve publications cited in Section 3.2. The interested reader can find the user experience questionnaire in the appendices of the paper. Note that to capture user satisfaction in the best possible way, responses must be spontaneous (Brooke, 2013). To this end, the questions should be presented one after the other and the students should have a time limit of 15 seconds per question.

At the end of this questionnaire, participants will receive their scores from the pre-test and post-test. They will be able to see the impact of the activity on their learning and debrief with the teacher. However, they will not have access to the correction, in order to limit the exchange of information between them and the next groups. During the debriefing, the players and the game master can then exchange on the logic of solving the riddles (Gordillo et al., 2020). A focus group will follow, where the participants wills complete their opinion in a semi-directed way, based on three questions.

After one month, all participants, regardless of the group to which they belong, will answer a questionnaire to assess their level of knowledge in the subject matter, so as to evaluate long-term retention. The questions will be different from those of the post-test in order to avoid the risk of automatic answers, which

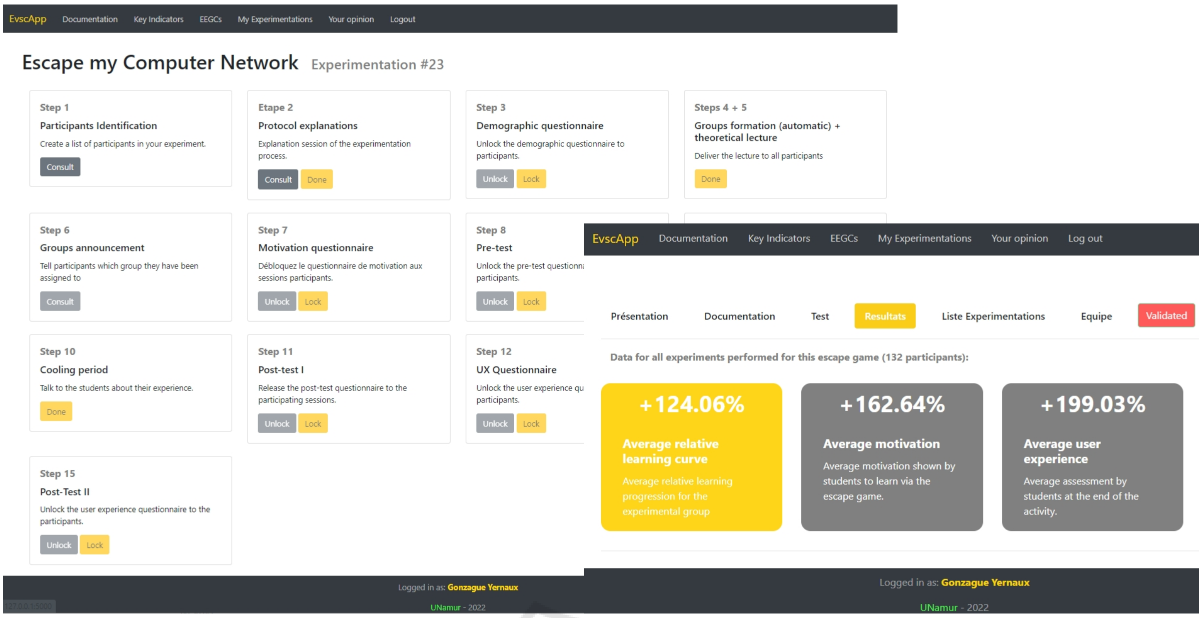

Figure 3: EvscApp Web app - experimentation dashboard (left) and results page (right).

this time will be real. This will allow to compare the level of retention between the two groups in the medium term (Connolly et al., 2008). However, one should be cautious about processing and interpreting the results from this phase as many factors may bias the results collected: exchanges between students, revision of the subject, modification of the sample, less participation in the survey, and so on (Lathwesen and Belova, 2021).

3.5 Application

We have transposed the evaluation process proposed above into a Web application. The artifact code and documentation is available online (Kabimbi Ngoy et al., 2022).

