A Pilot Study of a Digital Skill Tree in Gameful Education

Gustavo F. Tondello
HCI Games Group, Games Institute and School of Computer Science,
University of Waterloo, ON, Canada
gustavo@tondello.com

Lennart E. Nacke
HCI Games Group, Games Institute and Stratford School of Interaction Design and Business,
University of Waterloo, ON, Canada
lennart.nacke@acm.org

ABSTRACT

Gameful digital applications have been adopted in higher education to help increase student engagement and improve learning. However, many studies have only evaluated educational applications that combine some common game design elements—such as points, leaderboards, or levels. Consequently, we still lack studies exploring different ways of designing gameful learning experiences. Therefore, we introduce the design and implementation of a digital system employing a skill tree to mediate instructor feedback and assignment grading in a university course, Additionally, we present the results of a pilot evaluation with 16 students in which we summarized the positive and negative aspects of the experience to derive lessons learned for the use of digital skill trees in similar contexts. Finally, we suggest topics for further investigation.

Author Keywords

Gamification; Education; Skill Tree.

INTRODUCTION

Gamification is being adopted in education to improve learning and increase students' motivation and engagement. This tendency has been identified in systematic reviews focusing specifically on gamification for education [3, 9, 15] and gamification in general [16, 26]. Gamification is the use of game design elements in non-game contexts [8] or the use of affordances for gameful experiences to support users' value creation [12]. In education, elements of games can be used to make the content more interesting, to motivate students to complete more learning tasks, or to modify the way students are assessed and graded. Nonetheless, the majority of studies focus on a small subset of gamification elements, such as points, badges, leaderboards, levels, and avatars [9]. Therefore, we still lack studies exploring different ways of applying gamification elements to the learning experience.

At the same time, education scholars have been proposing new ways to improve students' motivation and performance in higher education, for example, by empowering them as self-regulated learners [23, 24]. Self-regulated students are able to regulate aspects of their thinking, motivation, and behaviour during learning, such as the setting of learning goals, the strategies to achieve these goals, the management of resources, the amount of effort exerted to study, and the reaction to external feedback. Two related mechanisms that can be used to foster self-regulation are open learner models (OLM) [4, 5, 6, 19] and self-assessment [22, 25, 27]. In intelligent tutoring systems, the learner model is the data about the student's current competence in the skills being taught [6]. When these models are opened to the students, they can better understand and self-regulate their learning journey. In turn, independent open learner models (IOLM) are similar representations, which are not connected to a specific tutoring system [5]. On the other hand, self-assessment refers to the student's judgement or rating of their own work [25, 27]. When the two concepts are combined, the IOLM supports self-assessment [19, 22] and represents a common tool that learners and instructors can use to discuss and plan the student's learning journey. Self-assessment and self-regulation have been shown to improve the learning outcomes [23, 25, 27].

In this work, we introduce a novel use of a digital skill tree as a gameful implementation of an independent learner model to support self-assessment and instructor assessment of student's work in a university-level computer science course. Skill trees are representations of progressive learning paths [29]. They have been used to organize lecture content (e.g., [1, 11, 18]), to provide structure and motivate students to complete additional learning tasks (e.g., [2, 10]), or as a visualization option for open learner models (e.g., [5, 7, 14, 20]). Uniquely, we describe a digital implementation of a skill tree to structure and mediate student self-evaluation and instructor assessment of the programming assignments completed by the students over the course of a four-month term. Additionally, we present a pilot evaluation of this design idea through a descriptive study with 16 students to understand their experiences with the skill tree. To conclude, we then summarize the positive and negative aspects of the students' experience with the gameful aspects of the course and derive general lessons learned.

Therefore, our work contributes to gamification and education by proposing a new way of implementing independent open learner models, self-assessment, and instructor assessment of students' work with a gameful approach using digital skill trees. This design concept can be further combined in future work with additional gameful design elements, such as badges or unlockable content, to provide a comprehensive gameful solution for higher education classrooms.

In: J. Arnedo-Moreno, C.S. González, A. Mora (eds.): Proceedings of the 3rd International Symposium on Gamification and Games for Learning (GamiLearn'19), Barcelona, Spain, 22-10-2019, published at http://ceur-ws.org

RELATED WORKS

Gamification Applied to Education

Studies of gamification in education comprise a considerable portion of the existing gamification literature [16, 26]. Gamification has been used in educational contexts with positive or mixed results to support a learning activity, improve an existing tutorial system, encourage participation, increase student motivation and engagement, and encourage students to do homework [26].

According to Kapp [13], gamification can be applied to education in two distinct ways. In *structural gamification*, the content is not altered and does not become game-like, but the structure around the content does. An example is using points, badges, and levels to track student progress. In *content gamification* on the other hand, the content itself is altered using game elements. An example is adding gameful story elements to modify the way the content is presented to learners. It is also important to note that the gamification of education is different from a serious game. A serious game is a full-fledged game with an instructional purpose, whereas gamification consists of inserting elements of games without turning the whole instruction into a full game [3, 9, 17].

