Exploring the Role of AI-Generated Feedback Tangential to Learning Outcomes

Steven C. Sutherland
University of Houston-Clear Lake
sutherland@uhcl.edu

Tiago Machado
Northeastern University
tiago.machado@nyu.edu

Shruti Mahajan
Worcester Polytechnic Institute
spmahajan@wpi.edu

Omid Mohaddesi
Northeastern University
mohaddesi.s@northeastern.edu

Camillia Matuk
New York University
cmatuk@nyu.edu

Gillian Smith
Worcester Polytechnic Institute
gmsmith@wpi.edu

Casper Harteveld
Northeastern University
c.harteveld@northeastern.edu

2023 IEEE Conference on Games (CoG) | 979-8-3503-2277-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/CoG57401.2023.10333239

***Abstract*—Students are often tasked in engaging with activities where they have to learn skills that are tangential to the learning outcomes of a course, such as learning a new software. The issue is that instructors may not have the time or the expertise to help students with such tangential learning. In this paper, we explore how AI-generated feedback can provide assistance. Specifically, we study this technology in the context of a constructionist curriculum where students learn about experimental research through the creation of a gamified experiment. The AI-generated feedback gives a formative assessment on the narrative design of student-designed gamified experiments, which is important to create an engaging experience. We find that students critically engaged with the feedback, but that responses varied among students. We discuss the implications for AI-generated feedback systems for tangential learning.**

***Index Terms*—creativity support, automated feedback, constructionist learning, game design**

I. Introduction

Game-based constructionist approaches to education have students make games (both analog and digital) in order to learn about the underlying topics that their game is about [1]. Students learn not only through playing an educational game, but through creating a game that deeply engages with the course material. The benefits and effectiveness of game-based learning or "constructionist gaming" are well explored in several fields [2]–[5]. This approach is especially relevant when designing curricula that integrate computer science learning into existing mathematics and science classes, as game programming is frequently used to introduce students to computer science [6]–[8].

For students to achieve the full potential of a constructionist game-based curriculum, they should be making games that are interesting enough for others to play and that meet student goals [9]. Students therefore need guidance in designing and developing games, especially where game design is not an educational goal for the course. Teachers may not necessarily have the expertise, time, or resources to provide such guidance.

This material is based upon work supported by the National Science Foundation under Grant No. 1736185 and No. 2142396.

To reduce the game design burden and address the expertise requirements, educators commonly turn to game design tools intended for novice designers so that students can create games while focusing sufficiently on other aspects of the course objectives. Popular tools include visual programming environments like *Scratch* [10] and interactive narrative tools, such as *Twine* [11] and *Ren'Py* [12]. These tools, though approachable and usable, still do not explicitly support students in following a reflective and iterative game design *process*, as is considered best-practice in game design education [13].

One potential solution to this problem is the incorporation of AI-generated feedback that can foster students to *tangentially* learn design concepts that can make their projects more enjoyable. Providing AI-based game design support, and how designers interact with AI systems during design, is an active area of research, including automated playtesting [14], [15] and co-creative systems [16]–[18]. Application of these techniques in educational contexts is underexplored, despite the potential to improve the experience that students have when designing games while also improving the outcomes of this learning process.

In this paper, we report on an exploratory effort of the effect of AI-generated *tangential feedback* on student behavior, where feedback targets learning outcomes (game design) not directly tied to the course's learning objectives (research design). The core question we aimed to address is: *Do students develop tangential learning behavior when exposed to an AI-generated feedback system?*

II. Related Work

A. Game-Based Constructionist Education

Constructionist learning theory provides the fundamentals for pedagogical approaches of learning through design [19]. According to constructionist principles, learning happens more effectively when creating public artifacts (e.g., computer programs, robotics constructions, or writing) [20]. Constructionist approaches put learners in the role of communicators, and not just recipients of knowledge; and of producers, and not just consumers of media [21]. They also draw on constructivism in the sense that they involve activities that engage students in a process of building knowledge [22]. Researchers indicate that under these circumstances, individuals learn how to ask meaningful questions, conduct investigations, reason about data, and apply what they learn in future situations [23]. Games gained attention from researchers because of how they attract and retain learners' engagement [24], and for their potential to embody constructionist ideas [1], [25].

The key issue for constructionist learning is broadly that creating artifacts is time-consuming and difficult [26], with a lot of time spent on topics that may be incidental to the curriculum material (e.g., programming and game design). In particular, creating effective games requires expert knowledge of game elements and mechanics, and of how to make use of these to produce desired player outcomes. To help our students create their gamified research, we considered the use of a *creativity support* tool (our feedback system) that embeds a theory of interactive narrative [27] to assist students in acquiring skills on how to design a gamified narrative experiment. This enables what we describe as *tangential learning*: the learning of content that is incidental or indirectly related to the curriculum.

B. AI Feedback in Education

The use of AI for the design of feedback systems came from the necessity researchers and educators have in providing smart interventions. More than quizzes and assessments, feedback systems powered by AI can attend to more students, work in real-time, and go over details that would be practically impossible for an instructor [28].

These necessities are clarified in other research. Trajkova [29] explains the design of an AI-based feedback system that deals with the issues ballet students have in acquiring proper feedback. Feedback from teachers and peers are a requirement to facilitate the acquisition of such skills. However, ballet classes have dozens of students per instructor. Because the instructor cannot attend to every student at the same time, the amount of information each student has may vary greatly.

