Coding peekaboom: a gaming mechanism for harvesting programming concepts

Yi-Ling Lin1 · Shih-Yi Chien1 · Wei-Cheng Su2 · Sharon Ihan Hsiao3

Received: 9 May 2022 / Accepted: 4 September 2022 / Published online: 4 October 2022 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
There are plentiful online programming resources that enable learners to develop an understanding of conceptual knowledge and practical implementation. However, learners, especially novices, often experience difficulties locating the required information to solve the programming problems. Differ from natural language in syntax and convention, answers for programming languages may not be found just by simple text information retrieval. To address this issue, Coding Peekaboom, a game-based tagging was developed to help adequately index the critical concepts of a code segment. An EEG device was applied to measure participants’ mental states to identify their engagement during the gameplay. Study results include the effectiveness of appropriate concepts collected by participants whereas 47.15 concepts were collected on average in a game. The brainwave analysis and the questionnaire results reveal that participants were highly engaged in the tagging task via Coding Peekaboom. Correlations were found between the state of flow and the number of concepts selected, score, and time. Finally, the results of the flow theory and personal traits were reported to reflect the user experiences in the game.

Keywords Programming concepts · EEG · Flow theory · Gamification · Engagement
1 Introduction

Wilson et al., (2010) suggested that “the future workforce picture provides a compelling case as to why schools need to move toward improving computer science education.” To tackle this digital trend, the U.S. Congress announced that computer science (CS) education has officially become an essential component of STEM (science, technology, engineering, and mathematics) education in 2015. These STEM-strategic programs aim to provide students with broader exposure and access to CS-related fields and better prepare them for careers, which project the demand and importance of quality CS materials and resources.

Learning programming languages inevitably involves various complex cognitive activities (e.g., conceptual knowledge, structural constructions, and mathematical/logical thinking) and system operations (e.g., program design, modification, and documentation). Learners often encounter difficulties in implementing conceptual knowledge into structural operations and thus easily become frustrated while learning (Gaspar et al., 2008). Although plentiful online resources are made available to help learners gain a better understanding, it is still difficult for them, especially novices, to locate the required information to solve their problems. For example, many programming learners join StackOverflow\(^1\), a question-and-answer site for programmers to find answers to their programming problems. New learners can only look for programming resources by simple keywords, however, these keywords are not enough to get desired information. While programming languages differ from natural languages in syntax and convention, simple text information retrieval and modeling approaches are not sufficient to support code search in different programming languages (Linstead et al., 2009). It is challenging to provide appropriate materials to support the information needs of programming learners with different knowledge levels. A more extreme case for a novice, s/he may not even equip with enough knowledge to have right keywords. Hence, it becomes important to provide proper indexes of programming concepts and their notions to fulfill various information needs for code search.

In the education domain, content analysis was widely used to gather and analyze online learning resources (Gerbic & Stacey, 2005). However, it is challenging to employ content analysis on a large and increasing volume of online learning resources hence the high cost (Hsiao, 2015). With the development of Web 2.0, several studies have attempted to focus on analyzing content with social power, such as Cohere (Shum, 2008), social learning analytics (Shum & Ferguson, 2012), computer-supported collaborative learning (Ingulfsen et al., 2018), and collaborative learning online (Altebarmakian & Alterman, 2019). Social tags are used by researchers as good alternatives to traditional keywords for representing the content of a document (Lee et al., 2012). Social tags not only represent the content but also bridge the gap between humans and machines by including the user aspect in the indexing task (Lee et al., 2012). They have been widely used to improve web search (Lin et al., 2016, 2019), especially in terms of indexing. This is also the motive to design the proposed system, Coding Peekaboom, with tags to effectively index the code materials manually collected from StackOverflow.

\(^1\)https://stackoverflow.com/.
To motivate and engage users in social tagging, gamification, a popular topic (Morschheuser et al., 2017) in which game-design elements and principles are applied in non-game contexts (Toth & Tovolgyi, 2017), has been adopted to develop Coding Peekaboom for the Java programming context in this article. Coding Peekaboom is a gaming platform that allows a two-person team to participate in indexing the assigned Java code, where one of the team members first annotates the Java program with tags (i.e., acts as a coder tagging the programming concepts in a code segment) and the other guesses the potential concepts of the tagged code from her partner (i.e., acts as a learner searching for relevant programming materials via the tags). To measure the correctness of the tags, three domain experts were recruited to review them. Pre- and post-questionnaires were used to verify participants’ attitudes toward the system’s usability. In addition, to identify whether the participants engage in the proposed game, an EEG (electroencephalography) device was applied to measure their mental states (relaxed or concentrating) during the experiments. The concept of using EEG to measure task engagement is not new. Research has been focusing on the evaluation of the state of flow and immersion by using EEG (McMahan et al., 2015; Plotnikov et al., 2012; Wang & Hsu, 2014). By checking if the participants immerse in the game or not, we could focus on whether the participants could engage in this learning programming concepts game. The results show that the collaborative tags (user-generated tags on Coding Peekaboom) effectively index the critical concepts of a code segment. Additionally, brainwave analysis and questionnaire results reveal that participants were engaged in the tagging task via Coding Peekaboom.

