Inspecting Quality of Games Designed for Learning Programming

Tihomir Orehovački[1(✉)] and Snježana Babić[2]

[1] Faculty of Organization and Informatics, University of Zagreb, Pavlinska 2, 42000 Varaždin, Croatia
tihomir.orehovacki@foi.hr

[2] Polytechnic of Rijeka, Trpimirova 2/V, 51000 Rijeka, Croatia
snjezana.babic@veleri.hr

Abstract. Educational games are specific piece of slow technology that enables students to enhance their competences through fun and play. Taking into account their features, games are widely used in diverse educational settings, including programming. This paper discusses findings of the empirical study that was carried out with an objective to examine quality of two educational games meant for learning programming concepts. Participants in the study were students from two Croatian higher education institutions. Subjective data was collected with pre- and post-use online questionnaires whereas objective data was gathered with reports in the form of in-game progress screenshots. The analysis of collected data uncovered: (1) to what extent students with different background knowledge differ in their perception of programming and attitudes towards using games for educational purposes, and (2) which pragmatic and hedonic facets of quality are relevant for the assessment of games designed for learning programming.

Keywords: Educational games · Quality evaluation · Learning programming · Subjective and objective measuring instruments · Empirical findings

1 Introduction

Quality refers to the extent to which a piece of software meets user needs or expectations [13]. In the current international standard on software quality evaluation [14], these needs are represented by two quality models. The first one is product quality model which consist of 31 attributes that constitute eight categories aimed for measuring static software properties. The second one is quality in use model which is decomposed into five categories that deal with the assessment of the outcomes of the interaction with a software in a particular context of use. Evaluation is a systematic process of measuring the value and importance of software features [9]. Apart from being used for the identification of certain usability issues during the interaction with the software, evaluation is also meant for testing the availability of software functionalities as well as for measuring the user experience [7]. Consequently, quality evaluation should not be reserved only for the end of the development process but has to be carried out continuously through all stages of the software life cycle.

P. Zaphiris and A. Ioannou (Eds.): LCT 2015, LNCS 9192, pp. 620–631, 2015.
DOI: 10.1007/978-3-319-20609-7_58

Computer games are specific breed of software. Given that their features reflect hedonic facets of quality, computer games are often referred to as an example of slow technology [24]. The employment of computer games in educational settings enables students to acquire new knowledge and skills through fun and play which increases their motivation throughout the learning process [30]. This is because computer games stimulate interaction which is one of the essential aspects of the learning process [31]. Educational computer games must be designed in a way to serve as a valuable pedagogical tool for enhancing learning outcomes [34]. According to Sharda [32], design of the educational content, the plot, and the game itself represents a foundation for the assessment of user experience. In that respect, when development of educational computer games is considered, special attention should be paid to the efficient interplay of these three design domains. Taking into account that educational games are useful means for learning and teaching programming concepts [31], a number of them have been developed specifically for programming courses. Detailed overview of educational games focused on teaching programming can be found in [19].

This paper has several objectives. First, to discover to what extent students with different educational background differ in their perception of programming as a discipline at university level and in their attitude towards employing games for the purpose of learning programming concepts. Second, to examine if there is a significant difference between programming novices and experienced programmers in terms of their in-game progress within given time interval. Third, to identify which subjective and objective quality attributes are relevant for the assessment of games designed for learning programming.

The remainder of the paper is organized as follows. Next section offers overview of relevant advances in the field. Brief description of all constructs which constitute research framework is provided in the third section. Findings of an empirical study are reported in the fourth section. Practical implications of our work and study limitations are discussed in the last section.

2 Rationale and Background

Recent HCI literature offers a number of studies aimed for evaluating the various aspects of quality in the context of web sites [8], Web 2.0 applications [27], cloud based applications [23], mashups [4], mashup tools [28], mobile applications [21], and educational artefacts [29]. On the other hand, current research related to educational games is mostly focused on exploring predictors of their adoption while studies on quality assessment are rather rare. Drawing on motivation theory [6], Long [18] found that intrinsic motivation factors such as expectation of fun and opportunity to learn new programming skills significantly contribute to users' behavior related to playing the educational game Robocode whereas extrinsic factors ("to win the game", "to win the prize in the con-test" and "to gain peer recognition") proved to be less important in that respect. With an objective to examine hedonic quality of slow technology (such as computer game Braid), Orehovački et al. [24] have proposed a model which reflects an interplay among constructs adopted from expectation-confirmation theory [20],