The amount of information and the personalization level that feedback can have is explored by Prajapati et al. [30]. In this work, the authors explore the communication students try to convey through sketches. Although considered as a critical skill, novice designers usually struggle to get fast and appropriate feedback from traditional learning methods, which brings limitations such as: few instructors for many students, limited time to attend to the students, and general information that does not address all the students' needs. The authors then developed a machine learning tool that compares students sketches to others in a data-set offering a visual correction assistant that teaches students how to avoid ambiguity.

Mirchi et al. [31] stated that the lack of transparency in AI-generated feedback systems can lead students to frustration and discomfort. To address this issue, the authors developed performance metrics of psycho-motor skills of students performing virtual surgeries. Then, they designed a new VR simulation tool that applied AI to calculate the students performance in real-time and provided feedback based on metrics coming from other students.

Vittorini et al. [32] designed an automatic grading system for an online data science course. Beyond offering accurate grades and reduce correction time, the system provided qualitative feedback to the students in natural language. Feedback was implemented based on Natural Language Processing (NLP) and summarized codes written in the R language to give short descriptions of the student progress in a given exercise. Students who took the course and used the feedback system got higher grades than those who did not use it.

III. Automated Narrative Feedback System

We adopted an automated narrative analysis system developed by Majahan et al. [33], which is built on top of StudyCrafter [34] and generates a feedback report[1]. StudyCrafter is a platform that enables users to create, play, and share interactive projects. Users can design various genres (e.g., action, puzzle, platforming), but its affordances support primarily the creation of interactive narratives [35]. A StudyCrafter project consists of scenes and scripts. Scenes are visual layouts in which users can put characters and objects. The scripts are used to control the behavior of elements in the scene and the whole flow of all the scenes in the project. Therefore, narrative feedback is provided after an analysis of these two elements (scenes and their scripts), see Figure 1.

A. Unified Graph Form

The scripts in StudyCrafter are a series of visual nodes connecting each other in the form of a connected graph. Therefore, the first step of the Narrative Feedback System is to create a representation that unifies all the scripts (that control the behavior of the scenes) in a single unified graph. This unification is done by using a graph-search based algorithm such as Depth-First or Breadth-First search. In the process, all the specificities of the nodes are captured. Some represent dialogues, some represent interaction opportunities, feedback to the player, and so on. Finally, the special node that represents the end of a scene is connected to the start node of the next scene, forming the intended graph unification.

B. Metrics

The Unified Graph Form is a mechanism that allows the Narrative Feedback System to properly calculate the metrics in which the feedback is based on. These metrics are divided in three factors: Narrative Structure Complexity, Interaction Affordances, and Interaction Point Affordances.

- **Narrative Structure Complexity** - This metric calculates the degree a script branches and loops and the average length of traversals inside these branches and loops. Remember that a script is essentially a graph and the basis of these measures comes from graph theory algorithms [36]–[38]. The goal is to present a measure for design complexity, which does not strictly relate to the level of challenge a player experiences or the narrative content.

[1]For consistency, we refer to this system as the "feedback system".

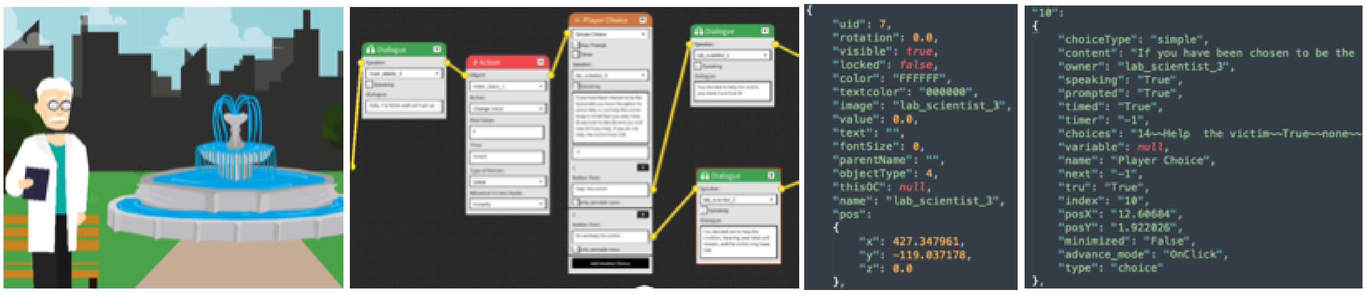

Fig. 1. From left to right: scene layout of a student project, script nodes in the StudyCrafter editor, description code for the scene layout, and the description code for the script nodes. The codes are automatically generated and form the input for the feedback system.

- **Interaction Affordances** - This metric considers the possibilities a player has and how it will affect the narrative progression. It is based on theoretical agency [36], [39], [40], which is the effects of player's actions without measuring their subjective experience of agency. Szilas and Ilea's playthrough-metrics [41] are the major influence for the single metric in this factor. The Narrative Feedback System uses simulated randomized playthroughs. These simulations run inside the graphs, simulating a player using a random choice policy. After several simulations, an average of the possible interactions is calculated.
- **Interaction Point Affordances** - Metrics in this factor are based on the theory of interaction structure and feedback [36], [42] and reveal how rich a scenario is in feedback to the player in two contexts: dependent and independent of player actions.