2 Theoretical background

2.1 Social tagging

Social tagging is known as *folksonomy*, a combination of *folk* and *taxonomy*, which is the result of personal free tagging of information and objects done in a social environment (Wal, 2005). This indicates a categorization system based on the identification of each entity from a group of similar users. It is different from a classification system based on the orderly and systematic assignment of each entity (Jacob, 2004). So in contrast to taxonomy classification, folksonomy is simpler: those without training or previous knowledge in classification and indexing can quickly learn to make useful contributions (Trant, 2009). Folksonomy is characterized by the following: (1) tags are generated based on personal spontaneous definitions; (2) tags are available and visible to all; and (3) tags are decided by the frequency of use by the crowd.

Social tagging has been widely used in different aspects in recent years, for instance in research about the process of combating COVID-19 with social tagging technologies (Cha, 2020), or studies on how social tagging is used on social networks such as Facebook or Instagram and video sharing tools like Youtube. These have shown that the attractiveness of these tools has the potential to attract more business and yield higher revenues (Allam et al., 2019). Also, tags can facilitate enhanced e-learning, as shown by how folksonomies work in programming tutoring systems (Klašnja-Milićević et al., 2018). The results show that tagging helps learners to memorize
better and encourages learners to think, and thus has proven to be a meta-cognitive strategy in educational processes (Klašnja-Milićević et al., 2018).

Despite the popularity of social tagging, related research focuses mainly on its utilitarian benefits and not factors related to user acceptance. This paper extends the research thread on motivation (Allam et al., 2019) to design a tagging system that encourages users to participate in tasks.

2.2 Gamification

Gamification has gained popularity in recent years (Morschheuser et al., 2017), attracting research from aspects ranging from marketing, health, to sports, and even education (Piteira et al., 2018). Gamification theory in education holds that learners learn best when they are having fun (Dichev & Dicheva, 2017), and learn best when they have goals, targets, and achievements to reach for, of course in a way the learner still perceives as being fun. Gamification in learning involves using game-based elements such as point scoring, peer competition, teamwork, and score tables to promote engagement (Piteira et al., 2018). It helps students assimilate new information and test their knowledge. It can also apply to school-based subjects, self-teaching apps, and courses, again showing that gamification is not just for children (Dichev & Dicheva, 2017).

Gamification has been incorporated with social tagging to encourage user involvement in tasks. In 2004, game-based tagging was successfully explored in ESP (von Ahn & Dabbish, 2004), a game that collects image tags from a crowd to improve the accuracy of image search results. Peekaboom is an extension of ESP that collects not only information about “what” objects are in the image but also “where” each object is located in the image (von Ahn et al., 2006). Therefore, Peekaboom tags can be used to develop image recognition algorithms or even advanced artificial intelligence-related algorithms. Another game-based crowdsourcing task is gamifying crowdsourced relevance assessments (Eickhoff et al., 2012), that is, using a game to attract reliable players (workers) to conduct relevance assessments and then clustering these with a Human Intelligence Task (HIT) on Amazon’s Mechanical Turk. In the education domain, a game-based crowdsourcing system was also developed to help students learn how to write a program by annotating example programs (Hsiao & Lin, 2008).

2.3 User engagement

When talking about a person who acts “with total involvement”, what is being referred to is the individual’s holistic sensation of the concept of flow. Csikszentmihalyi (2000) first proposed the concept of flow to represent a mental state of consciousness that is experienced by individuals who are fully focused, absorbed, and engaged in an activity. It is also referred to as “the engaged state” when users are navigating an online environment (Hoffman & Novak, 1996; Lin & Tsai, 2012). Several flow-related measurements have been investigated, including enjoyment (Kaur et al., 2016), attention (Harris et al., 2017), involvement (Jiang et al., 2010), time distortion (Pelet et al., 2017), and satisfaction (Ali, 2016). With the development of ICT, studies
have applied the concept of flow to human-computer interaction and network activity. Some studies explore people’s flow experience in online shopping environments (Bilgihan et al., 2014, 2015; Koufaris, 2002; Novak et al., 2000; Rong & Min, 2005), and some researchers focus on gamifying services to push the user’s mental state to reach the flow state to ensure that the users stay in the service continuously where they are immersed (Cowley et al., 2008). The flow experience is a critical lens for investigating whether a user is fully involved in a service.