Landers [17] proposed a theory of gamified learning, which indicates that gamification can affect learning via *moderation* when the instructor makes pre-existing content better in some way. An example is incorporating a gameful narrative into an existing learning plan. On the other hand, gamification affects learning via *mediation* when the instructor encourages a behaviour or attitude that itself should improve learning outcomes. An example is using gamification to increase the amount of time that students spend with the course material, which should cause greater learning.

Digital Skill Trees Applied to Education

Skill trees are used in games and gameful applications as a representation of progression [28, 29]. They have been used in gameful education with two different purposes: as a means to organize lecture content or to provide structure and motivate students to complete additional learning tasks. For example, Lee and Doh [18] suggest a design for a digital e-learning system that uses a skill tree to inform the user about what lectures they have already completed and what are the subsequent learning goals to pursue. Similarly, Anderson et al. [1], Turner et al. [30], and Hee et al. [11] describe gameful platforms for data science education that use skill trees to organize the lessons into a logical progression. Following the approach of using skill trees to organize task completion, Fuß et al. [10] describe a gameful system that groups related tasks with similar topics into lessons, then combines lessons into skills, which are then organized as a skill tree. Likewise, Barata et al. [2] used a skill tree to organize thematic tasks, which would earn students experience points (XP) upon completion. Regarding the use of skill (or competence) trees as open learner models, a few works [5, 7, 14, 20] describe hierarchical displays or pre-requisites views, which resemble skill trees.

Our approach differs from these prior works because we do not use the skill tree to organize lecture content, to motivate students to complete additional tasks, or to display a hierarchical structure of the learning topics. Instead, we use it to structure assessment and feedback regarding the skills needed to complete the programming assignments in a university course.

Self-Assessment and Open Learner Models

Research has shown that self-assessment improves student learning [25] and "is considered one of the most important skills that students require for effective learning and for future professional development and life-long learning" [27].

There are many ways to implement self-assessment of student practical work. The approach that we propose shares some similarities with the combined use of self-assessment and open learner models. For example, Long and Aleven [19] allowed student to self-assess their skills before displaying the tutoring system's OLM. Mitrovic and Martin [22] allowed students to inspect their OLM so they could self-assess their progress and choose the next tasks to solve. Another approach is that of persuadable OLMs [4, 21]. In this case, if the student does not agree with the assessment provided by the OLM, they can request a modification.

Our approach is similar to these prior works in the sense that it allows students to modify their self-assessed grades in the learner model (represented by the skill tree in our work). However, while previous works focused on letting students negotiate the values provided by intelligent tutoring systems or other classroom assessments, our work is focused on letting students self-assess their programming assignments on a computer science course.

SKILL TREE DESIGN AND IMPLEMENTATION

We implemented our gameful design in a third-year User Interfaces course of the Computer Science undergraduate program at the University of Waterloo during the Spring 2017 term (May–August 2017). Students spent the majority of the course learning how to implement user interfaces, with some course time dedicated to issues of design and usability. They were tasked with completing two major programming assignments and one small programming exercise:

A1: Implementing an interactive side scrolling game with a level editor in Java. This was the largest assignment, which consisted of three parts to be delivered on separate dates: (1) user interface design wireframes; (2) basic user interface and gameplay; and (3) level editor.

A2: Implementing a small animation with a timer in Java. This assignment was introduced after A1, but students were expected to work on it in parallel and complete it before finishing A1.

A3: Implementing a web client to retrieve information from an open data API using jQuery and AJAX.

We introduced gamification into the course by using a skill tree to mediate assignment assessment instead of plain numerical grades as is the common practice at our university. The skill tree was implemented as a small web application that was used by students, teaching assistants, and the instructor. It represented the skills that students were supposed to develop while working on the three programming assignments.