The feedback is formulated based on how a project scores on the metrics across the three factors mentioned above. The values for the project's metrics are situated inside a repository of previous projects. To understand how a project compares against others, its metrics, after calculated, are classified according to a cluster analysis based on the K-Medoids algorithm. The cluster categorization helps users understand how they are utilizing the metrics and how they are similar and dissimilar to other projects. These additional projects can be consulted by the users through links in the feedback form, and serve as examples to the users. It is important to always have in mind that the metrics do not "compute" narrative content. Aspects like character design or plot twists are out of scope. The metrics do compute narrative design structure, i.e. how linear it is, how it branches, how many loops it has, how fragmented it is. It also computes how this structure offers opportunities for interaction and feedback to the player.

For the final report, instead of showing numbers, users see short texts indicating how their projects are compared to previous ones. For each one of the factors, they receive two examples: a game that has similar metrics scores and another with dissimilar ones, allowing users the opportunity to explore the examples and see how to change their project structure.

C. Artificial Intelligence Clustering

Many applications have been developed in different fields using AI clusters, such as healthcare [43], [44], market segmentation [45], and education [46], [47]. All of them benefit from the elasticity offered by cluster techniques. In the particular case of this paper, the expansion or contraction of groups over time brings the student the opportunity to have a feedback that is "alive" and shows alternatives to explore for the interactive narrative design. The metrics' classification is guided by the K-Medoids cluster, as mentioned in the previous subsection. It is worth noting that the word "metric" does not mean a static evaluation or a deterministic formula that will grade students' narratives. As in other cluster analyses, the classification can change over time [48]. Different groups will emerge (or disappear) as more projects are part of the AI cluster. It is a direct consequence of how students' projects navigate into the three metrics' factors. Far from a "right" and "wrong" grader, the AI feedback acts as an assistant, showing what is possible to do structurally and visually to an interactive narrative, letting the student decide what applies best.

IV. Methods

This study was conducted as part of a module focused on teaching experimental research methods in an introductory game course for first-year graduate game design students. The learning outcomes of this course module are focused on experimental research, i.e. understanding what an experiment is, how to set up an experiment, deal with biases, etc. Each individual student is tasked to create a "gamified experiment" by making use of StudyCrafter, and is asked to make sure the experiment is both engaging for the player to participate in as well as that it has a rigorous research design so the influence of biases and confounds are minimized. Although we are interested in whether tangential feedback from the system may lead to higher achievement on primary learning outcomes, this paper focused only on how and whether students engaged with and implemented the feedback to assist in tangentially-related game design considerations. All materials of this study (Version 1 & 2 & Final, automated feedback reports, reflections, and instructions) are accessible at our repository.[2]

A. Research Design

To explore the role of automated narrative, we conducted a quasi-experiment where one section used the feedback system

[2]https://repository.library.northeastern.edu/sets/neu:h989rz04w

(i.e., the treatment group or T) and another section did not (i.e., the comparison group or C). Use of the feedback system by the treatment group was "forced" and "structured": at set intervals students received feedback.

B. Participants

The treatment group included 15 master students and the comparison group included 12 master students. The students have varying backgrounds: art, design, psychology, computer science, etc. They also vary in their level of programming skills, prior experience with research, and with game design.

C. Materials

Students made use of StudyCrafter (v2.4.1). We further used the feedback system by Mahajan et al. [33] with some minor modifications (e.g., language report, database with projects). For the treatment condition, students were asked to upload their projects to their online project folder. The research team provided feedback reports two times during the study. At the time of this study, the feedback system was external to StudyCrafter; however, it has since been implemented so that students can request the report themselves from inside the game. We provided the report, rather than allowing students to request it, to avoid self selection biases in our study. Once they received the feedback report, students were asked to provide a brief written reflection on how their interpretation of the feedback, where they are with their project, and what they planned to change. For the comparison group, we requested a similar reflection to avoid a possible bias instilled by this request; the instructions for the comparison group were similar except that it excluded the request for an "interpretation of the automated feedback report." Unfortunately, only a few students in the comparison group adhered to this request, most likely because they had less of an incentive to do so.

D. Procedure

Students were first asked to describe their experiment and make a storyboard. After, they were introduced to StudyCrafter and began programming their experiments. Following this, students were asked to upload *Version 1* to their project folder. Students in the treatment group provided a written reflection after looking at the feedback report, which the research team added to their project folders within 24 hours; the students in the comparison group could do this when uploading their Version 1, submitted as a progress check. After iterating their work, this process was repeated: students were asked to upload *Version 2*, and provide a written reflection. At the end, students presented their work and submitted their *Final Version*. As such, students in the comparison group received automated feedback twice: on their Version 1 and Version 2.

E. Data Analysis

We evaluated the metrics generated by the feedback system for the final versions and coded the student reflections.