While most studies on the flow experience use questionnaire surveys with self-reported responses, using neuroscientific equipment to collect psychophysiological evidence has gained much attention in the social and information sciences (Dimoka et al., 2012; Fugate, 2007; Pavlou et al., 2007; Riedl et al., 2010; Nacke & Lindley, 2008) adopt EEGs, galvanic skin response (GSR), and a facial electromyography (EMG) system to investigate the brain activity of game players. Their study shows that physiological responses are an important indicator of the psychological states of game players.

Of the many neuro-scientific methods available for social science research, EEG is relatively portable, offers better temporal resolution, and is useful for recording cognitive and emotional responses while subjects perform experimental tasks (Kuan et al., 2014). Therefore, this study uses an EEG headset to observe players’ brainwaves when they are interacting with Coding Peekaboom, our experimental system. In addition, as the testbed system is a two-person game, we also observe the relation between the players’ mental states and their performance during the game.

2.4 Brainwaves

Brainwaves are divided into bandwidths, including delta, theta, alpha, and beta. Brainwaves affect human activity and influence their reactions (de Melo, 2017). In this study, we focus on alpha and beta waves because they are the dominant waves when people are awake and conscious. Alpha waves, associated with meditation, are the main component of EEG rhythms, defined as rhythms in the frequency range from 8 to 12Hz, in the waking state (Shestyuk et al., 2019). Alpha waves are generated in calmness and learning when people are less alert and react slower to stimuli (Ülker et al., 2017). Beta waves are responsible for problem-solving or decision-making and are associated with attention. They are defined as rhythms in the frequency range from 12 to 38Hz (Shestyuk et al., 2019). Beta activities have a lower amplitude. Beta wave frequencies increase when the subject is focused or excited (Ülker et al., 2017). When beta waves are the dominant waves, the subject’s attention is directed towards the task at hand and he/she reacts quickly to external stimuli (Desai et al., 2015a; Neurosky Inc., 2009; Tiago-Costa et al., 2016).

Based on these definitions of brainwaves, in the following sections, we use the terms “meditation” to represent alpha waves and “attention” to represent beta waves. The literature has investigated the relationships between brainwaves and flow theory. Wang et al., (2014) find that attention (beta waves) is not equivalent to flow but only constitutes a component thereof, indicating that other brainwaves might also play a role.
role in reaching the status of flow. Yang et al., (2019) incorporate meditation (alpha waves) in addition to attention and evaluate their relationships with flow and painting behavior. Their results also suggest that attention is only one component of flow and that meditation influences painting behavior; however, they do not thoroughly investigate the relationship between meditation and flow. These research results reveal that even though attention is not a strong indicator of flow, it is related to a person’s mental state. In terms of meditation, there is still a lack of understanding about its relationship with the flow. Therefore, we adopt brainwave-based data on attention and meditation to analyze users’ mental states and investigate what brainwaves reveal about users’ interactions with the proposed system.

3 Coding peekaboom

“Coding Peekaboom” was developed as a game-based system to attract people to take part in tagging concepts of programs. Since programming is not general knowledge, the target crowd of “Coding Peekaboom” is those who possess basic programming abilities, particularly in Java. “Coding Peekaboom” was designed as a collaborative game with Challenger and Host player roles. Any player can start this game in a random role from the web.

In the game, the Host starts with a piece of source code (Fig. 1) and then he progressively highlights code segments and assigns proper programming concepts as a tagging question which Challenger will wait until the Host is done and answers later by guessing what concepts the segment contains (Fig.1). The highlighted code segment is displayed on the screen of the Challenger, who attempts to guess what concepts the Host has assigned for the code segment while the Host waits until the Challenger
finishes guessing (Fig. 2). After the Challenger submits her answers the first time, the upper left corner shows exactly how many concepts the Host selected, and the bottom right corner indicates which concepts match the Host’s selection (Fig.2). If the Challenger’s guesses match what the Host has selected, they pass this round and collect points as a reward. The more concepts are matched, the more points are gained. If the Challenger cannot guess what the Host has selected, or whenever the Challenger feels that no more suitable concepts can be applied to the code segment, the Challenger clicks the “Pass” button to finish the current round. After playing for 20min, to make the game more entertaining, we provide a “Change Role” button to allow the users to exchange roles (Fig.2).

In addition to the basic functions, the game provides various messages to increase players’ awareness of the game status and encourage them to make further guesses. For example, “Strike! Congratulations!” indicates that the Challenger has matched all of the Host’s concepts; likewise, “There are 2 matches, you can try again” means the Challenger has matched two of the Host’s concepts. In terms of the reward system, the Host and the Challenger are rewarded equal points when the Challenger guesses the correct concepts. Ten points are given for each Match concept. If the Challenger completes all the Matches, they are rewarded a bonus of 20 points. For example, given a code segment with three assigned concepts from the Host, they get 20 points if the Challenger correctly guesses two Matches but 50 points if the Challenger correctly selects all three concepts.
4 Research methodology

To assess whether the proposed Coding Peekaboom game harvests coding concepts and whether it is a feasible gaming mechanism to complete this task, we performed a controlled user study. This section describes the study design.