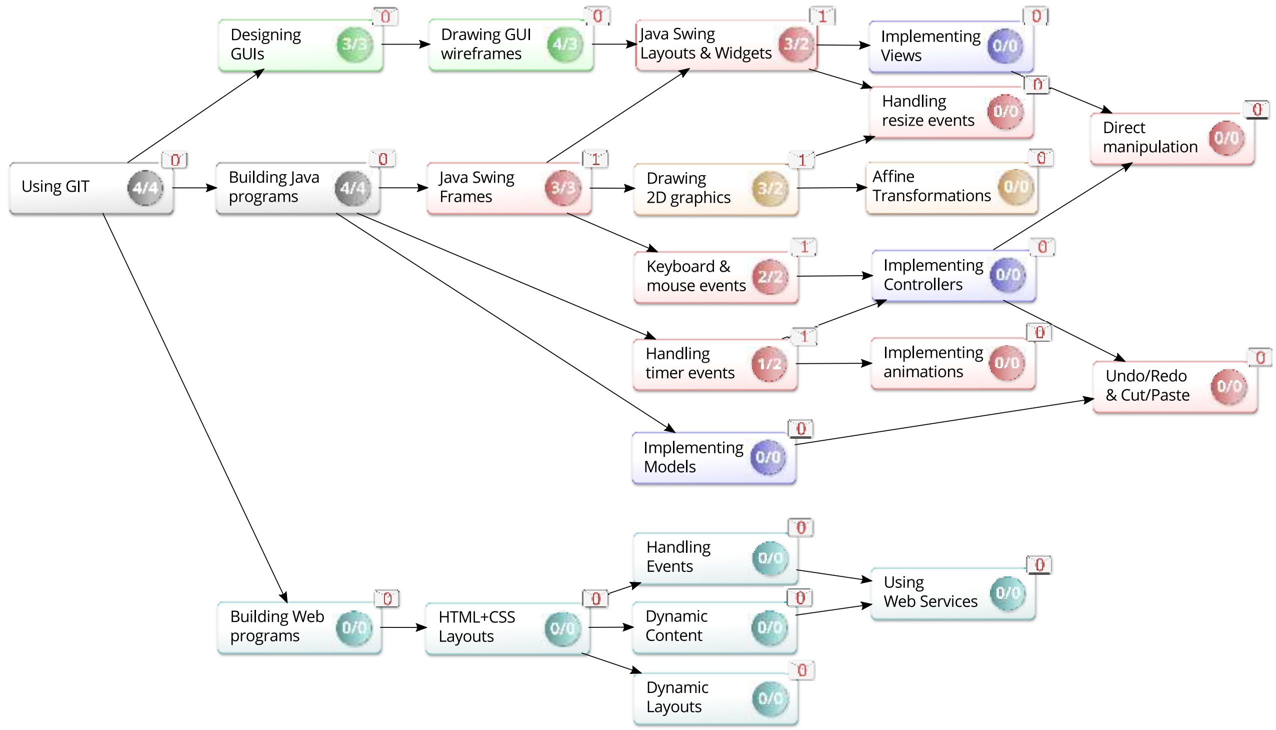

Figure 1. Skill tree used in the course (partly filled example).

Figure 1 presents an example of a partly filled skill tree used in the course. The precedence relationship between skills represented a suggested path for students to take while studying and working on the assignments. The colours represented different types of skills: grey for basic programming skills, green for design skills, red for Java Swing skills, orange for Java drawing skills, blue for the model-view-controller pattern, and aqua for web programming skills. The numbers in the circles represented the student's completed proficiency on each skill and were updated separately by the student (self-evaluation) and by the graders throughout the term. The first number represented the student's self-evaluated proficiency, whereas the second number represented the grader's evaluation of their work. For example, a "3/2" meant that students evaluated themselves at level 3 of proficiency, but the grader had evaluated them at level 2 out of 4 levels. Proficiency levels ranged from **0** (the student has not demonstrated any skill yet) to **4** (the student has achieved the top skill level expected for the end of the course). The numbers in the mail icons (top-right corner of each skill box) showed how many new evaluations from the graders were available for each skill and were updated as the graders registered new assessments.

By clicking on each skill, students and graders had access to a detailed information page. Students could update their proficiency for each skill at any time, by providing their current proficiency level and a free-text comment. They were also allowed to resubmit previously submitted work at any time. The goal of this practice was to allow students to focus on learning new skills more than on having to do everything perfectly the first time to receive good grades. Therefore, they could do their best work the first time and submit it for evaluation, then reflect on the feedback and fix their mistakes for improved learning and grades. Likewise, graders could update their evaluation of the students' skills at any time by providing a proficiency level and a free-text comment. Grading was carried out by six graduate students appointed by the course administrators as teaching assistants (TAs). The students' final assignments grade at the end of the course was calculated as the percentage of the skill tree that they had completed in the graders' evaluations. There were 23 skills in total with four levels each, thus adding up to 92 potential levels to be completed. Together, the three assignments accounted for 40 per cent of their final grade in the course; the remaining 60 per cent of the grade were distributed between two written exams.

Situating our gameful course design in the classifications proposed by the literature, our approach is an example of *structural gamification* according to Kapp [13] because we did not modify the content of the assignments with gamification, we merely provided a structure around it to improve grading and feedback. Considering Landers's theory of gamified learning [17], our gameful skill tree is supposed to affect learning via *mediation* because it was designed to encourage behaviours that could potentially improve learning: self-evaluation and continuous improvement. Moreover, our skill tree implemented the following design principles identified by Dicheva et al. [9]: *progress*, *feedback*, *accrual grading*, *visible status*, and *freedom to fail*. Finally, according to Taras's classification of self-assessment models [27], our implementation is an example of a *standard model*, in which learners use criteria to judge and grade their work prior to submission, then graders mark their work in the usual way while also providing feedback regarding the students' self-assessments.