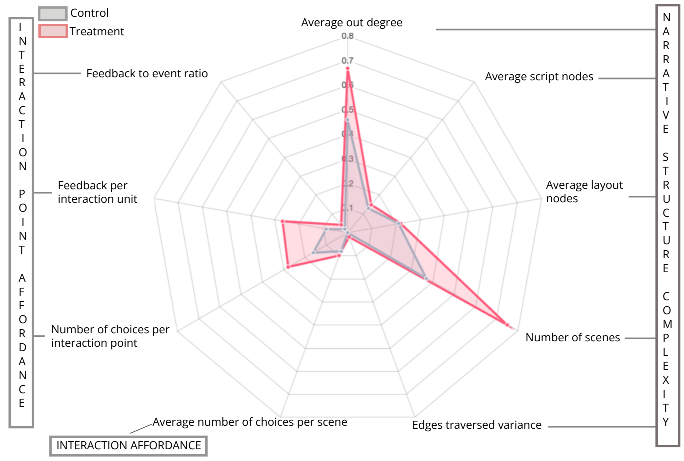

Fig. 2. Radar charts comparing medians between the treatment and control.

1) Metrics: We calculated the metrics (nine in total across three factors) using the same metrics underlying the feedback system. We calculated these metrics to compare the difference between the treatment group and comparison group. Since the metrics had different scales, we re-scaled each metric using the *min-max* normalization for the groups to be compared. For our comparisons, we used the median as a benchmark as this is a measure that is not affected by outliers.

2) Reflections: Using open coding, three raters coded independently the 15 written reflections associated with Version 1 and the 15 written reflections associated with Version 2. We then discussed the findings and established an initial codebook. Following this, we applied and iterated this codebook through consensus coding until we all agreed on the coding [49].

V. Results

The students developed in total 27 individual projects, each with a different topic. Some projects focused on studying real world topics, such as cognitive bias in seeking help from others; others focused on examining game topics, such as what game elements may influence player retention/frustration. We present the results of evaluating these projects using the metrics from the automated narrative feedback system, and compare the results from the treatment group and the comparison group. We also highlight major themes from student reflections related to how they interpreted the automated feedback, and triangulate qualitative findings with the metrics to provide further insight into the participants' behavior.

A. Treatment vs. Comparison Group

The medians of the treatment group were higher for all of the metrics (see Figure 2). Most notable are the values for average out of degree, feedback per interaction unit, and number of choices per interaction point. Average out of degree is a metric of the factor Narrative Structure Complexity, and means that the narrative structure has more branches and thus is not linear.

Feedback per interaction unit and number of choices per interaction point are part of the Interaction Point Affordances and Interaction Affordances factors, respectively. High values in these metrics mean that the projects offer many opportunities of interaction to the player (in the case of number of choices per interaction point), and also that the players will have more feedback about their actions (in the case of feedback per interaction unit). The number of scenes were also notably higher, however we believe this might be due to the fact that students in the treatment group adhered better to include a briefing and debriefing scene.

B. Student Responses

From the metrics analysis, we find that the treatment group was influenced by the feedback system; however, they were more influenced on particular metrics than others. From the student reflections, we found a potential explanation for this influence. We suggest this is a result of *metacognitive thinking*. The limited influence or the influence on specific metrics is a result of the task they received: students had to make *experiment considerations* and that involved ignoring or not addressing the suggested feedback. We describe these emergent themes in more detail below.

Metacognitive Thinking. From the student responses, it was apparent that the feedback report (and the request to reflect on this report) led students to thinking about what they made. Thinking about what they made was in part stimulated by the comparison with other student projects:

> I have since went and looked at that project, and have found it is also not quite a typical StudyCrafter project. — P4

For some it also made them reflect on the nature of the feedback report, which is focused on the narrative aspects of their projects, and what their project is about:

> Because of the unique platforming nature of my project, as opposed to the more interactive story based projects that are typical in StudyCrafter, I have some reservations about exactly how applicable some of this feedback is. — P4

The feedback further made students think about how others may experience their projects:

> I will rethink about the independent variable and choose more suitable items to represent them. These are what I am going to do to make the project engaging. — P8

> Going forward, I need to think about the actual experience of the game and possibly change the scenarios to make them more immersive. — P9

Common to these student reactions is that students engaged with what is called *metacognitive thinking* in education [50]. This form of "thinking about thinking" is critical for successful learning and, thus, educators seek strategies to foster such thinking. Additionally, from a game design perspective it is key to think about the player [51] and the student reflections suggest that students were encouraged to consider them.

Experiment Considerations. The most dominant observation regarding the student reflections pertains to the task of creating an experiment. The feedback system is context-independent, and thus ignores the experiment task, focusing only on the interactive narrative aspects of projects. Various students pointed this out:

> ...my first priority is still [sic] keep the experimental environment clear-cut, efficient and under controlled. — P15

> To be honest, this is not a game, it is an experiment; to make players better understand my research topic and research purpose, I choose [sic] to create everything linearly. However, the current design reduces the pleasure of playing experience; there is almost no interaction between the game and the player except choosing the tutor. — P10

As a result, in response to the report, which often indicated that projects were short, included few points of interaction, etc., students often stated that what they designed was "appropriate" (P10) for an experiment:

> I know my project is really one dimensional and I don't have many functions and actions and such, but I have a lot of options to pick from which forms the basis of my research and as long as I have those I feel like I have completed my objective. — P2

Others were willing to see how they could incorporate the feedback into their projects:

> Under this premise, I will try my best to add more narrative elements and scenes to increase the length of game playing. — P15

> I need to add more interaction between the player and the game. I will create more meaningful interactions [sic] will help participants to understand the contents and more actively join the processes of doing experiment. — P10

Students realized, however, that such integration may come at a *trade-off*. For example, P10 mentioned to "not want the experiment to be a large time commitment for my participants." A solution for resolving the experimental requirements versus the provided feedback was "to add some nonessential choices to the game to make it more interactive so that [is] more like a game" (P12). Thus, given the constraints of the task and the requirements students had in mind for their experiment, they seem to address certain aspects suggested by the feedback system. This explains why only particular metrics change.