theory of flow [3], and technology acceptance model [5]. Outcomes of the study carried out by Tao et al. [33] revealed that in the context of business simulation game: (1) perceived attractiveness significantly contributes to perceived ease of use and perceived usefulness; (2) perceived usefulness has a significant influence on learning performance; (3) perceived playfulness is significantly affected by perceived attractiveness and perceived ease of use; (4) learning performance impacts confirmation of expectations; and (5) learning performance and playfulness are significant predictors of students' satisfaction. As a follow up, Liao and Wang [17] discovered that learning motivation is a strong predictor of learning expectations which in turn together with perceived playfulness contribute to the perceived learning performance, learning confirmation is affected by perceived learning performance, while learning expectation, learning confirmation, and perceived learning performance have strong influence on learning satisfaction. By employing AttrakDiff2 questionnaire [11] and a scale adapted from ARCS motivational design model [15], Zaharias and Chatzeparaskevaidou [36] found that pragmatic quality, hedonic quality stimulation, hedonic quality identification, and attraction are significant predictors of students' motivation to learn. Finally, Ibrahim et al. [12] discovered that only two (effort expectancy and attitude) of four constructs introduced in the UTAUT [35] model, significantly affect students' preferences related to online educational games.

3 Research Framework

The success of certain technology greatly depends on users' continuance intentions and satisfaction that is influenced by confirmation of their expectations with respect to the technology employment [25]. The research framework introduced in this paper is composed of attributes that reflect students' attitudes towards playing games and programming as well as of attributes meant for measuring quality of educational games designed for learning programming concepts.

Based on the model proposed in [27] and refined in [26], a set of pragmatic and hedonic quality attributes was identified and adopted to the context of educational games. The extent to which students are ready to adopt games for the purpose of learning programming concepts was evaluated with attributes aimed for measuring playfulness, satisfaction, aesthetics, and loyalty. Quality of educational games considered from the perspective of a system was examined with items meant for assessing consistency. Attributes such as helpfulness, reliability, and feedback were employed for exploring quality of interaction between selected games and users. The extent to which the use of educational games enhances students' performances in learning programming concepts was investigated with attributes designed for measuring effectiveness, usefulness and efficiency. The effortlessness in employing educational games for learning programming was assessed with attributes meant for measuring minimal action, minimal memory load, accessibility, ease of use, learnability, memorability, and understandability.

Scales for measuring users' attitudes towards use, self-efficacy, gameplay anxiety, programming anxiety, perceived behavioural control, result demonstrability, and the perception of programming complexity were adopted from UTAUT model [35].

The degree to which educational games involved in an empirical study are distinctive among games with the same purpose was determined with items meant for measuring uniqueness [22]. The level to which students like to inform themselves and play new games was identified with items assigned to the attribute "resistance to changes" [22]. Dimensions of popularity in the context of educational games for learning programming were explored with items meant for measuring reputation [22]. The extent to which students believe that employment of games can provide them with opportunities for learning programming concepts was measured with items proposed in [12]. Preferences for using games in educational ecosystem, which appeared to be strong predictor of learning opportunities, were evaluated with items proposed by Bourgonjon et al. [1]. Finally, the intrinsic motivational factor which reflects the degree of difficulty and challenge a user experiences when playing a game was measured with items that were adopted from Hainey et al. [10].

4 Results

Participants. A total of 175 subjects were involved in the study. The sample was comprised of 61.71 % male and 38.29 % female participants. They ranged in age from 18 to 45 years (M = 19.72, SD = 2.286). Majority of them (70.86 %) were students at Polytechnic of Rijeka (POLYRI) while remaining 29.14 % studied at University of Zagreb, Faculty of Organization and Informatics (FOI). Most of the sample (96.57 %) were full-time students. Majority of study participants (56.01 %) play computer games at least once a week where 54.86 % of them spend between one and three hours a week on a game play. Study participants are also loyal players of mobile games. Most of them (51.43 %) play mobile games between once and twice a week where 36.57 % of students spend less than an hour on interaction with mobile games. Only 20.58 % and 11.43 % of research subjects play computer and mobile games, respectively, on daily basis.

Procedure. Prior to the implementation of the study, FOI students have completed one course related to the programming and at the time when the study took place they were enrolled in additional two programming courses. In that respect, they can be referred to as experienced programmers. On the other hand, general programming concepts were introduced to POLYRI students several weeks before the study was carried out. Therefore, they can be appointed as programming novices.