PILOT EVALUATION

Participants

The course had 159 registered students split into two sections. While all of them used the skill tree to submit and receive feedback for their assignments, participation in the study was voluntary and involved only a feedback survey. We sent an invitation by e-mail on the week after all the assignments had been submitted inviting all registered students to participate in the study by filling out an online form. Because the first author was the course instructor, the invitation was sent by a third party who was not involved in the course or the research; nevertheless, the invitation clearly stated the name of the researchers responsible for the study. Furthermore, students were assured that their participation would be anonymous and that the researchers would only access their responses after the course grades had been finalized. These measures were required by the ethical guidelines adopted by our institution and were intended to assure students that neither their decision to participate (or not), nor the answers they would provide, could affect their grades in any way. Participants did not receive any compensation. The study was reviewed and approved by the University of Waterloo Office of Research Ethics.

In total, 16 students answered the online survey (14 men), aged between 20 and 24 years old. The low response rate is not unexpected because students did not receive any incentive to participate and the invitation was sent only by e-mail from a third party unknown to them, without any mention or incentive for participation during presential lectures (as explained above, to avoid students feeling that their decision to participate could affect their grades).

Procedure

After following the link provided in the recruitment e-mail, participants were asked to complete an online informed consent form. The research was presented as a study to understand students' impressions of the gameful elements used in the course. The course included one lecture about gamification; thus, we can assume that the students were familiar with the term. In addition to demographic information (gender and age), the survey included the following questions:

Q1: What was your general impression regarding the skill tree system used in the course? (free-text)

Q2: How would you rate your overall experience with the skill tree system? (5-point Likert scale with a free-text comment)

Q3: How would you rate the experience of self-evaluating your skills? (5-point Likert scale with a free-text comment)

Q4: How would you rate the experience of receiving feedback from the markers via the skill tree? (5-point Likert scale with a free-text comment)

Q5: In comparison with other courses which you have taken at the University of Waterloo, which used a numeric grade system for assignments, how would you rate your preference? (5-point Likert scale only)

Q6: How much of the skill tree have you completed? (selection list with options corresponding to 10% ranges)

Q7: Would you like to make any additional comments or suggestions regarding the skill tree? (free-text)

RESULTS

Table 1 shows how many participants answered with each rating for the questions with Likert scales in the survey. Furthermore, all participants reported having completed at least 70% of the skill tree in response to **Q6**: 70-79%: 2; 80-89%: 8; 90-99%: 2; 100%: 2; N/A: 2; but we had no means of checking if their self-report was accurate because participation was anonymous.

Experience	**Q2** (overall)	**Q3** (self-evaluation)	**Q4** (TA feedback)
Very positive	2	3	3
Positive	9	6	7
Neutral	2	4	5
Negative	2	2	1
Very negative	1	1	0

Preference	**Q5** (general preference)
Strongly prefer a skill tree	5
Slightly prefer a skill tree	7
Strongly prefer numerical grades	3
N/A	1

Table 1. Number of participant responses for each rating.

Additionally, we read participants' responses to the free-text questions to understand their general impressions and the reason for their ratings. We summarize their answers in the following subsections. Due to the small sample size, we were able to include at least a partial quote from all meaningful responses (not all students provided a free-text follow-up to their quantitative answers). When quoting participants' free-text responses, we use the letter "P" followed by the participant's order in the dataset (e.g., P1). Moreover, we classified the free-text responses as positive, neutral, or negative, according to the participant's response to each question in the Likert scale (because each free-text question was a follow-up to a quantitative question, as described in the previous section).

General Impression of the Skill Tree

In answering **Q1**, 10 participants reported a positive impression, noting that the skill tree was "an innovative way to do grading" (P1), "a very unique way to evaluate my own skills and see what skills apply to which assignment" (P4), and a "very useful, transparent marking" (P16). P5 said "I liked the Skill Tree system a lot, since when I update my skills and write how I have achieved a rating for a skill, I actually think what I have done for that skill. It also helps me in thinking what else I can do to make sure I learn as much as possible about a skill. And of course, the feedback from TA's also helped a lot." Similarly, P15 said "Good and unique. It helped me clearly understand where my strengths and weaknesses were regarding the course content."

On the other hand, four participants reported a negative impression. P6 said it was an "Interesting idea but seemed a bit vague at times. There is a disconnect between assignment requirements and the point evaluation system of the skill tree. Felt like a separate element rather than something directly connected to the assignments / progression in the course." P11 did not like the fact that "requiring the student to go beyond in order to receive 25% of the marks is a terrible system to mark with," because achieving level 4 of proficiency for many skills required students to implement an enhancement that went beyond the minimal requirements. P12 felt that "it created more work for myself," and P14 felt that "the skill tree to me, was confusing, and I think not needed."

It is also noteworthy that P3 reported liking the skill tree, but added that "sometimes the expectations between 3/4 were too strict and trying for a 4 and almost getting it but not quite would make you get a 3 and not a 4 or anything in between," meaning that sometimes students would try to get the full 4 marks for a particular skill, but the grader thought it was not enough and only rated the student at level 3 for that skill.

Overall Experience with the Skill Tree

Participants who reported positive overall experiences when answering **Q2** said that they "enjoyed seeing my progress and being able to visualize by learning" (P1), that it "motivated me to actively seek and meet the requirements" (P3), that it was an "interesting and helpful visualization of what is being learned" (P7), and that it was "good for enticing me to add features" (P13).