4.1 Dataset

People can use a variety of keywords to describe a given programming concept. To determine whether a game-based system can be used for such a domain-specific purpose, we first used predefined programming concepts for annotations to account for player bias concerning concept description. The programming concepts in Coding Peekaboom are taken from Hsiao and Lin’s study (I-Han Hsiao & Lin 2017). Once the mechanism has been demonstrated to be a workable and reliable method, we will consider accepting open annotation answers from players.

As our game material, we chose StackOverflow posts that included code, resulting in a total of 23 game questions. Many programming learners join StackOverflow to find answers to their programming problems. Moreover, you can find different topics of experts in StackOverflow, however, novices without enough background knowledge may not be able to describe their dilemma properly. Therefore, how programming learners identify the concept and problem they are facing is important. By tags given in Coding Peekaboom, learners recognize those concepts faster and result in a better reaction when having questions on StackOverflow.

4.2 Participants

We recruited 24 groups of students from a university in Taiwan, for a total of 48 student participants. Six participants were from the department of computer science, and the remaining 42 students were from the department of information management. Nineteen students were graduate students and the rest were undergraduates. There are thirty males and eighteen females, and their average age lands between the range of twenty to twenty-two. Almost all the participants had over two years of programming experience and were used to searching for programming-related resources on the Internet. To ensure that every student had basic Java programming experience, the participants were requested to complete the Java programming test to evaluate whether he or she was suitable for the experiment. The test score threshold was 60 points; all the participants passed.

4.3 Apparatus

To observe the mental states of the participants, the participants were asked to wear an EEG headset. We used the MindWave Mobile EEG headset manufactured by Neurosky Inc. (Fig. 3) to detect participant brainwaves and transmit the data to a laptop via Bluetooth. This headset is a gel-free device with a dry sensor electrode on the forehead to read the weak electric currents of the brain. On another clip were two dry electrodes attached to the left earlobe as a ground wire.
Using the ThinkGear™ wafer to separate the brainwave signal from the noisy environment, it is amplified to produce a clear brainwave signal. Then, the brainwave is interpreted as an eSense™ parameter through the eSense™ patent algorithm to indicate the current mental state of the user. Previous research proved that this kind of wireless mobile EEG system can also provide reliable results compared with the traditionally lab-based systems (Rogers et al., 2016). With the “Neuroview” software developed by Neurosky Inc., it can provide the eSense™ parameter of Attention and Meditation indicators, a value ranging from 0 to 100, derived from the raw brainwave value to evaluate the participants’ mental focus (Neurosky Inc., 2009). The attention index also correlated with the situation when the challenge people faced and the skill people used got balanced, which would lead to the flow experiment.

4.4 Design and procedure

Once two participants had passed the test and proceeded to the official session, they were randomly assigned to different roles, one as Host and the other as Challenger. They were given about five minutes to familiarize themselves with the system via the experimental conductor, after which the official session began.

The official session was comprised of two phases. In the first phase, the participants operated the system in their initially-assigned roles; in the second phase, their roles were switched. Their task was to operate the system and achieve high scores. During the experiment, participants wore EEG headsets to record their brainwave data. After the official session, the participants filled out a post-experiment questionnaire to measure their experience of flow during the experiment. The questionnaire was designed on a 7-point Likert-type scale with constructs of enjoyment (Yager et al., 1997), focused attention (Steuer, 1992), involvement (Ghani, 1995; Novak et al., 2000), time distortion (Skadberg & Kimmel, 2004), and satisfaction (Wang & Hsu, 2014). Participants were given an NT$200 reward for completing the experiment. The overall experimental procedure is shown in Fig. 4.
In the experiment, 48 responses (24 groups) were received. Due to technical difficulties, the brainwave signals of the four groups were incomplete. Thus, in the results, we report the data of only 20 groups. Table 1 describes the participants’ performance. In total, 629 pieces of code were evaluated and 1595 programming concepts were highlighted by participants. All twenty-one concepts from Hsiao and Lin’s study (I-Han Hsiao & Lin 2017) provided in the game were assigned at least once. Table 2 shows the 21 concepts and how often each was assigned by participants.