The study was composed of three main parts. At the beginning, participants were briefly informed about the purpose of the study which was followed by the explanation of the procedure that was employed. The participants were then asked to complete a pre-use questionnaire that consisted of 60 items related to research subjects' demographics, frequency of using computer and mobile games, their interest for advances in the field as well as prior experience in playing games, their perception about programming and preferences related to the use of games in educational ecosystem, and perceived fear of programming and playing games. Thereafter, students were asked to play each of two educational games for 30 minutes. The first one was LightBot 2.0[1] in which study

[1] http://armorgames.com/play/6061/light-bot-20.

participants had to use programming logic to solve assignments that were presented in the form of puzzles. The second one was CodeCombat[2] in which research subjects had to write and execute snippets of code in Python programming language in order to solve implemented quests. Examples of assignments from both games are shown in Fig. 1. As soon as the predefined time for playing each game had elapsed, participants had to make a report in the form of screenshot that illustrated which level they managed to reach. Examples of these reports for both games are presented in Fig. 2. At the end of the study, participants were asked to complete a post-use questionnaire that was composed of 150 items related to 28 different dimensions of quality adapted to the context of games designed for learning programming.

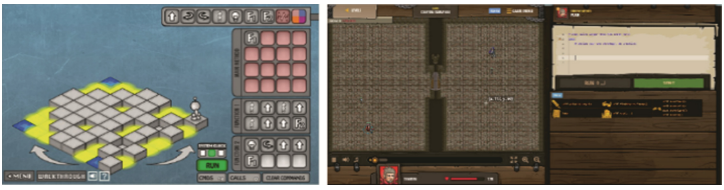

Fig. 1. Examples of assignments which constitute games aimed for learning programming (left: LightBot 2.0, right: CodeCombat).

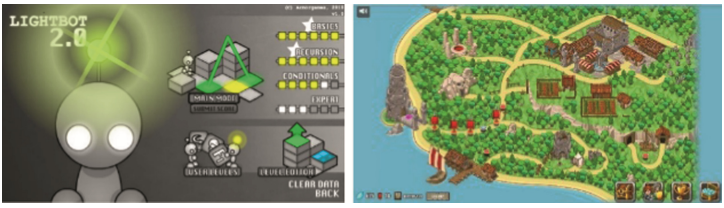

Fig. 2. Examples of screenshots that indicate which level particular student reached within predefined time interval (left: LightBot 2.0, right: CodeCombat).

Apparatus. Both pre- and post-use questionnaires were administrated online using the KwikSurveys[3] questionnaire builder. Responses to the questionnaire items were scored on a four point Likert scale (1– strongly agree, 4 – strongly disagree). The sum of responses to items assigned to particular attribute represent a composite measure which reflects relevant quality dimension. Differences between POLYRI and FOI students in terms of frequency of playing computer games, their prior experience with and preferences for computer games, and perceived anxiety related to programming and gameplay were explored with Mann-Whitney U statistics. The reason why we employed this non-parametric alternative to the independent t-test is because results of Shapiro-Wilk Tests uncovered that variables which constituted pre-use questionnaire together with variables that reflect game level which students reached within predefined time interval

[2] https://codecombat.com/.
[3] https://kwiksurveys.com.

significantly deviate from a normal distribution ($p < .05$). The analysis of data collected with post-use questionnaire adopted a within-subjects design contrasting two games meant for learning programming. Considering that the results of Shapiro-Wilk Tests revealed that at least one of the variables in a pairwise comparison violates the assumption of normality in data ($p < .05$), differences between evaluated games were examined by means of Wilcoxon Signed-Rank Tests. Taking into account the aforementioned, all reported results are expressed as median values. According to Cohen [2], values of .10, .30, or .50 for the size of an effect (r) can be, as a rule of thumb, interpreted as small, medium, or large, respectively.