1) Student Differences: The two aforementioned emergent themes from the student reflections, meta-cognitive thinking and experiment considerations, help explain the metrics analysis of comparing the treatment group with the comparison group. However, in analyzing the student responses to the feedback system we identified a third theme: *feedback responsiveness*. We find that students differ in how they respond to the feedback. We identified two main student types:

1) *Follower* [4 students]: a student who follows the provided feedback and mentions what changes they are

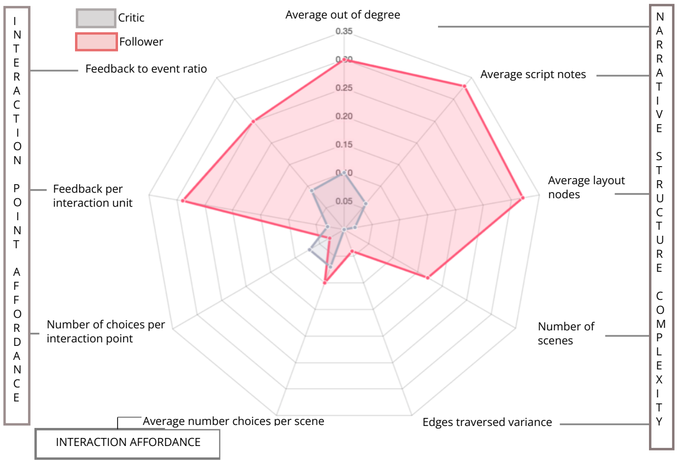

Fig. 3. Radar chart comparing medians for critics vs. followers.

going to make according to the feedback. Followers are explicit about what changes they will be making. For example: "According to the feedback, the first part of the project that needs to be fixed is the length of the narrative structure. I decide [sic] to make more clear explanation of the story and the background of this project." — P1

2) *Critic* [5 students]: a student who is aware of the benefits the feedback is trying to bring to their projects, however criticizes the feedback or has reservations because of the nature of their projects or experimental design considerations. They typically mention they are not going to make any changes. For example: "if the length of scripts and scenes doesn't need to be average, there is nothing to change." — P8

It was remarkable to observe how students stated that based on the "feedback from the machine" (P6) they "have to" (P13) or felt "required to" (P13) make changes, or a suggestion made "needs to be fixed" (P1). One student even ended their reflection with "hopefully it will be approved" (P10). It is remarkable because the instructor explicitly informed students that they are not *required* to consider the feedback and yet some students felt compelled to satisfy or adhere to what the "machine" was telling them. Based on these findings, we revisited the metrics analysis and explored how different types of students were influenced by the feedback system.

C. Metrics per Student Type

Following the identification of the student types, the medians of the metrics revealed that students categorized as followers were, with a few exceptions, considerably higher than the ones for the students categorized as critics (see Figure 3.) In the Narrative Structure Complexity factor, all the metrics had substantially higher medians, indicating that projects from students categorized as followers had more scenes, more nodes, more branches and loops. Consequently, the narratives presented were more fragmented and offer more ways to be followed by the players.

The Interactive Affordance factor has only one metric, average number of choices. Although the median was higher for follower students' projects, the difference was the smallest when we compared all the differences for the metrics. This is indicative that both critics and followers did not explore a balance in the number of choices throughout all the scenes.

The Interaction Point Affordances factor has the only metric whose median from students projects categorized as critics was higher. However, the other two metrics in this factor were considerably higher for followers. Even with number of choices per interaction point being higher in the critics' projects, feedback per interaction unit and feedback per event showed that followers' projects provided more feedback to the player and this feedback was also more balanced according to the player actions.

VI. Discussion

A. Feedback Influence on Student Projects

The students in the treatment group adjusted their projects. For all the metrics, their medians were higher compared to the control group. Overall, their projects were more complex in narrative and dialogue. It seems that students understood the importance of interaction and the need for reaction to players' choices from the feedback report. As a result, students changed their projects to be more interactive and, in many cases, increased the narrative structure complexity, adding more possibilities for players to select. Also, students critically evaluated the trade-offs of the feedback, an educational goal in the design of this kind of system [52]. Students evaluated and, often times, partially adopted feedback due to the constraints of their research experiments. .

We categorized students into two groups: *critics* and *followers*, based on their reflections on the feedback report. Followers appeared to be most influenced by the feedback system. They changed their projects, more strictly following the feedback, as opposed to critics who seem to have fewer changes in their projects and reservations about the feedback.