The tool allows to assist step by step the EEGC designers in the application of the EvscApp method and thus to guarantee the comparability between the results of the experiments that they will carry out. Concretely, our application allows designers to specify their own EEGCs, teachers interested in the use of EEGCs in their courses to consult the specification of the already-encoded EEGCs in order to experiment them and researchers to evaluate the EEGCs according to our method on aspects of motivation, user experience and learning.

The experimentation dashboard (left side of Figure 3) recapitulates the fifteen steps and automatizes a substantial part of these. For some steps, the experimentation is fully included in the tool. For instance, a link to the demographic questionnaire (step 3) is automatically send by email to the participants as soon as the researcher engages the corresponding button on the dashboard, allowing users to answer the demographic questions directly on the application. Pre- and post-tests are carried out just as easily. For other steps, such as the debriefing and focus group (steps 13 and 14), the application simply serves as a todolist-like reminder.

Other features include the administration of an escape game, such as the creation of pre/post-test questionnaires, the visualization of the key indicators for the EEGC in question (right side of Figure 3) and the encoding of useful information for anyone willing to replicate the EEGC.

At this stage of our work, the EvscApp Web application should be considered a minimum viable product that will evolve in a process of continuous improvement, based on feedback collected through a form on a dedicated page of the website.

4 EXPERIMENTATION

To get a first foretaste of the EvscApp approach we applied the method on an escape game developed at the University of Namur. The game is named Deskape. The idea of this EEGC is for (future) computer science students to try and access files that are hidden in a locked drawer of a desk. To open the desk, one has to pass a series of tests allegedly planted there by a mad teaching assistant. The tests basically boil down to finding the correct sequence of five (RFID)

Table 1: Some results of three EvscApp-based experiments.

	MC	ME	PC	PE	P1C	P1E	P2C	P2E	RLGC	RLGE	RRC	RRE	UX
1	62%	89%	8.2	9.0	14.3	16.5	13.2	15.5	+76.1%	+91.4%	91.8%	95.0%	74%
2	53%	88%	6.1	7.7	12.0	17.5	9.9	15.1	+94.5%	+129.7%	81.8%	88.5%	72%
3	59%	93%	10.9	9.4	15.6	15.8	12.1	13.6	+46.3%	+70.9%	76.9%	87.5%	87%

cards, all featuring the assistant. Each of the cards has a small detail that differs from the rest of the deck. Several riddles need to be solved in order to get information about which are the valid cards, and in which order the cards need be introduced so as to open the drawer. These riddles are scattered in and on the desk, sometimes in a cryptic form, sometimes displayed on an interactive screen. The players are also given summary sheets that expose chunks of a typical first-year logic and programming course's material, which are needed in order to solve the riddles. The riddles include:

- a logical formula that needs to be converted into the corresponding logical circuit, which points to the RFID card on which the correct circuit is printed;
- a small pseudo-code algorithm which is supposed to output a given string, which again is printed on a specific RFID card;
- a cipher that one needs to decipher using Caesar's algorithm in order to get an indication on the order in which the cards must be presented to the RFID receiver.

The experiments have been run on three groups, each composed of twenty high school senior students potentially willing to pursue their studies in an IT sector. Of these twenty students, ten were assigned to the control group and ten to the experimental group. The EvscApp method has been followed thoroughly. Similarly to the EEGC players group, the control group had access to summary sheets that could help solve some of the exercises.

The results are presented in Table 1. The following acronyms are used. MC, resp. ME, represents the average displayed motivation to learn through the EEGC, as measured by the motivation questionnaire for the control, resp. experimental, group. PC is the mean score on the pre-test (/20) for the students in the control group; PE is this score but for the experimental group. P1C is the mean score (with again 20 as the maximum) on the post-test 1 for the control group, while P1E is the same for the experimental group. P2C and P2E are the respective pendants for post-test 2. RLGC is the relative average learning growth for the control group whereas RLGE concerns the experimental group. Note that RLGC, respectively RLGE, are the mean of the ratios obtained for each control (resp. experimental) student's as the result on the first post test divided by the pre-test result. Similarly, RRC and RRE are the average long-term retention rates, computed as the mean of the individual ratios between P2(C/E) (for a single student) and P1(C/E) for the same student. UX represents the mean UX feedback score from the game played by the experimental group, as measured by the UX questionnaire.

