Mind the Gap: The Illusion of Skill Acquisition in Computational Thinking

Yeting Bao
Rochester Institute of Technology
Rochester, New York, USA
yetingbz@gmail.com

Hadi Hosseini
Pennsylvania State University
University Park, Pennsylvania, USA
hadi@psu.edu

ABSTRACT

With the advent of online educational platforms and the advances in pedagogical technologies, self-directed learning has emerged as one of the most popular modes of learning. Distance education—elevated by the COVID-19 pandemic—involves methods of instruction through a variety of remote activities which often rely on educational videos for mastery. In the absence of direct student engagement, the asynchronous nature of remote activities may deteriorate the quality of education for learners. Students often have an *illusion of skill acquisition* after watching videos, which results in overestimation of abilities and skills. We focus on the efficacy of skill acquisition through interactive technologies and assess their impact on computational thinking in comparison with delivery through other traditional media (e.g. videos and texts). In particular, we investigate the relationship between actual learning, perception of learning, and learners' confidence in adult learners. Our results reveal intriguing observations about the role of interactivity and visualization and their implications on the pedagogical design for self-directed learning modules.

CCS CONCEPTS

• **Social and professional topics** → **Computational thinking**; **Computational science and engineering education**; • **Applied computing** → *Interactive learning environments.*

KEYWORDS

Skill Acquisition; Interactive Learning; Computational Thinking

ACM Reference Format:
Yeting Bao and Hadi Hosseini. 2023. Mind the Gap: The Illusion of Skill Acquisition in Computational Thinking. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2023), March 15–18, 2023, Toronto, ON, Canada.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3545945.3569749

1 INTRODUCTION

Recent research in psychology of skill learning suggests that solely watching videos creates an *illusion of skill acquisition* for learners without any measurable impact on people's actual abilities [14]. Surprisingly, even intense repetition (repeatedly watching videos) can only enable learners to 'track the steps', but does not help in skill development itself.[1] These findings and their impacts on *self-efficacy* or meta-cognition have been investigated recently in a variety of contexts [26, 38].

With the advent of online educational technologies (e.g. MOOCs and Khan Academy), self-directed learning has emerged as one of the most popular modes of learning, overcoming national and societal boundaries. The pandemic of COVID-19 and its subsequent global restrictions, more than ever, has exacerbated access to in-person learning opportunities and, as a result, elevated the importance of effective self-directed learning technologies and their impact on skill acquisition. Self-directed learning, at its core, utilizes *asynchronous* remote instruction techniques such as *text-based instruction* and *educational videos* that heavily rely on learners' effort and self-discipline. Yet, the lack of learner engagement, for example when watching educational videos, has shown to be detrimental to effective learning and comprehension [6]. The use of interactivity in computer-supported education to teach concepts is arguably motivated by its connection to the learning theory, and primarily, the *self determination theory* [5]. Interactive modules target learners' intrinsic motivation by giving learners a sense of control and active involvement, and as a result, can improve cognition and critical thinking among learners in various age ranges [30, 35].

Computational thinking is a mental skill that utilizes multiple levels of abstraction for reasoning to solve problems in a variety of areas [41]. It is one of the core skills in training students in computing-related fields. In computational thinking, several recent papers have investigated *active learning* methods and their impact on perception of learning, actual learning comprehension, and engagement, focusing on in-classroom pedagogical interventions and delivery styles [1, 10–12, 31]. To date, however, little has been done in evaluating (the illusion of) skill acquisition in computational and algorithmic thinking and its relation to learning performance and the perception of learning.

We focus on self-directed learning techniques in teaching computational thinking to adult learners through asynchronous education.[2] In this vein, we investigate the interplay between perceived learning, actual learning, and learners' confidence in acquiring applied knowledge, and ask the following questions: *Is there an illusion of skill acquisition in teaching computational thinking through a variety of learning technologies? Is there a gap between the perception of learning, actual learning comprehension, and learners' confidence?*

[1] Note that *vicarious learning* may not necessarily fall into this category because it involves *active* and *deliberate* imagination of actions and/or steps.

[2] It is critical to emphasize the self-directed aspect of education since we do not consider synchronous teaching technologies such as content delivery through Zoom that enable (albeit limited) teacher-learner or learner-learner real-time interactions.

SIGCSE '23, March 15–18, 2023, Toronto, ON, Canada.

ACM ISBN 978-1-4503-9431-4/23/03…$15.00
https://doi.org/10.1145/3545945.3569749

1.1 Our Contributions

We initiate a systematic comparison between an interactive platform and two other modes of self-directed learning, namely educational videos and text-based instruction. To this end, we utilized an interactive online platform—MatchU.ai [8]—where learners can engage with new algorithms, manipulate the input data, and visually learn the steps of the algorithms.