Contrarily, P14 had a negative experience and noted that: "I had a difficult time understanding the connection between what we were asked to do and how it was going to be marked – the communication of that information was unclear to me. As well, it reminded me of mobile games where they have deals like '1530 coins for $4.22', where it is difficult to understand the impact of real life money on in-game currency. The features we were asked to implement did not get us marks, but the 'skills' that were expressed were marked, not to mention having a denominator of 23 skills, and 23*4 marks in total made it hard to gauge progress while working on the assignment."

Even with a positive experience, P1 argued that "sometimes the assignment requirements and the skill tree don't match up." Likewise, P13 had a positive experience, but said that "skills did not always match up to requirements in the assignment." P10 reported a neutral experience and said that "there was sort of a disconnect between the assignment expectations and the skill tree itself (some features in the assignment were not actually graded on the skill tree)." Furthermore, P6 said that the skill tree was a "Neat idea, fun to see how the skills connect. I regret not working on the assignments continuously and thus receiving feedback continuously. Instead I just did everything at once and submitted it." P7 had a positive experience, but thought that the "course was too structured around it." Finally, P10 suggested that "having individual trees for each assignment would be clearer."

Experience of Self-Evaluation

Participants who reported a positive self-evaluation experience when answering **Q3** stated that it "motivated me to check my work that I met the requirements" (P3) and that "there were clear explanations for what was expected, which helped gain an understanding of what the mark would look like" (P10). P1 mentioned that "I felt conflicted about this at first, what if I was too generous or too harsh on my own grades but the clear requirements for each level made this easier." Also, P6 explained that "I just looked at the outline for the points (i.e., 0 - did nothing, 1 - submitted something, 2 - implemented it incorrectly, 3 - implemented it correctly once, 4 - implemented it correctly twice) and submitted the appropriate evaluation."

On the other hand, participants who reported a neutral experience said that "Self-evaluating made it more obvious what I needed to do for each skill. However, since there were requirements listed on every page, it felt unnecessary at times," (P13) and that "It would have been easier to simply check off boxes for features that we did or did not implement" (P14). To the contrary, P11 reported a negative experience and asked "Why is it my responsibility to do my assignments, and mark them? I have no incentive to give myself anything but the highest mark." Additionally, P7 stated "Don't make 4/4 = going over and above. If you do the bare minimum you should still get 100%, if you do extra it should give you extra."

Experience of Receiving Feedback

Participants who reported positive experiences in response to **Q4** said that it was "nice to know exactly what I needed to work on" (P3), "The TA responsible for giving me feedback was very good, gave good advice on how to improve my skill as well as why I deserve a particular rating. That really helped in my improvements" (P5), and "comments were always detailed enough to act on" (P13). On the other hand, P4 stated that the experience was "Generally positive but had some issues. I would meet the mastery requirements for some skills [...] but the TAs would find an issue [...] and not give me full marks for the ones I have fully implemented..."

For the neutral experiences, participants said that "In my case, I never received very meaningful feedback, it was usually of a binary nature, a checkmark of sorts ('yes you submitted this and it's working and looks good')" (P6), "I would rather they just mark my assignments normally" (P11), and "there was very little incentive to finish assignments early" (P14). Contrarily, P9 seemed to have a negative experience due to an issue with the time that it took to receive feedback from the TA: "It also took them 2 weeks to give me feedback".

Additional Comments

In response to **Q7**, some participants made general suggestions for the improvement of the skill tree:

"As mentioned previously a more clear assignment to skill tree skill would be appreciated (e.g., setting menu in A1 is not covered in the skill tree)." (P1)

"Give room for some almost marks between two levels (i.e. 3 and 4) and have some bonus marks the instructor can give for efforts that don't exactly reflect on the skill tree." (P3)

"Skill tree system needs to somehow be worked in with the traditional grading system. Perhaps more gameful elements would make it more interesting / useful. Maybe something like, 'receive X/Y points on these skills to unlock advanced starter code for assignment Z.' Perhaps I would have been more motivated to finish my first assignment at an earlier time if I had some motivation. [...] I would have definitely appreciated if I could unlock a better starter code for the assignment." (P6)

DISCUSSION

In the previous section, we presented the results of an evaluation of the students' experience with the skill tree system implemented in a university-level course on user interfaces, in which the skill tree was used to provide feedback and replace numerical grades for the programming assignments. In this section, we summarize and discuss the findings from this study and the lessons learned.

Overall Experience

The results suggest that most students had an overall positive experience with the skill tree: 11 participants (69%) reported an overall positive experience, 9 participants (56%) reported a positive experience with self-evaluating their skills in the assignments, and 10 participants (62%) reported a positive experience with receiving feedback from the TAs through the skill tree. Additionally, 12 participants (75%) said they would prefer a skill tree grading system in their next course instead of a numerical grading system. These findings suggest that using a skill tree system might be a good idea. However, instructors should take additional precautions to mitigate the negative experience of the students who might not enjoy the skill tree or offer them alternative means of assessment.