### Table 1 Descriptive statistics of participant performance

| Performance | Average | Standard deviation |
|-------------|---------|--------------------|
| Brainwave Attention | 47.87 | 3.99 |
| Meditation | 54.21 | 3.35 |
| Performance Score | 485.25 | 185.36 |
| Time | 22.25 | 2.62 |
| Round | 11.53 | 4.3 |
| Number of concepts | 47.15 | 17.12 |
| Number of concepts per round | 4.14 | 1.42 |

### Table 2 Number of programming concepts assigned by participants

| Concept | # | Concept | # | Concept | # |
|---------|---|---------|---|---------|---|
| Arrays | 177 | IfElse | 206 | Objects | 103 |
| AccessControl | 15 | Inheritance | 13 | PrimitiveDataTypes | 129 |
| ArithmeticOperations | 121 | Interfaces | 7 | Strings | 122 |
| BooleanExpressions | 85 | LoopsDoWhile | 14 | Switch | 2 |
| Classes | 93 | LoopsFor | 145 | Thread | 11 |
| Constants | 8 | LoopsWhile | 115 | TwoDimesionalArrays | 4 |
| Exceptions | 28 | NestedLoops | 9 | Variables | 188 |

*Note. # stands for number*

5 Analysis and results

In the experiment, 48 responses (24 groups) were received. Due to technical difficulties, the brainwave signals of the four groups were incomplete. Thus, in the results, we report the data of only 20 groups. Table 1 describes the participants’ performance. In total, 629 pieces of code were evaluated and 1595 programming concepts were highlighted by participants. All twenty-one concepts from Hsiao and Lin’s study (I-Han Hsiao & Lin 2017) provided in the game were assigned at least once. Table 2 shows the 21 concepts and how often each was assigned by participants.

5.1 Quality of concepts

To evaluate the quality of the concepts identified, 60 pieces of code were randomly selected and assigned to three experts, who were asked to select the concepts that
they deemed correct. Experts were teachers of Java, compared to participants who had only learned the basic syntax of Java. The experts conducted the evaluation individually, after which their results were utilized as the quality evaluation baseline. We adopted three levels of agreement rate—union, intersection, and majority—from the experts’ results, due to the nature of our system and experiment. Our system was designed to collect concepts in a gaming manner: together with her partner, every user should select concepts that both considered appropriate for a certain code segment. Therefore, the concepts selected for a given code segment could vary widely from user to user. Each segment could be viewed at different levels of granularity, resulting in various concepts for a given code segment.

The majority agreement rate was calculated based on the number of concepts selected by at least two experts. Note that just because a concept was not selected does not necessarily mean that the concept is irrelevant to the code. Nevertheless, there are two potential reasons for a concept not being selected: (1) the concept is indeed simply wrong or irrelevant to the code; (2) the concept is relevant to the code but is not recognized by the users (participants or experts). This is because users view the code based on arbitrary levels of granularity, which is uncontrollable. For instance, one may view the code in Fig.5 as an example of a for loop but overlook concepts such as condition that are also present in the code. In this study, for a more comprehensive analysis, we address this via the adoption of three-level agreement rates.

In Table 3 we report the average number of concepts selected by participants and experts per round among the 60 code segments: the intersection agreement rate was 64.79% and the majority agreement rate was 93.42%. As defined in Table 4, four measurements—precision rate, sensitivity rate, miss rate, and false alarm rate—were utilized for detailed quality evaluation. Note that in our context, as no answer is guaranteed to be wrong, our measurement definitions are different from the general ones.3

---

Table 3 Descriptive statistics of concepts selected per round by different users

|                | Average | Standard deviation |
|----------------|---------|--------------------|
| Participants   | 4.35    | 1.25               |
| Expert         |         |                    |
| Intersection   | 5.10    | 1.76               |
| Majority       | 6.73    | 1.63               |
| Union          | 8.03    | 2.01               |

3 https://en.wikipedia.org/wiki/Sensitivity_and_specificity.

Fig. 5 For loop (outlined in red) and condition (outlined in blue) concepts.

```java
for (int count = 0; count < maxSoFar){
  maxSoFar = myArray[count];
}
```
The results are shown in Table 5. Based on the union of three experts, the precision rate is 96%, the sensitivity rate is 52%, the miss rate is 48%, and the false alarm rate is 10 out of 261. Based on the intersection agreement rate, the precision rate is 75%, the sensitivity rate is 65%, the miss rate is 35%, and the false alarm rate is 64 out of 261. Based on the majority agreement rate, the precision rate is 94%, the sensitivity rate is 74%, the miss rate is 26%, and the false alarm rate is 23 out of 261. The results show that the precision rate is at least 79% across the three levels and the false alarms are all less than 1, indicating the high quality of concepts selected by participants.