We conducted an empirical study with 150 participants on Amazon Mechanical Turk.[3] We measured learners' actual learning comprehension, perception of learning, confidence, and interest in learning on each of the treatments (interactive, video, text-based). Specifically, we highlight the role of interactivity as follows:

- **Learning comprehension**: learners in video and interactive groups outperformed those in the text-based group. While there was no statistically significant difference between the video and interactive groups, there was a larger variance between the participants in the video group compared to the interactive group. These findings, although not statistically conclusive, suggest that interactive technologies may be more suitable for a wider range of learners.
- **Perception of learning**: there is a significant difference between the groups with respect to the perception of learning. In particular, the perception of learning is significantly higher among participants of the interactive group. This is a surprising finding because (i) there was no significant difference between video and interactive groups with respect to actual learning and (ii) across all participants there is a strong positive correlation between actual learning and the perception of learning.
- **Confidence/interest in learning**: the participants in the video group were as confident as those in the interactive group, despite reporting significantly lower score for their perceived learning. This finding suggests that educational videos may be misleading with respect to confidence due to lack of interaction and missed opportunities to actively engage the learners. In addition, there was a significant difference between the three groups with respect to their interest in learning the material. Nonetheless, a follow up analysis was not able to show any statistical difference between the interactive and video groups.

While our study does not conclusively settle all aspects of skill acquisition in computational thinking, it provides several key insights and raises intriguing questions about the relationship between comprehension, perception, and confidence in algorithmic thinking. In Section 6 we discuss the limitations of our study, provide insights into next steps, and explain some of the lessons learned that may be of interest in conducting followup studies in the future.

1.2 Related Work

The emergence of novel web-based technologies has given rise to new possibilities in educational content development and has impacted the form of visualizations as pedagogical means (e.g. [19, 40, 42]). Online learning technologies increase accessibility while enabling students self-learn through interactions with the content [21]. In this vein, visualizations and interactive components can enhance learners' computational and algorithmic thinking, resulting in significant attention in computer-supported and cooperative education research [20, 27, 28]. In computer science education, visualization tools are often used by instructors in class in conjunction with other traditional-style lectures to demonstrate programming concepts or computational problem solving. Yet, the availability and spread of these tools have shown to be limited: approximately 20% of programming courses use software visualizations regularly [13]. Additionally, techniques such as programming visualizations for basic introductory CS courses and algorithmic visualizations for teaching data structures and algorithms, are generally of low quality with little pedagogical value [33]. Moreover, these techniques are generally passive and do not include interactive components, failing to engage students in the learning process. That is, visualization alone is insufficient for most self-directed learning environments.

Effective learning platforms should provide the ability to control the pace of visualization, focus on logical steps, provide effective examples, and allow learners to *manipulate* the input [32]. Naps et al. indicated that supporting student interaction and active learning makes visualization technologies more effective due to higher engagement levels [23]. They proposed five necessary forms of interactions in learning technologies: 1) animating the concepts (which is supported by most visualization systems, e.g., [7, 16, 39]), 2) responding to the system (e.g. [16, 36]), 3) ability to modify and interact by manipulating the input, 4) constructing visualizations manually, and 5) presenting and connecting with other learners.

Focusing on these interactivity forms, we design an interactive learning environment where learners can manipulate the algorithms, change instances, and interact with the visualizations. We evaluate learners' comprehension based on Blooms' taxonomy [2], perception of learning [14], confidence in applying the algorithms, and the usability of our computer-supported system [3]. The perception of learning and actual learning comprehension may have different contributing factors [4], but the relationship between them, in general (with few exceptions [17]), remains unstudied.

2 DESIGN AND DEVELOPMENT

2.1 Self-Directed Learning Methods

We considered the following learning strategies as primary methods for self-directed learning. These teaching strategies have different interaction levels based on the engagement categories developed by Naps et al. [23] (Table 1 provides a brief summary).

(1) Interactive platform: we developed online modules for interactive teaching of two algorithms. They contain animated visualizations for learning the algorithms, data manipulation, step by step progress and pace control, as well as additional manipulation features along with supportive text to explain the algorithmic concepts (see Section 2 for details).
(2) Educational videos: we selected well-produced educational videos on the topics that were carefully created by educators in the field and are publicly available. The videos are animated with voice-over and are about 6 minutes long.

[3] This study was conducted on Amazon Mechanical Turk due to restrictions during the COVID-19 pandemic. In Section 6, we discuss its limitations and potential future improvements under a controlled lab study.

(3) Text-Based instruction: we selected the instruction material from well-established courses/tutorials and enhanced them further by adding extra (static) graphical examples.

2.2 Interactive Design and Usability

We utilized a public academic platform, MatchU.ai [8], which is a not-for-profit effort with the mission of fostering the adoption of algorithmic decision-making and facilitating their learning by leveraging interactive design. We designed interactive modules each containing the following components: i) A *learning tab*, that briefly describes a sample problem, the algorithm, and the characteristics of the outcome, and ii) An *interactivity tab*, where users can interact with the algorithm and follow its steps visually. The interactive tab contains 1) input data components where users can manually enter the input or randomly generate input data, 2) visualization that animates each step of the algorithm, and 3) a control panel where users can change the pace of the steps, move backward or forward in every step (see Figure 1). The interactive modules do not include any audio explanation but include supportive text information.

Evaluating the Prototype. To assess the initial design of the interactive modules, we performed a *heuristic evaluation* [24, 25] by asking questions such as 1) Are users able to understand the language and figures used in the interactive platform? and 2) Can users create algorithm instances with minimum effort?