Strengths and Weaknesses

By examining participants' qualitative responses, we can identify the strengths and weaknesses of the skill tree design we employed in this study. The following aspects worked well:

- Students enjoyed that the skill tree was an "innovative" and "transparent" way of grading assignments.
- Students could better understand how they were learning skills that they could use to implementing each assignment.
- The skill tree helped students grasp their progress and work to meet all the requirements for a full grade.
- Students could understand if their work met the requirements and could have a good idea of what their grades would be once they completed their self-evaluation.
- The explanations given about how to self-evaluate each skill seem to have worked well for most participants who reported a positive experience.
- When the TAs gave students clear and detailed explanations about what they could do to improve their implementations, students appreciated and acted on the feedback.

To the contrary, the following aspects did not work as intended:

- There was a disconnection between the skills and the assignment requirements. The instructor and TAs had a table in which they established a relationship between programming requirements and the skills they would provide; however, this information was not disclosed to students. It would have been better if the skill tree provided a clear information about what programming requirements were associated with each skill. The requirements can be presented as a checklist within each skill.
- There were no grades in between proficiency levels 3 and 4. Thus, students who tried to do extra work to obtain the latest level, but failed for any reason, ended up not getting anything for their effort. It would have been better if the levels were more granular (for example, with 10 levels instead of 4), so that students could be better rewarded.
- Some students disliked that completing the minimum requirements would grant them level 3 for most skills (a grade of 75%), so they needed to implement enhancements of their choice to get 100%. However, this was a common practice for this course in prior terms, which was only made more apparent by the skill tree. Instructors could take this opportunity to help students understand that this gives them some flexibility to implement what they want to get full grades instead of giving them only fixed requirements.
- Having 23 skills with four levels each resulted in a total of 92 skills points. It would be better to have a number of skills and levels that lead to a round number (like 100) so that students can easily understand how much each point earned in the skill tree will represent to their final grade.
- Having just one large skill tree for the three assignments made it hard for students to separate the skills related with each assignment.
- Some students misused the possibility of resubmitting their projects, by delivering incomplete code initially and completing it later, whereas it was intended for students to use the feedback to improve their initial work, thus leading to improved learning. Thus, the freedom to resubmit improved projects needs to be better discussed at the beginning and safeguards must be used to avoid students abusing it.
- Some students did not understand why they had to evaluate their own work. This shows that the benefits of self-evaluation must be better discussed at the beginning of the term to help students understand why it is an important learning activity and a valuable ability for them to develop.
- The TAs gave feedback with different levels of detail, possibly due to their different time commitments and availability. Some students felt that the TAs' comments were not detailed or timely enough to help them. Thus, it is important to guarantee that feedback will be timely and that the amount and style of feedback given by the graders will be consistent.

Lessons Learned

Considering the findings from this study, we learned the following lessons:

- The tree must provide clear descriptions of each skill and what students are expected to accomplish.
- The tree must provide a clear mapping between assignment requirements and skills.
- The evaluation must use a sufficient grading granularity to adequately reward students for their efforts. Also, there must be a way of rewarding students for extra efforts that were not covered by the skills.
- It is better to use a number of skills and grading levels that allow students to clearly understand how much each skill point will represent to their final grade.
- It is better to clearly identify each different assignment on the skill tree or use a different tree for each assignment.
- Instructors should create ways to mitigate the negative experience or allow students who do not feel comfortable with the skill tree to be graded using a traditional method.

In addition, some lessons learned from our experiment are more related to the experience of self assessment rather than the skill tree, and echo best practices already identified in the education literature. However, we include them here for completeness, and to make them easily available for educators following the design concept proposed in this work.

- At the beginning of the course, it is beneficial to discuss the benefits of self-evaluation and of improving one's work from the instructor feedback for the students' learning and development of important life skills.
- If resubmission of improved work is allowed, safeguards should be employed to avoid misuse of this freedom and prevent students from submitting incomplete initial work.
- It is important to ensure that all graders will provide timely feedback, which is detailed enough to help students improve their work. Moreover, graders should be available to respond to students' questions as needed.

Limitations and Future Work

Despite the small sample size, the free-text responses provided a rich source of data that allowed us to interrogate, and posit reasons for, the students' positive or negative experiences, congruent to our intentions for this study. Nevertheless, although the majority of participants in the cohort reported positive experiences, our sample might have been affected by a self-selection bias because students who had a positive experience might have been more inclined to participate in the study. Furthermore, some of the concerns described by the students represent factors that are not directly related to the skill tree, such as the quality and timeliness of TA feedback. Future studies should try to better control for these factors.

Moreover, the course used in this study was heavily based on a set of large programming assignments, which could be mapped to a skill tree. Future studies will need to investigate if this approach can be generalized to different styles of courses and courses in different disciplines.

Additionally, we evaluated the skill tree isolated in our course, so we could gather students' feedback about this element specifically. However, future works could combine it with other design elements, such as badges to reward the student's progress or unlockable content based on skill tree completion, to design a comprehensive gameful learning experience.