From these preliminary experiments, we can draw the following early conclusions.

- There is a tendency for the students that are going to participate to Deskape to be more motivated than their control group comrades, i.e. $ME > MC$ in a significant way.
- Although the test scores highly vary from one experiment to the other, the post-test 1 score seem to be relatively higher for the experimental groups. The relative average learning growth is also higher for the experimental group. In other words $P1E > P1C$ and $RLGE > RLGC$.
- Similarly, the difference between the scores of MP1C and MP2C is smaller than that between P1E and P2E, as measured by RRC and RRE respectively. This higher retention rate for the experimental group seems to point towards the experimental students being more marked, i.e. Mathematically $RRE > RRC$.

These observations indicate that the experimental groups, during a Deskape-based learning, are both more enthusiast at the beginning of the experiment and more skilled than the control students during the post-test phases. Informal interviews of several students also point towards the fact that the experimental students have had a better time and had a tendency to better incorporate the concepts that have been seen. Overall, these informal conversations seem to confirm the results obtained by EvscApp, which is a first promising feedback of our novel method.

Note that in the experiments, we have not announced to the students the date at which the retention test (post-test 2) would be carried out, so that the students could not have easily cheated by revising the course material. Also note that it has been verified that the students were relatively evenly distributed among the control and experimental groups, based on their high school grades in mathematics.

5 CONCLUSIONS AND FUTURE WORK

Rethinking learning methods has been identified as a potential way to make computer science more accessible and attractive. In this respect, EEGCs are the subject of particular enthusiasm. However, to our knowledge no general-purpose empirical study allows to conclude as to their real pedagogical effectiveness.

In this work, we proposed to take a step towards filling this gap by the definition of an EEGC evaluation framework that has been transposed into a Web application: EvscApp. A first version of the application is available for researchers to use and collect local data on their experiments, the long-term idea being to aggregate and take these results online. The tool as it is offers the possibility to EEGC designers to use a standardized and structured process for evaluating escape games, to make replicable experiments and to compare the results achieved by different EEGCs. The application implements various evaluation criteria that are processed to quantitatively assess motivation, user experience and learning. While the definition of these factors is based on a comprehensive literature review, there is room for improvement in the exact computations that compose each of the factors. We have built EvscApp as a parametric framework relative to these factors, allowing to easily change or adapt their definition if the need arises. The framework shows good results on some pioneering experiments and more of those should be carried out to better understand the outcome.

Note that in the experiments, the students playing the Deskape game were dispatched in small groups of two or three students. An interesting extension of the framework would be one that takes into account the impact of the pairs (or trios) that are (randomly) formed as riddle-solving teams, e.g. considering the scores obtained by the students during pre- and post-test depending on their team.

Although still in its infancy, we firmly believe that EvscApp can have some interest in a context of IT education. Future research will focus on further submitting EvscApp for evaluation. To do this we will proceed at two levels. First of all, we will submit the evaluation questionnaires to exploratory and confirmatory factorial analysis. These analyses should allow to confirm the coherence, relevance and reliability of the selected evaluation elements, notably thanks to Cronbach's alpha. In a second step, we will organize more experimental sessions with EEGC designers by e.g. conducting an ethnographic analysis, collecting information based on semi-structured interviews and UEQ user experience and usability forms. We will then be able to empirically confirm the interest of our method, improve the app's user experience and develop complementary functionalities that seem the most appreciated by our peers. After this, we intend to develop and deploy an improved and online version of EvscApp to centralize experiments on EEGCs. Given the relatively broad scope of our approach, we also plan to investigate whether the EvscApp method could be applied to other sectors than computer science alone.

REFERENCES

Aranda, D., Towler, A., Ramyaa, R., and Kuo, R. (2021). The usability of using educational game for teaching foundational concept in propositional logic. In *2021 International Conference on Advanced Learning Technologies (ICALT)*, pages 236–237.