### 5.2 Brainwave analysis

Brainwaves have proved to influence, or be correlated with, human behavior; therefore, a robust analysis should include brainwave observations during the experiment. To fully understand how the brainwaves change during the experiment, we conducted different levels of analyses. To improve the readability of the following paragraphs, regarding the change of working status, Challengers are discussed as progressing from idle to working and Hosts progress from working to idle, reflecting how the experiment was conducted. While working is when the Challenger answers the quiz and idle is when the Host waits for the Challenger ends the quiz. To evaluate the participants’ mental focus, the raw brainwave values were interpreted using eSense™ software with attention and meditation indicators ranging from 0 to 100 (Neurosky Inc., 2009). Attention indicates the intensity of a user’s level of mental “focus” or “attention”; mediation indicates the level of a user’s mental “calmness” or “relaxation” (Neurosky Inc., 2009).

| Table 4 | Quality measurements and their definitions |
|---------|-----------------------------------------------|
| Measurement | Definition |
| Precision rate | (participants ∩ experts) / (participants) | Percentage of participants’ results relevant to those selected by experts |
| Sensitivity rate | (participants ∩ experts) / (experts) | Percentage of total relevant results correctly selected by experts |
| Miss rate | 1 – sensitivity rate | Percentage of concepts selected by participants but not by experts |
| False alarm rate | (participants) – (participants ∩ experts) | Concepts selected by participants but not experts |

| Table 5 | Concept quality for three agreement rates |
|---------|------------------------------------------|
|          | Precision | Sensitivity | Miss | False alarm rate |
| Majority | 0.94      | 0.74        | 0.26 | 23/261           |
| Intersection | 0.75 | 0.65        | 0.35 | 64/261           |
| Union    | 0.96      | 0.52        | 0.48 | 10/261           |

Note. The reported precision, sensitivity, and miss rates are calculated as the average of the 60 selected questions. The false alarm rate is reported as the fraction of the total concepts selected by the participants.
In this analysis, we categorized our data into different parts to look for correlations in different groups (Fig. 6). First, we decided to look for correlations in whole participants, without grouping them by Hosts or Challengers (Fig. 6 (a)). Second, we further divided our participants into two groups by their roles. To check out how brainwave level changes when switching working status (Fig. 6 (b)), in the last part, we focused on attention level and meditation level differences held by Hosts and Challengers (Fig. 6 (c)(d)).

Taking the participants, their attention level changed significantly ($p < .001$) from working ($M = 48.13, SD = 5.48$) to idle ($M = 47.62, SD = 5.25$). However, their meditation level showed no difference when their working status changed. Considering the effect of the role, we categorized participants as Host or Challenger. For Challengers, we expected that when their status changed from idle to working, the meditation level would decrease and the attention level would increase. Conversely, for Hosts, we assumed that when changing from working to idle, their meditation level would increase and their attention level would decrease. The results (see Table 6)

**Table 6** Meditation and attention statistics results of all participants, challengers, and hosts

|        | Meditation | Attention |
|--------|------------|-----------|
|        | p-value    | Avg_w     | Avg_i     | p-value    | Avg_w     | Avg_i     |
| All    | 0.80       | 54.22     | 54.21     | 0.00       | 48.13     | 47.62     |
| Challenger | 0.94       | 54.06     | 54.07     | 0.014      | 47.87     | 47.47     |
| Host   | 0.69       | 54.39     | 54.35     | 0.00       | 48.53     | 47.88     |

*Note. Avg_w stands for the average brainwave value when the participants were in the working state. Avg_i is for when they were in the idle state.*
showed that the attention level of Challengers became significantly higher \( (p = .014) \) when changing from idle \( (M=47.47, SD=4.67) \) to working \( (M=47.87, SD=4.45) \); for Hosts, it became significantly lower \( (p < .001) \) when changing from working \( (M=48.53, SD=6.37) \) to idle \( (M=47.88, SD=5.83) \). For both Host and Challenger, the meditation level did not change significantly between the two states.

We further analyzed the relationship between roles and attention/meditation differences with brainwave data. For Challengers whose attention level increased, their meditation level decreased marginally significantly \( (p = .09) \). For Challengers whose attention level decreased, their change in meditation level was significantly higher \( (p = .04) \). For Challengers whose meditation level increased, their attention level decreased significantly \( (p < .001) \). For Hosts whose meditation level increased, their attention level decreased marginally significantly \( (p = .09) \). For Hosts whose meditation level decreased, their attention level decreased significantly \( (p = .01) \). These results are shown in Table 7; no other significant differences were found.

### 5.3 Questionnaire analysis

The degree of flow for participants was evaluated using questionnaires via several measurements—enjoyment \( (M=5.94, SD=0.81) \), focused attention \( (M=4.93, SD=0.96) \), involvement \( (M=6.11, SD=0.67) \), time distortion \( (M=5.41, SD=1.35) \) and satisfaction \( (M=6.15, SD=0.87) \)—along with the overall value \( (M=5.71, SD=0.58) \). We calculated the correlations between the five indexes of flow (focused attention, enjoyment, involvement, time distortion, and satisfaction) and attention as well as meditation: we did not find any correlations. We also calculated the correlations between brainwave and score as well as the number of concepts selected. Again, we did not find any correlations. We further investigated whether a relationship exists between the number of concepts selected by each participant and the four subsets of brainwave data, including working attention, idle attention, working meditation, and idle meditation. The results also revealed no relationship between these factors and the brainwave data subsets.