For the purpose of this initial study, four students majoring in human-computer interaction with no prior knowledge about these algorithms were recruited. We conducted a *semi-informal interview* where they were able to freely explore and interact with the platform and were encouraged to speak out loud and explain their thought process. Upon completion, participants filled out a post-session survey. We used the well-adopted System Usability Scale (SUS) test to measure the user experience.

After taking these consideration into account, we updated our design and recruited three additional college students to test the platform in person. Figure 1 illustrates the final design for one of the two algorithms. The SUS score for new prototype is 90.83, which indicates a desirable level of usability.

3 EXPERIMENTAL DESIGN

We conducted a randomized between-subject study on Amazon Mechanical Turk (MTurk)—a popular crowdsourcing marketplace—through a single Human Intelligence Task (HIT). We recruited 150 participants (turkers) to complete a 20-minute learning task. The experiment consisted of a demographic survey, a pre-knowledge assessment test, a learning task, a questionnaire about perception of learning, a post-assessment test with 5 questions, and a usability questionnaire (only for the interactive treatment).

The payment for the HIT is divided into two parts—a *base* payment, and a *bonus* payment. If a participant answers all the questions within the HIT, she gets a base payment of 50¢ upon submission. Depending on the consistency of responses a participant may receive an additional bonus payment of 50¢.

Treatments. To evaluate the efficacy of the learning methods in algorithmic/computational thinking, we conducted our experiment on two well-studied algorithms that are foundational in artificial intelligence research. The first algorithms is the Deferred Acceptance algorithm [9] for finding stable matching between two disjoint sets. This algorithm has been in the center for attention in National Residency Matching program as well as several related organ/kidney exchange programs. We also evaluated a second algorithm, called the Top Trading Cycle algorithm [34], that has been widely used for assigning courses to students or students to public schools.[4]

Table 1: Interaction level of three teaching strategies.

Learning Strategy	Mode			Manipulation
	Text	Graphic	Animation	
Text-Based Instruction	✓	✓		
Educational Videos	✓	✓	✓	
Interactive Module	✓	✓	✓	✓

The choice of these two algorithms, in contrast to other algorithms, is intentional for several reasons: First, these algorithms are often not taught in most typical undergraduate courses on data structures or analysis of algorithms. As a result, our data is more likely not to suffer from any *selection* or *anchoring bias*. Later, we will discuss a pre-knowledge test to check for such biases. Second, these two algorithmic methods are sufficiently intuitive and straightforward, and thus, require little prior knowledge.

Demographic Survey. Prior to the start of the task, each participant was asked demographic questions about gender, age, level of education, and prior exposure to one of the algorithms. To avoid any type of *selection bias*, we only included adult participants with no prior knowledge about these algorithms. The breakdown of our participants is as follows: 50.67% female and 48.67% male; 56.67% have some computing-related background; 43.33% did not have computing-related background; 56.67% were undergrads; 32% postgraduates; and 8% were PhD students. The participants' age group were distributed as follows: 28% aged between 19–22; 22.67% between 23–26; 12.67% between 27–30; and 34.67% were over 30.

Assignment and Workflow. Each participant was randomly assigned to one of the treatment groups. Each treatment group received about 50 responses. In each treatment group, participants completed a learning task according to the assigned learning mode (interactive, educational video, or text-based instruction). Upon finishing a learning task, each participant answered a set of questions about their perceived level of learning, followed by a post-assessment test. The participants in the interactive group also completed an additional usability questionnaire (see Figure 2).

Pre-knowledge Assessment. To test participant's prior knowledge of the material, we conducted a short pre-knowledge test that included 5 questions. These questions are meant to assess participants' knowledge about well-known data structures and algorithmic concepts. The first question is participants self-evaluation of their knowledge. We asked participants "How do you rate your knowledge of the following concepts?" where they rated through a 5-point Likert scale from "Not familiar" to "Very Familiar" their familiarity with "Arrays, Linked-List, Stacks and Queues, Binary Trees, Graph Structures, and Hash Tables". The remaining questions formed a quick quiz, for example, as follows:

What is the run-time of the following code snippet (in Big-O notation?)

[4]Due to space constraints, we only discuss the results of the first algorithm. The results of the second algorithm are qualitatively similar, and further confirm our findings.

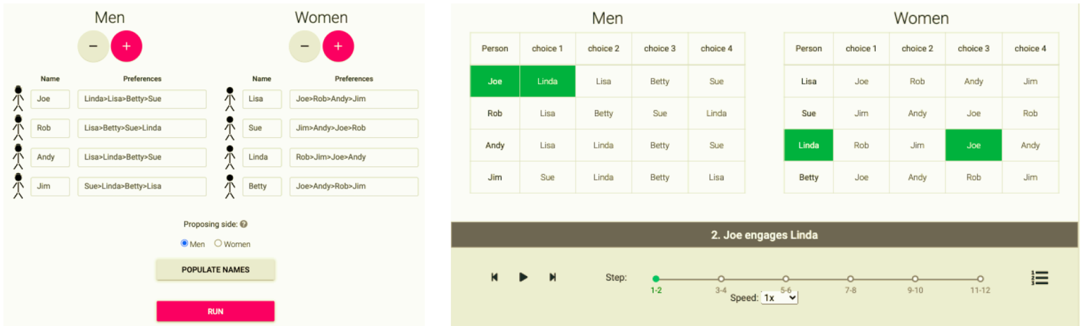


Figure 1: A snapshot of the interactive visual design that enables learners to interact and manipulate the input.