Findings. The analysis of data collected with the pre-use questionnaire revealed that FOI students (Mdn = 13) are in general significantly less resistant ($U = 2456.50$, $p < .05$, $r = -.18$) to technological advances in terms of novel computer games than POLYRI students (Mdn = 14). More specifically, findings indicate that FOI students (Mdn = 2) like to discuss ($U = 2505.50$, $p < .05$, $r = -.17$) and browse information ($U = 2498.00$, $p < .05$, $r = -.17$) about new computer games significantly more often than POLYRI students do (Mdn = 3). On the other hand, there was no significant difference between these two groups of students (Mdn = 2) in terms of the extent to which they are interested in novel computer games ($U = 2801.50$, ns, $r = -.09$), the degree to which they are willing to inform themselves about new computer games ($U = 2605.50$, ns, $r = -.15$), the level to which they like to try novel computer games ($U = 2785.50$, ns, $r = -.10$), and the extent to which they are inclined to change a game they play in their leisure time ($U = 2657.00$, ns, $r = -.13$). According to the study results, FOI students (Mdn = 11) have significantly more experience ($U = 2291.00$, $p < .005$, $r = -.22$) in playing computer games than POLYRI students (Mdn = 14) have. Namely, compared to POLYRI students (Mdn = 2, 3, and 3, respectively), FOI students (Mdn = 2, 3, and 2, respectively) like to play computer games significantly more ($U = 2424.00$, $p < .01$, $r = -.20$), spend significantly more time in interaction with computer games ($U = 2436.50$, $p = .01$, $r = -.19$), and play significantly more diverse genres of computer games ($U = 2586.50$, $p < .05$, $r = -.22$). However, no significant difference was found between these two groups of students (Mdn = 2 and 3, respectively) in terms of the perceived frequency of playing computer games ($U = 2726.00$, ns, $r = -.12$) and perception of themselves as game play addicts ($U = 2599.00$, ns, $r = -.15$). Significant difference between FOI (Mdn = 4) and POLYRI students (Mdn = 6) was also found in terms of the extent to which people surrounding them play computer games ($U = 1774.00$, $p < .0001$, $r = -.36$). The set forth is particularly true for colleagues ($U = 1481.50$, $p < .0001$, $r = -.48$), friends ($U = 2250.00$, $p = .001$, $r = -.26$), and acquaintances ($U = 2578.50$, $p < .05$, $r = -.17$) of FOI students (Mdn = 1, 1, and 2, respectively).

The study results imply that FOI and POLYRI students (Mdn = 9) do not differ significantly ($U = 2637.50$, ns, $r = -.13$) in terms of the degree to which the reputation of a computer game affects their decision to play it. This is specifically true for the number of active game players ($U = 3031.00$, ns, $r = -.04$) as well as for position of a particular game on top lists ($U = 3046.00$, ns, $r = -.03$). However, it appeared that developer's name has significantly higher impact ($U = 2277.50$, $p = .001$, $r = -.24$) on

POLYRI students' decision (Mdn = 3) to play a particular game than it has on decision making process in that respect of FOI students (Mdn = 4). It was also discovered that FOI students (Mdn = 5) in general have significantly stronger ($U = 1898.00$, $p < .0001$, $r = -.32$) preference for the implementation of computer games in the educational settings than POLYRI students (Mdn = 6). Namely, significantly more FOI students (Mdn = 2, 2, and 2, respectively) was delighted with the idea of employing computer games for the educational purposes ($U = 2400.00$, $p < .01$, $r = -.20$) and would enroll ($U = 1763.50$, $p < .0001$, $r = -.37$) and actively participate ($U = 2085.50$, $p < .0001$, $r = -.28$) in courses on which games are played as a part of a syllabus. In general, there is no significant difference ($U = 3038.50$, ns, $r = -.03$) between FOI and POLYRI students (Mdn = 16) in terms of their reasoning about programming proficiency which is particularly affected by the fact that both groups of students (Mdn = 3) believe that intelligence plays an important role in the process of learning programming concepts ($U = 2745.00$, ns, $r = -.12$). However, significantly more POLYRI students (Mdn = 2, 2, and 2, respectively) believe that programming is a complex discipline ($U = 2421.50$, $p < .01$, $r = -.21$), that is difficult to learn programming concepts ($U = 2001.00$, $p < .0001$, $r = -.31$), and that one have to write large amount of programming code in order to solve relatively simple assignments ($U = 1453.00$, $p < .0001$, $r = -.46$). On the other hand, significantly more FOI students (Mdn = 1, 1, 1, and 1, respectively) think that programming is a comprehensive discipline ($U = 2539.00$, $p < .05$, $r = -.18$), that it requires a special way of thinking ($U = 2045.50$, $p < .0001$, $r = -.31$), that it must be learned with understanding ($U = 2302.00$, $p = .001$, $r = -.25$), and that in cannot be learned by heart ($U = 1609.50$, $p < .0001$, $r = -.43$).