Despite the inconclusive correlation analysis results, we did find some interesting results when we used partial least squares analysis (PLS). We used PLS because the correlations analyses considered the flow theory indexes separately; however, the indexes should be considered when representing the flow. We used PLS to investigate whether flow influences the number of concepts selected, the score, or the time duration in Table 7.

Considering all participants, the results show that flow positively influenced the number of concepts selected \( (p < .001) \), score \( (p < .001) \), and time \( (p < .01) \). We further focused on four data subsets: Hosts whose meditation increased/decreased when changing from working to idle, and Challengers whose meditation increased/decreased when changing from idle to working (see Table 8). The results showed that for Hosts whose meditation decreased, their flow positively influenced their score \( (p < .001) \). For Hosts whose meditation increased, their flow positively influenced the number of guesses \( (p < .01) \). For Challengers whose meditation decreased, their flow positively influenced the number of concepts selected \( (p < .001) \), score \( (p < .001) \), and the number of guesses \( (p < .001) \). For Challengers whose meditation increased, their
Table 7  Breakdown of brainwave analysis

|                  | Challenger Decreased meditation |  | Challenger Increased meditation |  | Host Decreased meditation |  | Host Increased meditation |  |
|------------------|---------------------------------|---|---------------------------------|---|---------------------------|---|---------------------------|---|
|                  | p-value | Avg_w | Avg_i | p-value | Avg_w | Avg_i | p-value | Avg_w | Avg_i | p-value | Avg_w | Avg_i | p-value | Avg_w | Avg_i |
| Attention        | 0.56     | 47.61  | 47.46 | <0.001  | 47.85  | 47.23 | 0.01     | 48.34  | 47.56 | 0.09     | 48.75  | 48.34 |
| Meditation       | 0.04     | 53.61  | 54.05 | 0.09    | 54.06  | 53.53 | 0.79     | 54.10  | 54.06 | 0.75     | 55.06  | 55.01 |

*Note.* Avg_w stands for the average brainwave value when the participants were in the working state. Avg_i is for when they were in the idle state.
flow positively influenced their score ($p<.001$). The rest of the analysis revealed no significant influence.

6 Discussion and conclusion

This study proposed Coding Peekaboom, a gaming system to investigate whether a crowdsourced tagging approach is an adequate indexing mechanism to engage crowds to collect concepts of code segments. Our empirical results indicate that Coding Peekaboom effectively collects the collaborative tags (i.e., quality concepts of a code segment); further brainwave analysis and flow questionnaire results reveal that participants were engaged in the tagging task via Coding Peekaboom. Several interesting results were found:

1) The concepts collected by Coding Peekaboom are of reasonable quality in terms of correctness. Participants in Coding Peekaboom did not blindly assign concepts to the code segments. Under the strictest condition, based on the intersection of three experts, the concepts assigned by the participants achieved a precision rate of 0.79 and a sensitivity rate of 0.65 (see Table 5). Although the participants were not experts (e.g., students who only learned Java basic syntax before) in the Java programming language, they were still able to generate high-quality concepts in the assigned code segments. The false alarm rate among the three conditions indicates that only a few concepts selected by a small group of participants were not selected by the experts (see Table 5). This is strong evidence that programming concepts are appropriately collected by Coding Peekaboom, the proposed collaborative mechanism.

2) Brainwave analysis demonstrates that attention exhibits significant changes from the working ($M=48.13$, $SD=5.48$) to idle ($M=47.62$, $SD=5.25$) states (see Fig. 6). As expected, the different roles in Coding Peekaboom led participants to shift their attention in different ways. For example, when players immersed themselves in the working status, their attention increased. In comparison with attention, meditation is a weak indicator for specifying status change. Further analyses of brainwaves regarding user roles and attention/meditation changes confirm that Hosts decreased attention and increased meditation from working to idle status. However, the results also revealed that several Hosts possessed increased attention and decreased meditation when shifting from working to idle status, which suggests that they were aggressively participating in their task as well as their partners’ guessing processes (see Fig. 6 (c), Table 6). In addition, as expected, Challengers with contrary attention and meditation could be those who were concentrating on familiarizing themselves in the gaming environment so they could efficiently guess what the Hosts selected. The result shows that different individuals have different purposes for the same task (see Fig. 6 (d)). We might regard idle as a relaxing time, however, participants with hardworking passion would also be exhausted while waiting for the result with a higher score expected. In comparison, those who enjoy the game from beginning to the end might just show out they are only having fun without specific achievement,
and also be less stressed. As above mentioned, showing different attitudes on the same task can be further studied with personal traits.