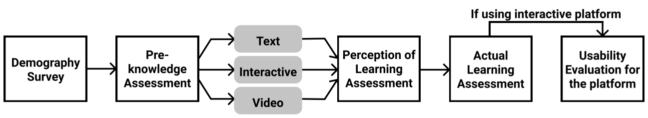


Figure 2: The workflow of the experimental design.

```
int sum = 0;
for (int i = 1; i < n; i = i * 2){
    sum++;}
```

(1) $O(log(n))$, (2) $O(2^n)$, (3) $O(n^2)$, (4) $O(n)$, (5) Do not know.

There was no significant difference between the groups with respect to self evaluation (Kruskal-Wallis: $H(2) = 1.58, p = .454$). Similarly, the remaining four assessment questions showed no difference between the prior knowledge among the groups (one-way ANOVA: $F(2, 147) = 0.82, p = .441$). There was no difference between participants that were assigned randomly to different groups with respect to the prior knowledge. Therefore, our remaining analysis is reliable with respect to potential selection biases.

Perception Questionnaire. Inspired by the experiments conducted by Kardas and O'Brien [14] to measure the *perception of skill acquisition*, we developed six questions that ask a learner to report their perceived learning experience as well as their confidence in applying the learned topic to solve similar or new problems. The questions are categorized based on *Bloom's taxonomy of learning* [2] and address the first three objectives level of the *cognitive domain*.

Each question is designed specifically to measure how much a participant believes to have learned the material (knowledge acquisition), comprehend the details (comprehension), and can solve new problems by applying the acquired knowledge and techniques (application). Below is an example of the question statement:

I have comprehended the following aspects of this algorithm (5-point scale from "Strongly Disagree" to "Strongly Agree"):
(1) What problems it can solve; (2) How it works on those problems; (3) What properties it guarantees; (4) Its application to other problems.

Additional questions elicit participants confidence in their answers. For instance, participants were asked to rate the following: *You are asked to run this algorithm on a given problem. What do you feel are the chances that you'd successfully execute this algorithm?* Select from a 5-point Likert scale: "I feel there's no chance at all I'd succeed." to "I feel I'd definitely succeed without a doubt."

Post Assessment. Upon completion of the perception questionnaire, participants go through a post-assessment quiz with five standard questions about the algorithm. The questions were designed to test their knowledge according to the Bloom's taxonomy of learning ranging from recalling concepts to application of the algorithm on new sample inputs.

Usability Questionnaire. To assess the relation between learning and usability of the interactive modules, the interactive group completed an additional questionnaire based on the System Usability Scale (SUS). This is a widely adopted test—proposed by John Brooke [3]—that measures the perceptions of usability in technological platforms where users interact with computer systems. The SUS questionnaire contains five positively worded statements and five alternate negatively worded statements to reduce *framing bias*. The total score is 100; any score above 50 and below 70 suggests that the usability is fair but needs some mild improvements. For instance, one question asked participants to rate the following statement "*I think that I would like to use the platform frequently.*".

For our interactive platform, the mean SUS score was 53.75, suggesting that the usability of the educational modules are acceptable but needs some mild improvements (see the result section for an extensive discussion). In the final question, we asked participants to leave any comments about the experiment or their experience.

Response Qualification. To ensure high-quality responses, we deployed several 'sanity checks' along the way. First, we restricted our HIT on MTurk to only be visible to participants who meet the following conditions: (a) completed at least 100 HITS successfully, (b) above 85% approval rate on previous HITs, (c) completed high school, and (d) the region restricted to the US and Canada.[5]

We performed two additional 'sanity checks' in the middle and at the end of the experiment by repeating a question towards the end of the experiment and including a simple memory retrieval question to check attention. This process ensures that MTurk participants put minimal effort in going through the learning task. About 65% of participants (150 out of 231) successfully passed our sanity checks.

Time Spent. To ensure that our analysis is not influenced by the time participants have spent on the learning material, we also collected the task completion time. A one-way ANOVA analysis showed no statistically significant difference between any of the interactive ($M = 5.43$ min, $SD = 9.98$), video ($M = 5.839$ min,

[5]We restricted location to ensure language proficiency and prevent any potential issues due to linguistic barriers.

$SD = 6.492$), and the text ($M = 4.952$ min, $SD = 5.059$) groups ($p = 0.838$). Moreover, we found no significant (positive or negative) correlation between time spent and other parameters such as actual learning, perception, or interest.