CONCLUSION

In this paper, we presented a novel design and implementation of a digital skill tree as a mediator of self-evaluation, feedback, and grading of assignments in higher education. The results showed that the experience was generally positive. Therefore, the digital system we described can be used to mediate the communication between graders and students in a gameful way, particularly for larger classes, when communication in person with all students is not viable. The lessons learned that we presented can guide educators who are willing to implement this design idea. They can also help researchers devise new ways to combine the digital skill tree with additional gameful design elements for a more complete gameful learning experience.

ACKNOWLEDGMENTS

We would like to thank all the students who took the User Interfaces course at the University of Waterloo on the Spring 2017, especially those that contributed with suggestions and the 16 anonymous participants of this study. We also thank the teaching assistants who carried out the work of grading and providing feedback to students. Moreover, we thank Marcela Bomfim for helping with participant recruitment.

This work was supported by the CNPq, Brazil; SSHRC [895-2011-1014, IMMERSe]; NSERC Discovery [RGPIN-2018-06576]; NSERC CREATE SWaGUR; and CFI [35819].

REFERENCES

1. Paul E. Anderson, Thomas Nash, and Renée McCauley. 2015. Facilitating Programming Success in Data Science Courses Through Gamified Scaffolding and Learn2Mine. In *Proceedings of ITiCSE '15*. ACM, 99–104. DOI: http://dx.doi.org/10.1145/2729094.2742597
2. Gabriel Barata, Sandra Gama, Joaquim Jorge, and Daniel Gonçalves. 2017. Studying student differentiation in gamified education: A long-term study. *Computers in Human Behavior* 71 (June 2017), 550–585. DOI: http://dx.doi.org/10.1016/j.chb.2016.08.049
3. Simone de Sousa Borges, Vinicius H. S. Durelli, Helena Macedo Reis, and Seiji Isotani. 2014. A Systematic Mapping on Gamification Applied to Education. In *Proceedings of SAC '14*. ACM, 216–222. DOI:http://dx.doi.org/10.1145/2554850.2554956
4. Susan Bull, Blandine Ginon, Clelia Boscolo, and Matthew Johnson. 2016. Introduction of learning visualisations and metacognitive support in a persuadable open learner model. In *Proceedings of LAK '16*. ACM, 30–39. DOI:http://dx.doi.org/10.1145/2883851.2883853
5. Susan Bull, Matthew. D. Johnson, Mohammad Alotaibi, Will Byrne, and Gabi Cierniak. 2013. Visualising Multiple Data Sources in an Independent Open Learner Model. In *Artificial Intelligence in Education. AIED 2013. LNCS 7926*. Springer, Berlin, Heidelberg, 199–208. DOI: http://dx.doi.org/10.1007/978-3-642-39112-5_21
6. Susan Bull and Judy Kay. 2010. Open Learner Models. In *Advances in Intelligent Tutoring Systems. Studies in Computational Intelligence, vol 308*. Springer, Berlin, Heidelberg, 301–322. DOI: http://dx.doi.org/10.1007/978-3-642-14363-2_15
7. Ricardo Conejo, Monica Trella, Ivan Cruces, and Rafael Garcia. 2012. INGRID: A Web Service Tool for Hierarchical Open Learner Model Visualization. In *Advances in User Modeling. UMAP 2011. LNCS 7138*. Springer, Berlin, Heidelberg, 406–409. DOI: http://dx.doi.org/10.1007/978-3-642-28509-7_38
8. Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart E Nacke. 2011. From Game Design Elements to Gamefulness: Defining "Gamification". In *Proceedings of the 15th International Academic MindTrek Conference*. ACM, Tampere, Finland, 9–15. DOI: http://dx.doi.org/10.1145/2181037.2181040

9. Darina Dicheva, Christo Dichev, Gennady Agre, and Galia Angelova. 2015. Gamification in Education: A Systematic Mapping Study. *Journal of Educational Technology & Society* 18, 3 (2015), 75–88. http://www.jstor.org/stable/jeductechsoci.18.3.75

10. Carsten Fuß, Tim Steuer, Kevin Noll, and André Miede. 2014. Teaching the Achiever, Explorer, Socializer, and Killer – Gamification in University Education. In *Games for Training, Education, Health and Sports: GameDays 2014. Proceedings*, Stefan Göbel and Josef Wiemeyer (Eds.). Springer International Publishing, 92–99. DOI: http://dx.doi.org/10.1007/978-3-319-05972-3_11

11. Kim Hee, Roberto V. Zicari, Karsten Tolle, and Andrea Manieri. 2016. Tailored Data Science Education Using Gamification. In *Proceedings of CloudCom 2016*. IEEE, 627–632. DOI: http://dx.doi.org/10.1109/CloudCom.2016.0108

12. Kai Huotari and Juho Hamari. 2017. A definition for gamification: anchoring gamification in the service marketing literature. *Electronic Markets* 27, 1 (2017), 21–31. DOI: http://dx.doi.org/10.1007/s12525-015-0212-z

13. Karl M. Kapp. 2012. *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education*. Pfeiffer, San Francisco, CA.

14. Judy Kay, Z. Halin, T. Ottomann, and Z. Razak. 1997. Learner know thyself: Student models to give learner control and responsibility. In *Proc. of International Conference on Computers in Education*. 17–24.