Both POLYRI and FOI students (Mdn = 9) agree that educational background has an important role in learning programming concepts at university level ($U = 3039.00$, ns, $r = -.03$). The set forth especially refers to the knowledge of mathematics (Mdn = 2, $U = 3137.50$, ns, $r = -.01$). However, significantly more POLYRI students (Mdn = 2) believe that prior knowledge strongly affects the understanding of programming concepts ($U = 2251.50$, $p = .001$, $r = -.24$). On the other hand, significantly more FOI students (Mdn = 2) perceive programming as an interdisciplinary mastery ($U = 2457.00$, $p < .005$, $r = -.23$). The analysis of collected data revealed that POLYRI students (Mdn = 15) generally have significantly higher level of perceived programming anxiety ($U = 1841.50$, $p < .0001$, $r = -.33$) than FOI students (Mdn = 18). This is because significantly more POLYRI students (Mdn = 3, 3, 3, 3, and 3, respectively) feel nervous ($U = 2400.00$, $p < .01$, $r = -.20$), discomfort ($U = 1938.00$, $p < .0001$, $r = -.33$), agitation ($U = 2222.00$, $p = .001$, $r = -.25$), worry ($U = 2513.50$, $p < .05$, $r = -.17$), and aversion ($U = 1547.00$, $p < .0001$, $r = -.43$) when thinking about programming. It was also discovered that POLYRI students (Mdn = 12) in general feel significantly higher level ($U = 1891.00$, $p < .0001$, $r = -.32$) of game play anxiety than FOI students (Mdn = 15). Namely, significantly more POLYRI students (Mdn = 2, 2, 2, and 3, respectively) were worried that they will not be able to successfully complete assignments which constitute games aimed for learning programming ($U = 1803.00$, $p < .0001$, $r = -.36$), that the level of their prior knowledge will reduce their productivity in completing game assignments ($U = 1816.00$, $p < .0001$, $r = -.35$), that they will not be able to complete game assignments from the first attempt ($U = 2268.50$, $p = .001$, $r = -.24$), and that they will not be able to

concentrate on solving quests implemented in games (U = 2262.00, p = .001, r = -.24). Finally, FOI and POLYRI students (Mdn = 3) were equally worried that, compared to them, their peers will achieve better results in solving game assignments (U = 2634.50, ns, r = -.14).

The analysis of data collected from screenshots revealed that FOI students (Mdn = 11) were significantly more successful (U = 555.50, p < .0001, r = -.65) in completing assignments which constitute Light Bot 2.0 game than POLYRI students (Mdn = 8) were. It was also discovered that POLYRI students (Mdn = 7) completed significantly lower count of levels (U = 1108.00, p < .0001, r = -.51) of CodeCombat game than FOI students (Mdn = 14) did.

A Wilcoxon Signed-Rank Test showed that study participants have positive attitude towards the employment of both LightBot 2.0 and CodeCombat (Z = -.536, p = .592) for the purpose of learning programming concepts at the university level. It was also found that LightBot 2.0 (Mdn = 18) enhances the perceived self-efficacy of players to the significantly higher extent (Z = -2.573, p = .01, r = -.14) than CodeCombat (Mdn = 14). Moreover, it appeared that is significantly easier (Z = -3.417, p = .001, r = -.18) to figure out how to solve quests which constitute CodeCombat (Mdn = 6) than to learn how to complete assignments that are implemented into LightBot 2.0 (Mdn = 6). On the other hand, a Wilcoxon Signed-Rank Test did not elicit a statistically significant difference (Z = -.454, p = .650) between LightBot 2.0 (Mdn = 7) and CodeCombat (Mdn = 7) in terms of the degree to which is easy to memorize and recall how to solve game assignments. The analysis of data revealed that study participants had significantly less difficulties (Z = -3.039, p = .002, r = -.16) in completing quests integrated into CodeCombat (Mdn = 8) than solving assignments that are included into LightBot 2.0 (Mdn = 9). Study results also imply that players had to make significantly less physical effort related to the frequency of using a keyboard and mouse (Z = -9.191, p = .000, r = -.49) when they were completing tasks which are part of the CodeCombat (Mdn = 13) than when they were addressing assignments that are implemented into LightBot 2.0 (Mdn = 15). On the other hand, the study participants had to invest significantly more (Z = -4.722, p = .000, r = -.25) mental effort in terms of thinking and decision making when they were dealing with quests that constitute CodeCombat (Mdn = 17) than when they were solving assignments that are included into LightBot 2.0 (Mdn = 16). It was also discovered that user interface of CodeCombat (Mdn = 10) has been perceived by players as significantly more visually appealing (Z = -1.996, p = .046, r = -.11) than those of LightBot 2.0 (Mdn = 10). Moreover, the analysis of collected data yielded that CodeCombat (Mdn = 6) and LightBot (Mdn = 6) do not differ significantly (Z = -.652, p = .515) in terms of distinctive features. However, it was discovered that LightBot 2.0 (Mdn = 10) employs significantly more uniform interface structure, design, and terminology (Z = -2.934, p = .003, r = -.16) than CodeCombat (Mdn = 10) does. Furthermore, it appeared that players perceived CodeCombat (Mdn = 9) as significantly less dependable, stable, and bug-free game (Z = -6.753, p = .000, r = -.36) than LightBot 2.0 (Mdn = 7).