B. Reliance on AI assistance

The reliance on AI assistance is part of what researchers call algorithmic experience [53] and refers to people's involvement with a technology whose results are no longer static and are based on a series of inputs, rather than on a single user. Algorithmic experience can be observed in how the feedback seemed to foster more activity from the students in the treatment group. This can be noted with students revealing different levels of compliance towards the given feedback. Some participants were more eager to change their projects based on the recommendations provided, while others used them to evaluate the trade-offs of incorporating suggested changes or keeping their initial design. This fluctuation around compliance takes the system out of the danger of being a negative persuasive interface [54] that pushes its users to take extreme decisions, in this case, to totally comply or totally neglect the suggestions [55]. Actually, the results show that the participants reacted in a more reflexive way after getting the feedback, with many of them presenting traits of metacognitive thinking. It shows that the system has traits that belong to the inspirationalist school of creativity because it pushes students to critically explore alternatives [56].

However, it is important to identify particular events in the utilization of the feedback system. For example, it is not clear why a participant stated to seek approval from the system even with instructions that the system was not responsible for approving anything. This reliance needs to be better addressed in order to not escalate in a study with more participants and potentially harm them in taking decisions based on incorrect assumptions. This behavior can be related to misuse or errors in how to present the feedback, in that case, it can be adjusted when the user gets more experienced with the system [57].

C. Pedagogical Considerations

Our results suggest that students are encouraged to change their games more if given automated feedback. It promotes engagement in critical and metacognitive thinking and reflection, which is pointed out as good outcomes in the design of feedback systems [52], [58]. Indeed, applying an automated feedback system, similar to the one in this paper, needs some considerations. To obtain the maximum possible outcome from such a system, instructors need to take on a more active role.

Guidance on how to best utilize the automated feedback is required, especially in circumstances such as the case presented in this paper where the feedback has to be traded off against the assignment. In such cases where an automated feedback system is employed to analyze creative assignments developed by students, creative freedom could be restricted. To avoid this, the instructor must emphasize creative freedom and direct students to use the feedback as a tool.

This emphasis is also important for preventing unnecessary increases in cognitive load, which can make students fail to understand what the feedback is saying [59], [60]. For some unique projects, such as the one in this paper with a platforming nature, the automated feedback of this kind might be less applicable, and as such, the instructor can guide the student to effectively interpret how to utilize the feedback.

Guidance could also be in the form of how to address the feedback. The report provides examples of other projects as recommendations that the students might benefit from playing. However, there are no guidelines on what changes to make and how to make them according to the feedback. Therefore, examining an example of a project developed with consideration of the feedback might be beneficial to students.

D. Limitations

For a typical experimental study, the number of participants per condition is relatively low. However, we harnessed the design of an experiment to systematically explore the influence of automated assistance, and not to find evidence for specific hypotheses. In fact, our work did not present hypotheses for this particular reason. Another issue pertains to the student artifacts. The current paper examines in total 27 student-designed artifacts. Despite that many projects are fairly similar in terms of their structure, in part because of the specific task requirements and in part because of the affordances of StudyCrafter, students had creative freedom and could develop widely diverging projects, illustrated for example by a platforming game (a genre not common to see associated with narrative-based engines such as StudyCrafter). Due to the low number of considered projects, any outlier may still be influential, despite using medians as a measure.

VII. Conclusion

In this paper, we explored how students may be influenced by AI-supported feedback while designing a gamified project. The AI-feedback had the intent to help students in designing more compelling interactive experiences. Our results suggest that when a student is exposed to the feedback, they are more likely to engage in tangential learning behavior. As such, our work suggests that AI-generated feedback can play a meaningful role in helping students with learning activities tangentially related to the learning outcomes of a course.

References

[1] Y. B. Kafai and Q. Burke, "Constructionist gaming: Understanding the benefits of making games for learning," *Educational psychologist*, vol. 50, no. 4, pp. 313–334, 2015.

[2] C. B. Eiben, J. B. Siegel, J. B. Bale, S. Cooper, F. Khatib, B. W. Shen, B. L. Stoddard, Z. Popovic, and D. Baker, "Increased diels-alderase activity through backbone remodeling guided by foldit players," *Nature biotechnology*, vol. 30, no. 2, p. 190, 2012.

[3] J. Cechanowicz, C. Gutwin, B. Brownell, and L. Goodfellow, "Effects of gamification on participation and data quality in a real-world market research domain," in *Proceedings of the first international conference on gameful design, research, and applications*, pp. 58–65, 2013.

[4] S. Kim, K. Song, B. Lockee, and J. Burton, "What is gamification in learning and education?," in *Gamification in learning and education*, pp. 25–38, Springer, 2018.

[5] T. Barko and T. D. Sadler, "Learning outcomes associated with classroom implementation of a biotechnology-themed video game," *The american biology Teacher*, vol. 75, no. 1, pp. 29–33, 2013.

[6] J. Denner, L. Werner, and E. Ortiz, "Computer games created by middle school girls: Can they be used to measure understanding of computer science concepts?," *Computers & Education*, vol. 58, no. 1, pp. 240–249, 2012.

[7] L. Werner, S. Campe, and J. Denner, "Children learning computer science concepts via alice game-programming," in *Proceedings of the 43rd ACM technical symposium on Computer Science Education*, pp. 427–432, 2012.

[8] L. Zhang and J. Nouri, "A systematic review of learning computational thinking through scratch in k-9," *Computers & Education*, vol. 141, p. 103607, 2019.

[9] C. Harteveld, *Triadic game design: Balancing reality, meaning and play*. Springer Science & Business Media, 2011.