15. Ana Carolina Tomé Klock, Aline Nunes Ogawa, Isabela Gasparini, and Marcelo Soares Pimenta. 2018. Does gamification matter? A systematic mapping about the evaluation of gamification in educational environments. In *Proceedings of SAC 2018: Symposium on Applied Computing*. ACM, 2006–2012. DOI: http://dx.doi.org/10.1145/3167132.3167347

16. Jonna Koivisto and Juho Hamari. 2019. The rise of motivational information systems: A review of gamification research. *International Journal of Information Management* 45 (Apr 2019), 191–210. DOI: http://dx.doi.org/10.1016/J.IJINFOMGT.2018.10.013

17. Richard N. Landers. 2014. Developing a Theory of Gamified Learning. *Simulation & Gaming* 45, 6 (2014), 752–768. DOI: http://dx.doi.org/10.1177/1046878114563660

18. Haksu Lee and Young Yim Doh. 2012. A Study on the Relationship between Educational Achievement and Emotional Engagement in a Gameful Interface for Video Lecture Systems. In *2012 International Symposium on Ubiquitous Virtual Reality*. IEEE, 34–37. DOI: http://dx.doi.org/10.1109/ISUVR.2012.21

19. Yanjin Long and Vincent Aleven. 2017. Enhancing learning outcomes through self-regulated learning support with an Open Learner Model. *User Modeling and User-Adapted Interaction* 27, 1 (Mar 2017), 55–88. DOI: http://dx.doi.org/10.1007/s11257-016-9186-6

20. Andrew Mabbott and Susan Bull. 2004. Alternative Views on Knowledge: Presentation of Open Learner Models. In *Intelligent Tutoring Systems. ITS 2004. LNCS 3220*, J.C. Lester, R.M. Vicari, and F. Paraguaçu (Eds.). Springer, Berlin, Heidelberg, 689–698. DOI: http://dx.doi.org/10.1007/978-3-540-30139-4_65

21. Andrew Mabbott and Susan Bull. 2006. Student Preferences for Editing, Persuading, and Negotiating the Open Learner Model. In *Intelligent Tutoring Systems. ITS 2006. LNCS 4053*, M. Ikeda, K. D. Ashley, and TW. Chan (Eds.). Springer, Berlin, Heidelberg, 481–490. DOI: http://dx.doi.org/10.1007/11774303_48

22. Antonija Mitrovic and Brent Martin. 2007. Evaluating the Effect of Open Student Models on Self-Assessment. *International Journal of Artificial Intelligence in Education* 17, 2 (2007), 121–144.

23. David J. Nicol and Debra Macfarlane-Dick. 2006. Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Studies in Higher Education* 31, 2 (2006), 199–218. DOI: http://dx.doi.org/10.1080/03075070600572090

24. Paul R. Pintrich and Akane Zusho. 2007. Student Motivation and Self-Regulated Learning in the College Classroom. In *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, Raymond P. Perry and John C. Smart (Eds.). Springer Netherlands, Dordrecht, 731–810. DOI: http://dx.doi.org/10.1007/1-4020-5742-3_16

25. Philip M. Sadler and Eddie Good. 2006. The Impact of Self- and Peer-Grading on Student Learning. *Educational Assessment* 11, 1 (2006), 1–31. DOI: http://dx.doi.org/10.1207/s15326977ea1101_1

26. Katie Seaborn and Deborah I. Fels. 2014. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies* 74 (2014), 14–31. DOI: http://dx.doi.org/10.1016/j.ijhcs.2014.09.006

27. Maddalena Taras. 2010. Student self-assessment: processes and consequences. *Teaching in Higher Education* 15, 2 (2010), 199–209. DOI: http://dx.doi.org/10.1080/13562511003620027

28. Gustavo F. Tondello, Alberto Mora, and Lennart E. Nacke. 2017a. Elements of Gameful Design Emerging from User Preferences. In *Proceedings of CHI PLAY ’17*. ACM, 129–142. DOI: http://dx.doi.org/10.1145/3116595.3116627

29. Gustavo F. Tondello, Rina R. Wehbe, Rita Orji, Giovanni Ribeiro, and Lennart E. Nacke. 2017b. A Framework and Taxonomy of Videogame Playing Preferences. In *Proceedings of CHI PLAY ’17*. ACM, 329–340. DOI: http://dx.doi.org/10.1145/3116595.3116629

30. Clayton A. Turner, Jacob L. Dierksheide, and Paul E. Anderson. 2014. Learn2Mine: Data Science Practice and Education through Gameful Experiences. *International Journal of e-Education, e-Business, e-Management and e-Learning* 4, 3 (2014), 243–248. DOI: http://dx.doi.org/10.7763/IJEEEE.2014.V4.338