According to the results of the analysis of data collected with the post-use questionnaire, LightBot 2.0 (Mdn = 5) is significantly more (Z = -3.339, p = .001, r = -.18) usable to players with the widest range of characteristics and capabilities than

CodeCombat (Mdn = 5). However, LightBot 2.0 (Mdn = 7) and CodeCombat (Mdn = 6) do not differ significantly (Z = -.895, p = .371) in the extent to which their interface functionalities are clear and unambiguous to players. On the other hand, it was found that CodeCombat (Mdn = 10) offers significantly better and more diverse help materials (Z = -2.741, p = .006, r = -.15) than LightBot 2.0 (Mdn = 11) does. In addition, the quality of messages provided by CodeCombat (Mdn = 13) is significantly higher (Z = -2.679, p = .007, r = -.14) than those shown by LightBot 2.0 (Mdn = 14). Games meant for learning programming (Mdn = 9) do not differ significantly (Z = -.262, p = .793) in the degree to which they reduce the amount of time needed for learning specific programming concept. There was also no significant difference (Z = -1.854, p = .064) between evaluated games (Mdn = 14) in the extent to which they improve player's effectiveness in learning programming concepts. Moreover, LightBot 2.0 (Mdn = 10) and CodeCombat (Mdn = 10) do not differ significantly (Z = -.441, p = .659) in the degree to which their use contributes to the improvement of programming skills and habits. The level of perceived external control over addressing assignments was not significantly affected (Z = -.654, p = .513) by educational game (Mdn = 8) that was employed for that purpose. Evaluated games (Mdn = 7) do not differ significantly (Z = -1.102, p = .270) in terms of perceived learning outcomes that are result of their employment. No significant difference (Z = -1.567, p = .117) exist between evaluated games (Mdn = 13) regarding learning opportunities (e.g. evolution of logical and critical thinking, personalization of learning process, etc.) they offer to their players. Moreover, it was found that quality of implemented assignments is not significantly influenced (Z = -1.290, p = .197) by the game (Mdn = 12) that was used for learning programming concepts. However, it appeared that LightBot 2.0 (Mdn = 7) was significantly more challenging (Z = -4.593, p = .000, r = -.25) for study participants than CodeCombat (Mdn = 6).

Games aimed for learning programming (Mdn = 8) do not differ significantly (Z = -.941, p = .347) in the extent to which they have met expectations of research subjects. Nevertheless, the study results uncovered that LightBot 2.0 (Mdn = 15) was significantly more successful (Z = -2.260, p = .024, r = -.12) in arousing participants' imagination and stimulating they creativity in completing assignments than CodeCombat (Mdn = 16) was. There was no significant difference (Z = -.729, p = .466) between LightBot 2.0 (Mdn = 11) and CodeCombat (Mdn = 12) in terms of the enjoyment the study participants experienced when they were dealing with implemented assignments. CodeCombat (Mdn = 9) and LightBot 2.0 (Mdn = 8) do not differ significantly (Z = -1.714, p = .086) in the extent to which they made an overall impression on research subjects. Finally, it appeared that LightBot 2.0 (Mdn = 16) and CodeCombat (Mdn = 17) do not differ significantly (Z = -1.478, p = .139) in the degree to which study participants are willing to play them regularly and recommend them to others. Considering all the aforementioned, no significant difference (Z = -.076, p = .939) was found between CodeCombat (Mdn = 285) and LightBot 2.0 (Mdn = 287) in the overall perceived quality.

5 Discussion and Conclusion

This paper provides several contributions and implications to both scientific and professional communities. To begin with, the concept of quality introduced in current and relevant international standard [14] has been enhanced and adapted to the context of educational games. In addition, the validity of the employed research design and measuring instruments (reports in form of screenshots, pre- and post-use questionnaires) was empirically confirmed. Following the guidelines suggested by Lewis [16], the sensitivity of the pre-use questionnaire was explored through comparison of two groups of students which differed in terms of their prior knowledge related to programming whereas the sensitivity of reports and post-use questionnaire was examined by benchmarking two games meant for learning programming. All constructs that have met the criteria of sensitivity have shown small, medium, or large effect in size thus confirming the validity of measuring instruments.