4 RESULTS

4.1 Perception of Learning

We collected responses about the *perception of learning* as well as *confidence* about applying the learned algorithm to solve similar problems (See Section 3). Since the collected responses are ordinal, we used a Kruskal-Wallis test to compare differences among three groups and an additional Mann-Whitney test to compare between each of the two groups. The results show a statistically significant difference between all groups on questions pertaining to the *perception of learning* ($H(2) = 7.47, p = .024$) and *confidence* ($H(2) = 9.01, p = .011$). A non-parametric Mann-Whitney test between each pair of the treatments revealed an intriguing observation: there is a statistically significant difference between the interactive group ($mdn = 3$) and the video group ($mdn = 2, U = 2858.5, p = .018$), and between the interactive group and the text-based group ($mdn = 3, U = 2193.5, p = .019$).

With respect to confidence, although there was a significant difference between the interactive group and the text-based group ($mdn = 3, U = 2120.5, p = .003$), there was no difference between the interactive and the video group ($mdn(both) = 3, U = 2770.50, p = .072$). Figures 3a and 3b illustrate the distributions for perception of learning and confidence.

These findings reveal an intriguing observation: while the learners of the interactive group reported significantly higher score for perceived comprehension—in contrast to video and text-based groups—the participants in the video group were as confident about their knowledge as those in the interactive group. This finding suggests that educational videos may lead learners to develop overconfidence about the subject matter. In contrast, interactivity provides an opportunity for learners to 'get their hands dirty' by manipulating the input and directly interacting with the algorithms; which in turn aligns learners' perceived comprehension with confidence.

4.2 Actual Learning Comprehension

We measured the learning comprehension of the participants through a summative assessment that contained five questions. All participants (irrespective of the treatment) answered the same set of questions. These answers were manually graded according to a predefined rubric. The descriptive statistics reveal less variation in responses of the interactive group ($M = 2.13, SD = 1.13$) compared to the text ($M = 2.32, SD = 1.32$) and the video ($M = 2.61, SD = 1.37$) groups (Figure 3c). A one-way ANOVA test showed that there is no significant difference in actual learning comprehension among the video and the interactive groups ($F(2, 147) = 1.79, p = .171$).

We postulate that these inconclusive findings are due to challenges in measuring actual learning in education, which requires significant control on the environment and learners' attitude. Actual learning is often attributed to a variety of internal and external factors [29], resulting in low or inconclusive efficacy of pedagogical interventions [18].

Table 2: Positive and negative comments.

Group	Positive	Negative	Total
Text-Based	19 (90.48%)	2 (9.52%)	21 (26.25%)
Educational Video	29 (96.67%)	1 (3.33%)	30 (37.50%)
Interactive	29 (100.00%)	0 (0.00%)	29 (36.25%)
Total	77 (96.25%)	3 (3.75%)	80 (100%)

4.3 Interest in Learning

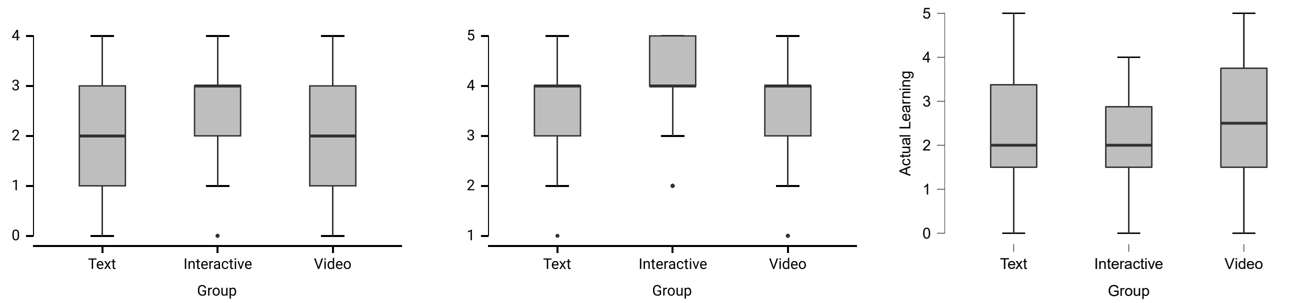
To compare the engagement level and interest among the groups, we asked a few additional questions in the post-assessment survey. A sample question is "*I want to learn more about matching algorithms in the future*" ranging from 'strongly disagree' to 'strongly agree' (5-point). A Kruskal-Wallis test showed a significant difference between the groups ($H(2) = 6.44, p = .04$). A follow up pairwise Mann-Whitney test between the video ($mdn = 5$) and the text-based group ($mdn = 4$) indicates that more participants in the video group have increased interest in learning the material ($U = 2196, p = .011$). However, there was no difference between the video group ($mdn = 5$) and the interactive group ($mdn = 5, U = 2399.5, p = .304$).