Armstrong, S. (2022). A shortage of tech workers could be on the horizon. is it time for you to upskill? https://www.euronews.com/next/2022/12/21/a-shortage-of-tech-workers-could-be-on-the-horizon-is-it-time-for-you-to-upskill. Accessed: 2023-02-03.

Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *J. Usability Studies*, 4(3):114–123.

Beguin, E., Besnard, S., Cros, A., Joannes, B., Leclerc-Istria, O., Noel, A., Roels, N., Taleb, F., Thongphan, J., Alata, E., and Nicomette, V. (2019). Computer-security-oriented escape room. *IEEE Security and Privacy*, 17:78–83.

Ben-Ari, M. (1998). Constructivism in computer science education. In *Proceedings of the Twenty-Ninth SIGCSE Technical Symposium on Computer Science Education*, SIGCSE '98, pages 257–261, New York, NY, USA. Association for Computing Machinery.

Borrego Iglesias, C., Fernández-Córdoba, C., Blanes, I., and Robles, S. (2017). Room escape at class: Escape games activities to facilitate the motivation and learning in computer science. *Journal of Technology and Science Education*, 7:162.

Brooke, J. (2013). Sus: a retrospective. *Journal of usability studies*, 8(2):29–40.

Caldiera, V. R. B. G. and Rombach, H. D. (1994). The goal question metric approach. *Encyclopedia of software engineering*, pages 528–532.

Campbell, D. T. and Stanley, J. C. (2015). *Experimental and quasi-experimental designs for research*. Ravenio books.

Caponetto, I., Earp, J., and Ott, M. (2014). Gamification and education: A literature review. In *European Conference on Games Based Learning*, volume 1, page 50. Academic Conferences International Limited.

Chaves, R. O., von Wangenheim, C. G., Furtado, J. C. C., Oliveira, S. R. B., Santos, A., and Favero, E. L. (2015). Experimental evaluation of a serious game for teaching software process modeling. *ieee Transactions on Education*, 58(4):289–296.

Christoph, L. H. et al. (2006). *The role of metacognitive skills in learning to solve problems*. SIKS.

Clarke, S., Peel, D., Arnab, S., Morini, L., Keegan, H., and Wood, O. (2017). Escaped: A framework for creating educational escape rooms and interactive games to for higher/further education. *International Journal of Serious Games*, 4.

Connolly, T., Stansfield, M. H., and Hainey, T. (2008). Development of a general framework for evaluating games-based learning. In *Proceedings of the 2nd European conference on games-based learning*, pages 105–114. Universitat Oberta de Catalunya.

Coquard, E. (2021). The impact of the global tech talent shortage on businesses. *Medium*. Accessed: 2023-02-03.

de Carvalho, C. V. (2012). Is game-based learning suitable for engineering education? In *Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON)*, pages 1–8.

Deeb, F. A. and Hickey, T. J. (2019). Teaching introductory cryptography using a 3d escape-the-room game. In *2019 IEEE Frontiers in Education Conference (FIE)*, pages 1–6.

Fu, F.-L., Su, R.-C., and Yu, S.-C. (2009). Egameflow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education*, 52:101–112.

Gibson, B. and Bell, T. (2013). Evaluation of games for teaching computer science. In *Proceedings of the 8th Workshop in Primary and Secondary Computing Education*, WiPSE '13, pages 51–60, New York, NY, USA. Association for Computing Machinery.

Gordillo, A., López-Fernández, D., López-Pernas, S., and Quemada, J. (2020). Evaluating an educational escape room conducted remotely for teaching software engineering. *IEEE Access*, 8:225032–225051.

Ho, A. (2018). Unlocking ideas: Using escape room puzzles in a cryptography classroom. *PRIMUS*, 28.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9):641–661.

Kabimbi Ngoy, R., Yernaux, G., and Vanhoof, W. (2022). Artifact documentation and code (online). https://github.com/rkabimbi/evscapp.

Laal, M. and Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia - Social and Behavioral Sciences*, 31:486–490. World Conference on Learning, Teaching & Administration - 2011.

Lathwesen, C. and Belova, N. (2021). Escape rooms in stem teaching and learning—prospective field or declining trend? a literature review. *Education Sciences*, 11:308.