Drawing on the results of validity testing, relevance of dimensions in testing differences among students with different educational background and examining games meant for learning programming was determined. Measure that reflects the number of levels which can be completed within predefined time interval revealed large in size differences between the groups of study participants. Items meant for measuring the influence of subjective norms on adoption of a game, programming anxiety, preference for the implementation of games in educational settings, and game play anxiety elicited medium in size differences between the groups of research subjects. Constructs that measure how often participants play games and to what extent they are interested in novel games uncovered small in size differences between groups they belong to. Measures that indicate to what degree students perceive programming as complex and interdisciplinary proficiency as well as to what degree reputation of a game affects their decision to play it have not detected significant differences between groups of study participants.

In the context of the quality assessment, it appeared that: (a) 3.57 % of proposed constructs revealed large (minimal action) and 3.57 % medium (reliability) in size differences between evaluated games; (b) 7.14 % of introduced constructs (minimal memory load and challenge) uncovered between small and medium in size effects between educational games; (c) 32.14 % of employed constructs (learnability, accessibility, ease of use, consistency, helpfulness, feedback, self-efficacy, and playfulness) elicited small in size differences between games; and (d) 53.57 % of remaining constructs (attitude towards behavior, memorability, uniqueness, understandability, efficiency, effectiveness, usefulness, external control, result demonstrability, learning opportunities, quality of assignments, confirmation of expectations, pleasure, satisfaction, and loyalty) did not show significant differences between evaluated games.

As in the case of most empirical studies, work presented in this paper has limitations. The first one is related to the homogeneity of study participants. Although students are representative users of educational games, heterogeneous sample could have importantly different perception about the quality of games that are used for learning programming concepts. The second one concerns the generalizability of reported findings. Considering that each genre of educational games has specific features which might

affect one or several quality dimensions, the empirical results should be interpreted with caution. Taking the aforementioned into account, further studies should be carried out in order to draw sound conclusions.

References

1. Bourgonjon, J., Valcke, M., Soetaert, R., De Wever, B., Schellens, T.: Parental acceptance of digital game-based learning. Comput. Educ. **57**(1), 1434–1444 (2011)
2. Cohen, J.: A power primer. Psychol. Bull. **112**(1), 155–159 (1992)
3. Csikszentmihalyi, M.: Beyond Boredom and Anxiety. Jossey-Bass, San Francisco (1975)
4. Daniel, F., Matera, M.: Quality in mashup development. In: Daniel, F., Matera, M. (eds.) Mashups: Concepts, Models, and Architectures, pp. 269–291. Springer, Heidelberg (2014)
5. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)
6. Deci, E.L.: Intrinsic Motivation. Plenum Press, New York (1975)
7. Dix, A., Finlay, J., Abowd, G.D., Beale, R.: Human-Computer Interaction, 3rd edn. Prentice-Hall, Haddington (2004)
8. Fogli, D., Guida, G.: A practical approach to the assessment of quality in use of corporate web sites. J. Syst. Softw. **99**, 52–65 (2015)
9. Gena, C., Weibelzahl, S.: Usability Engineering for the Adaptive Web. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 720–762. Springer, Heidelberg (2007)
10. Hainey, T., Westera, W., Connolly, T.M., Boyle, L., Baxter, G., Beeby, R.B., Soflano, M.: Students' attitudes toward playing games and using games in education: Comparing Scotland and the Netherlands. Comput. Educ. **69**, 474–484 (2013)
11. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. In: Ziegler, J., Szwillus, G. (eds.) Mensch & Computer 2003: Interaktion in Bewegung, pp. 187–196. B.G. Teubner, Stuttgart (2003)
12. Ibrahim, R., Wahab, S., Khalil, K., Jaafar, A.: Student perceptions of educational games in higher education: An empirical study. Issues Inf. Syst. **12**(1), 120–133 (2011)
13. Institute of Electrical and Electronics Engineers (IEEE) std. 610.12-1990. IEEE standard glossary of software engineering terminology (1990)
14. ISO/IEC 25010: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models (2011)
15. Keller, J.M.: Development and use of the ARCS model of motivational design. J. Instr. Dev. **10**(3), 2–10 (1987)
16. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int. J. Human-Comput. Interact. **7**(1), 57–78 (1995)
17. Liao, Y.W., Wang, Y.S.: Investigating the factors affecting students' continuance intention to use business simulation games in the context of digital learning. In: International Conference on Innovation, Management and Service, pp. 119–124. IACSIT Press, Singapore (2011)
18. Long, J.: Just for fun: using programming games in software programming training and education. J. Inf. Technol. Educ.: Res. **6**, 279–290 (2007)
19. Malliarakis, C., Satratzemi, M., Xinogalos, S.: Educational games for teaching computer programming. In: Karagiannidis, C., Politis, P., Karasavvidis, I. (eds.) Research on e-Learning and ICT in Education, pp. 87–98. Springer, Heidelberg (2014)