4.4 Usability

To measure the usability of the interactive modules, we collected open-ended comments from the participants and manually analyzed their positive or negative sentiments (Table 2). Overall, the interactive and video groups received substantially more positive comments compared to the text group. However, a follow up chi-square analysis found no significant difference between the groups (chi-square value = 3.08). We attribute this result to two important factors: First, the interactive treatment required participants to proactively engage in the learning process by manipulating the input. Thus, more *effort* is required in this mode compared to the video group. Second, the educational videos contained audio in addition to the visual explanation. We hypothesize that audio can have an important impact on the overall learning experience. As expected, those in the text-based group were less satisfied and found the instructions complicated and sometimes confusing. For example, two participants commented: *"Very confusing. I'm sorry but I did put genuine effort into the assignment. I'm absolutely not computer language literate. Not adept at all, however thank you for giving me the opportunity to be a part of this HIT."*, and *"it was kind of complicated."*

5 POSITIVE CORRELATIONS

Actual Learning and Perception. Although we found no significant difference in actual learning, the perception of learning between the groups varied. An interesting question for us was to find whether the perception of learning and actual learning are statistically correlated. To investigate this relation, we conducted a Pearson correlation test across all treatment groups. Inspired by Kardas and O'Brien [15], our goal was to measure how participants perceive their acquired knowledge. Specifically, we were interested in measuring applied knowledge corresponding to the first three levels of cognitive domain in the *Bloom's taxonomy* of learning [2]. Therefore, once the learning task was completed, participants



(a) Perceived comprehension (b) Confidence about learning (c) Actual learning comprehension

Figure 3: Box plots for perception of learning, confidence, and actual learning comprehension.

responded to the following question: "*You are asked to run this algorithm on a given problem. What do you feel are the chances that you'd successfully execute this algorithm?*". It turns out that the perception of learning is positively correlated with the actual learning comprehension ($r(148) = .326, p < .001$), and the positive correlation remains the same within the groups.

Actual Learning and Usability. A key factor in successfully designing learning technology for self-directed learning is usability of the systems. Previous studies showed that the usability of digital technology for education improves student engagement and learning [37]. Therefore, we conducted a Pearson correlation test between the usability scores (SUS) and the actual learning comprehension of the post-assessment quiz. The results reveal that usability and learning are positively correlated ($r(48) = .381, p = .006$), suggesting that the usability of the interactive platform is an important factor in learners' ability to comprehend new knowledge.

6 LIMITATIONS & LESSONS LEARNED

Our preliminary study uncovered a series of insights in design and assessment of effective self-directed learning modules. Below, we briefly outline these limitations, insights, and lessons learned.

Lab Experiments vs. Crowdsourcing. The positive correlation between usability score (SUS) and actual learning ($r(148) = 0.381, p = .006$) emphasizes the importance of learner-focused designs. The SUS score of our modules through MTurk was 53.75, while the in-person semi-informal assessment of the same resulted in the score of 90.83. While both are within the acceptable range, the discrepancy between the two reveals a potential negative impact of conducting experiments on online platforms such as MTurk. Since participants are often less motivated in crowdsourcing platforms–with no oversight–our assessment of actual learning may be suffering from a *floor effect*, which prevents distinguishing between the groups. In fact, despite 20 minutes allocated time, participants in all groups spent an average of less than 6 minutes to complete the task (see Section 3 for details). We believe future studies should include formal in-person assessments with real learners in addition to a thorough study with students involved in self-directed learning.

Learner Agency. A crucial factor in educational technologies is providing learners with means of control for pace and step by step progression while interacting with the modules. The stop of the animation has shown to be a key factor in active learning [22]. Non-stop demonstrations, in forms of animated features or videos, do not provide an opportunity for taking a break, stepping back, and reflecting on the learning material. In several critical steps, providing learners the opportunity to pause and showing detailed explanations can be an effective way for encouraging deep thinking.

Active Interventions. To improve understanding of concepts, the interactive platforms should be designed to actively *intervene* and prompt learners by asking questions, showing new features, or verifying some answers during the visualization process. Asking the user to predict the next step of the visualization has shown to be an effective pedagogical tool in previous studies [16, 36]. These interventions are crucial in self-directed learning technologies to promote mastery and further engage learners with the platform.

Instance generation. Our findings postulate that merely constructing input instances for visualization may not be a sufficient learning strategy. From the qualitative results, we learned that the educational modules should cover a wide range of instances as well as several interesting edge cases. In our experiments, we found that learners would like to see more 'edge cases' in addition to randomly generated examples and the ones created by themselves. Thus, a feature that allows learners to switch between a list of useful examples may help learners discover all 'interesting' cases.

7 CONCLUDING REMARKS

While interactive platforms improve the perceived learning [10, 12] compared to other methods, educational videos also provide a sense of confidence among learners. Although our findings do not show any impact on actual learning, the gap between confidence in applying algorithmic techniques and perceived knowledge suggests that educational videos (when used in isolation) may be detrimental to learning by providing an illusion of confidence. The interactive visualizations enhance learners' perception of learning and improve their confidence in the ability to solve problems. While perception and interest are positively correlated with comprehension, further research is required to identify the causal relationship between actual learning, perception, and confidence in self-directed learning.

ACKNOWLEDGMENTS

We acknowledge support from NSF grants #1915404 and #2144413. We are grateful to the anonymous reviewers for their valuable and constructive feedback.

REFERENCES

[1] Anders Berglund, Anna Eckerdal, Arnold Pears, Philip East, Päivi Kinnunen, Lauri Malmi, Robert McCartney, Jan-Erik Moström, Laurie Murphy, Mark Ratcliffe, et al. 2009. Learning computer science: Perceptions, actions and roles. *European Journal of Engineering Education* 34, 4 (2009), 327–338.