López-Pernas, S., Gordillo, A., Barra, E., and Quemada, J. (2019a). Analyzing learning effectiveness and students' perceptions of an educational escape room in a programming course in higher education. *IEEE Access*, 7:184221–184234.

López-Pernas, S., Gordillo, A., Barra, E., and Quemada, J. (2019b). Examining the use of an educational escape room for teaching programming in a higher education setting. *IEEE Access*, 7:31723–31737.

Marín, B., Frez, J., Cruz-Lemus, J., and Genero, M. (2018). An empirical investigation on the benefits of gamification in programming courses. *ACM Trans. Comput. Educ.*, 19(1).

Michaeli, T. and Romeike, R. (2021). Developing a real world escape room for assessing preexisting debugging experience of k12 students. In *2021 IEEE Global Engineering Education Conference (EDUCON)*, pages 521–529.

Nicholson, S. (2015). Peeking behind the locked door: A survey of escape room facilities.

Oroszi, E. D. (2019). Security awareness escape room - a possible new method in improving security awareness of users. In *2019 International Conference on Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)*, pages 1–4.

Pawlowski, S. and Datta, P. (2001). Organizational responses to the shortage of it professionals: A resource dependence theory framework.

Petri, G. and Gresse von Wangenheim, C. (2016). How to evaluate educational games: a systematic literature review. *Journal of Universal Computer Science*, 22:992.

Petri, G., Gresse von Wangenheim, C., and Borgatto, A. (2016). Meega+: An evolution of a model for the evaluation of educational games.

Phan, M. H., Keebler, J. R., and Chaparro, B. S. (2016). The development and validation of the game user experience satisfaction scale (guess). *Human factors*, 58(8):1217–1247.

Queiruga-Dios, A., Santos, M., Dios, M., Gayoso Martínez, V., and Encinas, A. (2020). A virus infected your laptop. let's play an escape game. *Mathematics*, 8:166.

Rêgo, M. and de Medeiros, I. (2015). Heeg: Heuristic evaluation for educational games. *Proceedings of SBGames*.

Sahami, M., Roach, S., Cuadros-Vargas, E., and LeBlanc, R. (2013). Acm/ieee-cs computer science curriculum 2013: Reviewing the ironman report. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education*, SIGCSE '13, pages 13–14, New York, NY, USA. Association for Computing Machinery.

Santos, A. M., Sá, S., Costa, L. F. C., and Coheur, L. (2021). Setting up educational escape games: Lessons learned in a higher education setting. In *2021 4th International Conference of the Portuguese Society for Engineering Education (CISPEE)*, pages 1–8.

Seebauer, S., Jahn, S., and Mottok, J. (2020). Learning from escape rooms? a study design concept measuring the effect of a cryptography educational escape room. In *2020 IEEE Global Engineering Education Conference (EDUCON)*, pages 1684–1685.

Taladriz, C. C. (2021). Flipped mastery and gamification to teach computer networks in a cybersecurity engineering degree during covid-19. In *2021 IEEE Global Engineering Education Conference (EDUCON)*, pages 1624–1629.

Tan, J. L., Goh, D. H.-L., Ang, R. P., and Huan, V. S. (2010). Usability and playability heuristics for evaluation of an instructional game. In Sanchez, J. and Zhang, K., editors, *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2010*, pages 363–373, Orlando, Florida, USA. Association for the Advancement of Computing in Education (AACE).

Towler, A., Aranda, D., Ramyaa, R., and Kuo, R. (2020). Using educational game for engaging students in learning foundational concepts of propositional logic. In *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*, pages 208–209.

Veldkamp, A., van de Grint, L., Knippels, M.-C. P., and van Joolingen, W. R. (2020). Escape education: A systematic review on escape rooms in education. *Educational Research Review*, 31:100364.

Wang, Y.-Q., Liu, X., Lin, X., and Xiang, G. (2011). An evaluation framework for game-based learning.

Wiemker, M., Elumir, A., and Clare (2015). *Escape Room Games: Can you transform an unpleasant situation into a pleasant one?*