20. Oliver, R.L.: A cognitive model for the antecedents and consequences of satisfaction. J. Mark. Res. **17**(4), 460–469 (1980)
21. Olsina, L., Santos, L., Lew, P.: Evaluating mobileapp usability: a holistic quality approach. In: Casteleyn, S., Rossi, G., Winckler, M. (eds.) ICWE 2014. LNCS, vol. 8541, pp. 111–129. Springer, Heidelberg (2014)
22. Orehovački, T.: Methodology for evaluating the quality in use of Web 2.0 applications, Ph.D. thesis. University of Zagreb, Faculty of Organization & Informatics, Varaždin (2013)
23. Orehovački, T.: Perceived quality of cloud based applications for collaborative writing. In: Pokorny, J., et al. (eds.) Information Systems Development – Business Systems and Services: Modeling and Development, pp. 575–586. Springer, Heidelberg (2011)
24. Orehovački, T., Al Sokkar, A.A., Derboven, J., Khan, A.: Exploring the hedonic quality of slow technology. In: CHI 2013 workshop on Changing Perspectives of Time in HCI, http://bib.irb.hr/datoteka/617623.workshop_paper_final3.pdf (2013)
25. Orehovački, T., Babić, S.: Predicting students' continuance intention related to the use of collaborative Web 2.0 applications. In: Proceedings of the 23rd International Conference on Information Systems Development, pp. 112–122. Faculty of Organization and Informatics, Varaždin (2014)
26. Orehovački, T., Babić, S., Jadrić, M.: Exploring the validity of an instrument to measure the perceived quality in use of Web 2.0 applications with educational potential. In: Zaphiris, P., Ioannou, A. (eds.) LCT 2014, Part I. LNCS, vol. 8523, pp. 192–203. Springer, Heidelberg (2014)
27. Orehovački, T., Granić, A., Kermek, D.: Evaluating the Perceived and Estimated Quality in Use of Web 2.0 Applications. J. Syst. Softw. **86**(12), 3039–3059 (2013)
28. Orehovački, T., Granollers, T.: Subjective and objective assessment of mashup tools. In: Marcus, A. (ed.) DUXU 2014, Part I. LNCS, vol. 8517, pp. 340–351. Springer, Heidelberg (2014)
29. Orehovački, T., Žajdela Hrustek, N.: Development and validation of an instrument to measure the usability of educational artifacts created with Web 2.0 applications. In: Marcus, A. (ed.) DUXU 2013, Part I. LNCS, vol. 8012, pp. 369–378. Springer, Heidelberg (2013)
30. Prensky, M.: Computer games and learning: digital game-based learning. Handb. comput. Game stud. **18**, 97–122 (2005)
31. Seng, W.Y., Yatim, M.H.M.: Computer game as learning and teaching tool for object oriented programming in higher education institution. Procedia – Soc. Behav. Sci. **123**, 215–224 (2014)
32. Sharda, N.K.: Designing, Using and Evaluating Educational Games: Challenges, Some Solutions and Future Research. http://ceur-ws.org/Vol-386/p08.pdf (2008)
33. Tao, Y.H., Cheng, C.J., Sun, S.Y.: What influences college students to continue using business simulation games? Taiwan Experience. Comput. Educ. **53**(3), 929–939 (2009)
34. Torrente, J., Del Blanco, Á., Marchiori, E.J., Moreno-Ger, P., Fernández-Manjón, B.: < e-Adventure >: Introducing educational games in the learning process. In: 1st Annual Engineering Education Conference (EDUCON), pp. 1121–1126. IEEE, Madrid (2010)
35. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: toward a unified view. MIS Q. **27**(3), 425–478 (2003)
36. Zaharias, P., Chatzeparaskevaidou, I.: Hedonic and pragmatic qualities as predictors for motivation to learn in serious educational games. In: 8th International Conference on the Foundations of Digital Games (FDG) (2013). http://www.fdg2013.org/program/workshops/papers/G4L2013/g4l2013_04.pdf