[2] Benjamin Samuel Bloom, Committee of College, and University Examiners. 1964. *Taxonomy of educational objectives*. Vol. 2. Longmans, Green New York.

[3] John Brooke. 2013. SUS: A Retrospective. *Journal of Usability Studies* 8, 2 (2013), 29–40.

[4] Shana K. Carpenter, Miko M. Wilford, Nate Kornell, and Kellie M. Mullaney. 2013. Appearances can be deceiving: instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review* 20, 6 (2013), 1350–1356. https://doi.org/10.3758/s13423-013-0442-z

[5] Edward Deci and Richard M Ryan. 1985. *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

[6] Louis Deslauriers, Logan S McCarty, Kelly Miller, Kristina Callaghan, and Greg Kestin. 2019. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences* 116, 39 (2019), 19251–19257.

[7] Mohammed F. Farghally, Kyu Han Koh, Hossameldin Shahin, and Clifford A. Shaffer. 2017. Evaluating the Effectiveness of Algorithm Analysis Visualizations. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (Seattle, Washington, USA) *(SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, 201–206. https://doi.org/10.1145/3017680.3017698

[8] James Ferris and Hadi Hosseini. 2020. MatchU: An Interactive Matching Platform. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 13606–13607.

[9] D. Gale and L. S. Shapley. 1962. College Admissions and the Stability of Marriage. *The American Mathematical Monthly* 69, 1 (1962), 9–15. https://doi.org/10.1080/00029890.1962.11989827 arXiv:https://doi.org/10.1080/00029890.1962.11989827

[10] Maxwell Hartt, Hadi Hosseini, and Mehrnaz Mostafapour. 2020. Game On: Exploring the Effectiveness of Game-based Learning. *Planning Practice & Research* 35, 5 (2020), 589–604.

[11] Hadi Hosseini, Maxwell Hartt, and Mehrnaz Mostafapour. 2019. Learning IS Child's Play: Game-Based Learning in Computer Science Education. *Journal of Planning Education and Research, ACM Transactions on Computing Education (TOCE)* forthcoming (2019).

[12] Hadi Hosseini and Laurel Perweiler. 2019. Are You Game?. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (Minneapolis, MN, USA) *(SIGCSE '19)*. Association for Computing Machinery, New York, NY, USA, 866–872. https://doi.org/10.1145/3287324.3287411

[13] Essi Isohanni and Hannu-Matti Järvinen. 2014. Are Visualization Tools Used in Programming Education? By Whom, How, Why, and Why Not?. In *Proceedings of the 14th Koli Calling International Conference on Computing Education Research*. Association for Computing Machinery, New York, NY, USA, 35–40.

[14] Michael Kardas and Ed O'Brien. 2018. Easier seen than done: Merely watching others perform can foster an illusion of skill acquisition. *Psychological science* (2018).

[15] Michael Kardas and Ed O'Brien. 2018. Easier Seen Than Done: Merely Watching Others Perform Can Foster an Illusion of Skill Acquisition. *Psychol Sci* 29, 4 (Apr 2018), 521–536. https://doi.org/10.1177/0956797617740646

[16] Zoltan Katai. 2014. Selective Hiding for Improved Algorithmic Visualization. In *Proceedings of the 2014 Conference on Innovation & Technology in Computer Science Education* (Uppsala, Sweden) *(ITiCSE '14)*. Association for Computing Machinery, New York, NY, USA, 33–38. https://doi.org/10.1145/2591708.2591734

[17] Kerri-Ann Kuhn and Sharyn Rundle-Thiele. 2009. Curriculum alignment: Student perception of learning achievement measures. *International Journal of Teaching and Learning in Higher Education* 21, 3 (2009), 351–361.

[18] Anders Larrabee Sønderlund, Emily Hughes, and Joanne Smith. 2019. The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology* 50, 5 (2019), 2594–2618.

[19] Meng-Yun Liao, Ching-Ying Sung, Hao-Chuan Wang, Wen-Chieh Lin, and Fu-Yin Cherng. 2019. Embodying Historical Learners' Messages as Learning Companions in a VR Classroom. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312861

[20] Junhua Liu, Lionell Loh, Ernest Ng, Yijia Chen, Kristin L. Wood, and Kwan Hui Lim. 2020. Self-Evolving Adaptive Learning for Personalized Education. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) *(CSCW '20 Companion)*. Association for Computing Machinery, New York, NY, USA, 317–321. https://doi.org/10.1145/3406865.3418326

[21] Marc Najork. 2001. Web-Based Algorithm Animation. In *Proceedings of the 38th Annual Design Automation Conference* (Las Vegas, Nevada, USA) *(DAC '01)*. Association for Computing Machinery, New York, NY, USA, 506–511.