[10] M. Resnick, J. Maloney, A. Monroy-Hernández, N. Rusk, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. Silver, B. Silverman, *et al.*, "Scratch: programming for all," *Communications of the ACM*, vol. 52, no. 11, pp. 60–67, 2009.

[11] H. Engström, J. Brusk, and P. Erlandsson, "Prototyping tools for game writers," *The Computer Games Journal*, vol. 7, no. 3, pp. 153–172, 2018.

[12] H.-L. Chen and Y.-C. Chuang, "Using ren'py as a digital storytelling tool to enhance students' learning," in *Multi Conference on Computer Science and Information Systems*, p. 352, 2019.

[13] T. Fullerton, C. Swain, and S. Hoffman, *Game design workshop: Designing, prototyping, & playtesting games*. CRC Press, 2004.

[14] N. Shaker, M. Shaker, and J. Togelius, "Ropossum: An authoring tool for designing, optimizing and solving cut the rope levels," in *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2013.

[15] I. Borovikov, Y. Zhao, A. Beirami, J. Harder, J. Kolen, J. Pestrak, J. Pinto, R. Pourabolghasem, H. Chaput, M. Sardari, *et al.*, "Winning isn't everything: Training agents to playtest modern games," in *AAAI Workshop on Reinforcement Learning in Games*, 2019.

[16] A. Liapis, G. N. Yannakakis, C. Alexopoulos, and P. Lopes, "Can computers foster human users' creativity? theory and praxis of mixed-initiative co-creativity," 2016.

[17] H. Saragih, "Co-creation experiences in the music business: a systematic literature review," *Journal of Management Development*, 2019.

[18] M. Davidovici-Nora, "The dynamics of co-creation in the video game industry: the case of world of warcraft," *Communications & Strategies*, no. 73, p. 43, 2009.

[19] S. Papert, *Mindstorms: Computers, children, and powerful ideas*. Basic Books, 1980.

[20] Y. B. Kafai, "Constructionist visions: Hard fun with serious games," *International Journal of Child-Computer Interaction*, vol. 18, pp. 19–21, 2018.

[21] I. Harel, *Children designers: Interdisciplinary constructions for learning and knowing mathematics in a computer-rich school.* Ablex, 1991.

[22] S. Papert and I. Harel, "Situating constructionism," *Constructionism*, vol. 36, no. 2, pp. 1–11, 1991.

[23] J. L. Kolodner, P. J. Camp, D. Crismond, B. Fasse, J. Gray, J. Holbrook, S. Puntambekar, and M. Ryan, "Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design (tm) into practice," *The journal of the learning sciences*, vol. 12, no. 4, pp. 495–547, 2003.

[24] S. Papert, "Does easy do it? children, games, and learning," *Game Developer*, vol. 5, no. 6, p. 88, 1998.

[25] G. M. Troiano, Q. Chen, Ángela Vargas Alba, G. Robles, G. Smith, M. Cassidy, E. Tucker-Raymond, G. Puttick, and C. Harteveld, "Exploring how game genre in student-designed games influences computational thinking development," in *Proceedings of the 2020 CHI Conference*, pp. 1–13, 2020.

[26] C. Girvan, B. Tangney, and T. Savage, "Slurtles: supporting constructionist learning in second life," *Computers & Education*, vol. 61, pp. 115–132, 2013.

[27] E. Carstensdottir, N. Partlan, S. Sutherland, T. Duke, E. Ferris, R. M. Richter, M. Valladares, and M. Seif El-Nasr, "Progression maps: Conceptualizing narrative structure for interaction design support," in *Proceedings of the 2020 CHI Conference*, CHI '20, (New York, NY, USA), p. 1–13, Association for Computing Machinery, 2020.

[28] H. Yu, C. Miao, C. Leung, and T. J. White, "Towards ai-powered personalization in mooc learning. npj science of learning, 2 (1), 15," 2017.

[29] M. Trajkova, "Designing ai-based feedback for ballet learning," in *Extended Abstracts of the 2020 CHI Conference*, pp. 1–9, 2020.

[30] S. P. Prajapati, R. Bhaumik, T. Kumar, and U. Sait, "An ai-based pedagogical tool for creating sketched representation of emotive product forms in the conceptual design stages," in *ICT Systems and Sustainability*, pp. 649–659, Springer, 2021.

[31] N. Mirchi, V. Bissonnette, R. Yilmaz, N. Ledwos, A. Winkler-Schwartz, and R. F. Del Maestro, "The virtual operative assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine," *PloS one*, vol. 15, no. 2, p. e0229596, 2020.

[32] P. Vittorini, S. Menini, and S. Tonelli, "An ai-based system for formative and summative assessment in data science courses," *International Journal of Artificial Intelligence in Education*, vol. 31, no. 2, pp. 159–185, 2021.

[33] S. Mahajan, L. Bunyea, N. Partlan, D. Schout, C. Harteveld, C. Matuk, W. Althoff, T. Duke, S. Sutherland, and G. Smith, "Toward automated critique for student-created interactive narrative projects," in *Proceedings of the... AAAI Conference on Artificial Intelligence*, 2019.

[34] C. Harteveld, N. Manning, F. Abu-Arja, R. Menasce, D. Thurston, G. Smith, and S. C. Sutherland, "Design of playful authoring tools for social and behavioral science," in *Intelligent User Interfaces Companion Proceedings*, pp. 157–160, 2017.