[22] Thomas L. Naps, James R. Eagan, and Laura L. Norton. 2000. JHAVÉ—an Environment to Actively Engage Students in Web-Based Algorithm Visualizations. In *Proceedings of the Thirty-First SIGCSE Technical Symposium on Computer Science Education* (Austin, Texas, USA) *(SIGCSE '00)*. Association for Computing Machinery, New York, NY, USA, 109–113. https://doi.org/10.1145/330908.331829

[23] Thomas L. Naps, Guido Rößling, Vicki Almstrum, Wanda Dann, Rudolf Fleischer, Chris Hundhausen, Ari Korhonen, Lauri Malmi, Myles McNally, Susan Rodger, and J. Ángel Velázquez-Iturbide. 2002. Exploring the Role of Visualization and Engagement in Computer Science Education. In *Working Group Reports from ITiCSE on Innovation and Technology in Computer Science Education* (Aarhus, Denmark) *(ITiCSE-WGR '02)*. Association for Computing Machinery, New York, NY, USA, 131–152. https://doi.org/10.1145/960568.782998

[24] Jakob Nielsen. 1994. Usability inspection methods. In *Conference companion on Human factors in computing systems*. 413–414.

[25] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 249–256.

[26] Ed O'Brien. 2019. Enjoy it again: Repeat experiences are less repetitive than people think. *Journal of Personality and Social Psychology* 116, 4 (2019), 519.

[27] Jungkook Park, Yeong Hoon Park, Suin Kim, and Alice Oh. 2017. Eliph: Effective Visualization of Code History for Peer Assessment in Programming Education. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 458–467.

[28] Samir Passi and Steven Jackson. 2017. Data Vision: Learning to See Through Algorithmic Abstraction. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2436–2447. https://doi.org/10.1145/2998181.2998331

[29] Paul R Pintrich, Robert W Roeser, and Elisabeth AM De Groot. 1994. Classroom and individual differences in early adolescents' motivation and self-regulated learning. *The Journal of Early Adolescence* 14, 2 (1994), 139–161.

[30] Jan L Plass, Bruce D Homer, and Charles K Kinzer. 2015. Foundations of game-based learning. *Educational Psychologist* 50, 4 (2015), 258–283.

[31] Phil Sands, Aman Yadav, and Jon Good. 2018. Computational thinking in K-12: In-service teacher perceptions of computational thinking. In *Computational thinking in the STEM disciplines*. Springer, 151–164.

[32] Purvi Saraiya, Clifford A. Shaffer, D. Scott McCrickard, and Chris North. 2004. Effective Features of Algorithm Visualizations. *SIGCSE Bull.* 36, 1 (March 2004), 382–386. https://doi.org/10.1145/1028174.971432

[33] Clifford A. Shaffer, Matthew L. Cooper, Alexander Joel D. Alon, Monika Akbar, Michael Stewart, Sean Ponce, and Stephen H. Edwards. 2010. Algorithm Visualization: The State of the Field. *ACM Trans. Comput. Educ.* 10, 3, Article 9 (2010), 22 pages.

[34] Lloyd Shapley and Herbert Scarf. 1974. On cores and indivisibility. *Journal of Mathematical Economics* 1, 1 (1974), 23 – 37. https://doi.org/10.1016/0304-4068(74)90033-0

[35] Elizabeth Simpson. 1971. Educational objectives in the psychomotor domain. *Behavioral Objectives in Curriculum Development: Selected Readings and Bibliography* 60, 2 (1971).

[36] David Scot Taylor, Andrei F. Lurie, Cay S. Horstmenn, Menko B. Johnson, Sean K. Sharma, and Edward C. Yin. 2009. Predictive vs. Passive Animation Learning Tools. In *Proceedings of the 40th ACM Technical Symposium on Computer Science Education* (Chattanooga, TN, USA) *(SIGCSE '09)*. Association for Computing Machinery, New York, NY, USA, 494–498. https://doi.org/10.1145/1508865.1509038

[37] Jaime Urquiza-Fuentes and J. Ángel Velázquez-Iturbide. 2009. A Survey of Successful Evaluations of Program Visualization and Algorithm Animation Systems. *ACM Trans. Comput. Educ.* 9, 2, Article 9 (June 2009), 21 pages. https://doi.org/10.1145/1538234.1538236

[38] Lisa Vangsness and Michael E Young. 2020. More isn't always better: when metacognitive prompts are misleading. *Metacognition and Learning* (2020), 1–22.

[39] Peng Wang, Roman Bednarik, and Andrés Moreno. 2012. During Automatic Program Animation, Explanations after Animations Have Greater Impact than before Animations. In *Proceedings of the 12th Koli Calling International Conference on Computing Education Research* (Koli, Finland) *(Koli Calling '12)*. Association for Computing Machinery, New York, NY, USA, 100–109. https://doi.org/10.1145/2401796.2401808

[40] Qiaosi Wang, Shan Jing, Ida Camacho, David Joyner, and Ashok Goel. 2020. Jill Watson SA: Design and Evaluation of a Virtual Agent to Build Communities Among Online Learners. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3382878

[41] Jeannette M Wing. 2006. Computational thinking. *Commun. ACM* 49, 3 (2006), 33–35.

[42] Soojeong Yoo, Sunkyung Kim, and Youngho Lee. 2020. Learning by Doing: Evaluation of an Educational VR Application for the Care of Schizophrenic Patients. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3334480.3382851