[35] M. O. Riedl and V. Bulitko, "Interactive narrative: An intelligent systems approach," *Ai Magazine*, vol. 34, no. 1, pp. 67–67, 2013.

[36] N. Partlan, E. Carstensdottir, S. Snodgrass, E. Kleinman, G. Smith, C. Harteveld, and M. S. El-Nasr, "Exploratory automated analysis of structural features of interactive narrative," in *Fourteenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2018.

[37] J. A. Bondy, U. S. R. Murty, *et al.*, *Graph theory with applications*, vol. 290. Macmillan London, 1976.

[38] R. Tarjan, "Depth-first search and linear graph algorithms," *SIAM journal on computing*, vol. 1, no. 2, pp. 146–160, 1972.

[39] N. Wardrip-Fruin, M. Mateas, S. Dow, and S. Sali, "Agency reconsidered.," in *DiGRA conference*, Citeseer, 2009.

[40] D. Thue, V. Bulitko, M. Spetch, and T. Romanuik, "Player agency and the relevance of decisions," in *Joint International Conference on Interactive Digital Storytelling*, pp. 210–215, Springer, 2010.

[41] N. Szilas and I. Ilea, "Objective metrics for interactive narrative," in *International Conference on Interactive Digital Storytelling*, pp. 91–102, Springer, 2014.

[42] E. Carstensdottir, E. Kleinman, and M. S. El-Nasr, "Towards an interaction model for interactive narratives," in *International Conference on Interactive Digital Storytelling*, pp. 274–277, Springer, 2017.

[43] Y.-C. Lee, S.-C. Huang, C.-H. Huang, and H.-H. Wu, "A new approach to identify high burnout medical staffs by kernel k-means cluster analysis in a regional teaching hospital in taiwan," *Inquiry: The Journal of Health Care Organization, Provision, and Financing*, vol. 53, p. 0046958016679306, 2016.

[44] A. Lemsara, S. Ouadfel, and H. Fröhlich, "Pathme: pathway based multi-modal sparse autoencoders for clustering of patient-level multi-omics data," *BMC bioinformatics*, vol. 21, pp. 1–20, 2020.

[45] D. Kamthania, A. Pawa, and S. S. Madhavan, "Market segmentation analysis and visualization using k-mode clustering algorithm for e-commerce business," *Journal of computing and information technology*, vol. 26, no. 1, pp. 57–68, 2018.

[46] D. Yang, T. Sinha, D. Adamson, and C. P. Rosé, "Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses," in *Proceedings of the 2013 NIPS Data-driven education workshop*, vol. 11, p. 14, Lake Tahoe, NV, 2013.

[47] S. Bharara, S. Sabitha, and A. Bansal, "Application of learning analytics using clustering data mining for students' disposition analysis," *Education and Information Technologies*, vol. 23, pp. 957–984, 2018.

[48] H. Mao, M. Schwarzkopf, S. B. Venkatakrishnan, Z. Meng, and M. Alizadeh, "Learning scheduling algorithms for data processing clusters," in *Proceedings of the ACM special interest group on data communication*, pp. 270–288, 2019.

[49] M. B. Miles and A. M. Huberman, *Qualitative Data Analysis: An Expanded Sourcebook*. SAGE, Jan. 1994.

[50] D. J. Hacker, J. Dunlosky, and A. C. Graesser, *Handbook of metacognition in education*. Routledge, 2009.

[51] T. Fullerton, *Game design workshop: a playcentric approach to creating innovative games*. CRC press, 2014.

[52] R. Moreno and A. Valdez, "Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback," *Educational Technology Research and Development*, vol. 53, no. 3, pp. 35–45, 2005.

[53] C. Oh, T. Lee, Y. Kim, S. Park, S. Kwon, and B. Suh, "Us vs. them: Understanding artificial intelligence technophobia over the google deepmind challenge match," in *Proceedings of the 2017 CHI Conference*, pp. 2523–2534, 2017.

[54] K. Torning and H. Oinas-Kukkonen, "Persuasive system design: State of the art and future directions," in *Proceedings of the 4th International Conference on Persuasive Technology*, Persuasive '09, (New York, NY, USA), Association for Computing Machinery, 2009.

[55] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human factors*, vol. 39, no. 2, pp. 230–253, 1997.

[56] B. Shneiderman, "Creativity support tools: A grand challenge for hci researchers," in *Engineering the User Interface*, pp. 1–9, Springer, 2009.

[57] S. C. Sutherland, C. Harteveld, and M. E. Young, "Effects of the advisor and environment on requesting and complying with automated advice," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 6, no. 4, pp. 1–36, 2016.

[58] R. S. Baker, S. K. D'Mello, M. M. T. Rodrigo, and A. C. Graesser, "Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments," *International Journal of Human-Computer Studies*, vol. 68, no. 4, pp. 223–241, 2010.

[59] T. De Jong, "Cognitive load theory, educational research, and instructional design: some food for thought," *Instructional science*, vol. 38, no. 2, pp. 105–134, 2010.

[60] T. Hatziapostolou and I. Paraskakis, "Enhancing the impact of formative feedback on student learning through an online feedback system.," *Electronic Journal of E-learning*, vol. 8, no. 2, pp. 111–122, 2010.