3) The post-experiment questionnaire collected participant feedback, revealing that the perceived flow experience averaged 5.71 points (SD:0.58) on a 7-point scale questionnaire. The high rating indicates that participants were immersed in ample flow states during the experiment, demonstrating that Coding Peekaboom not only effectively encourages users to concentrate on the assigned tasks but also allows participants to enjoy the developed collaborative platform.

4) The perceived flow positively influenced participant performance in terms of the number of selected concepts, the task score, and the number of guesses. For Hosts whose meditation increased from working to idle status, their flow experiences positively influenced the number of guesses, indicating they were fully immersed in the game-like environment. For Hosts whose meditation decreased from working to idle, their flow experiences positively influenced task scores, indicating that these participants were concentrated on scoring points (see Table 8). For Challengers whose meditation decreased, their perceived flow positively influenced the number of selected concepts, task score, and interaction time; whereas for some Challengers whose meditation increased (indicating immersion in the game), their flow positively influenced score (see Table 8). In other words, there is nothing that would make the score increase, different states with different personalities might also end up scoring high scores. No matter how participants choose their purpose in the game, somehow hard workers would end up the same as the ones who enjoy the most.

5) These results suggest that meditation is not a strong indicator of flow experience. We argue that this finding originates from the definition of alpha waves (meditation). As mentioned, alpha waves are dominant during meditative or relaxed states, when people are less alert or act slower to stimuli. However, if people are in the flow, that inherently means that they are working on something. Therefore, it should be difficult for people to be in the flow when they are relaxed and not thinking about things. Even though meditation does not directly reflect flow, our results do show that meditation can influence participant performance. While various studies which investigate meditation focus on its relationships with the mental health (Desai et al., 2015b; Kan & Lee, 2015; Peng et al., 2011), we argue that these relationships played a role in influencing performance. If participants were stressed or depressed when using Coding Peekaboom, it is unlikely that they would outperform their counterparts.

| Table 8 PLS results between flow and performance measurements among different groups |
|--------------------------------|----------|----------|----------|----------|
|                                | All      | HMD      | HMI      | CMD      |
| Number of concepts             | ***3.503(+) | 1.707 (+) | 1.101 (+) | ***9.96(+) | 1.015(+) |
| Score                          | ***4.489(+) | ***7.343(+) | 0.732 (-) | ***7.216(+) | ***4.19 (+) |
| Number of guesses              | ***2.942 (+) | 1.69 (-) | ***3.101 (+) | ***4.578(+) | 0.619(+) |

Note. HMD stands for Host meditation decreased, HMI stands for Host meditation increased, CMD stands for Challenger meditation decreased, and CMI stands for Challenger meditation increased.
In short, empirical evaluations prove that Coding Peekaboom is effective in encouraging users to concentrate on the collaborative framework. In terms of the quality of concepts, the gaming mechanism can be used in specific domains to collect programming language-related concepts. Non-expert participants with basic programming experience can quickly produce high-quality, reliable programming concepts for code segments. Although Coding Peekaboom has not gone online, the offline results shed light on the power to encourage participants to engage in such a gaming context. Although brainwave activities do not correspond exactly to flow states, this result still reflects participants’ experience in the game. The experience of flow is a situation with multiple combinations instead of a single or pure indicator of attention and meditation. More effort will be needed to measure and investigate flow using physical and psychological methods.

**Funding** This research was supported in part by MOST 110-2410-H-004-026-MY3 to the first author, and MOST 109-2410-H-004-067-MY2 to the second author.

**Data Availability** The datasets used during the current study are available from the corresponding author on reasonable request.

**Declarations**

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

**References**

Ali, F. (2016). Hotel website quality, perceived flow, customer satisfaction and purchase intention. *Journal of Hospitality and Tourism Technology, 7*(2), 213–228

Allam, H., Bliemel, M., Spiteri, L., Blustein, J., & Ali-Hassan, H. (2019). Applying a multi-dimensional hedonic concept of intrinsic motivation on social tagging tools: A theoretical model and empirical validation. *International Journal of Information Management, 45*(April 2018), 211–222

Altebarmakian, M., & Alterman, R. (2019). Cohesion in online environments. *International Journal of Computer-Supported Collaborative Learning, 14*(4), 443–465. https://doi.org/10.1007/s11412-019-09309-y

Bilgihan, A., Nusair, K., Okumus, F., & Cobanoglu, C. (2015). Applying flow theory to booking experiences: An integrated model in an online service context. *Information & Management, 52*(6), 668–678

Bilgihan, A., Okumus, F., Nusair, K., & Bujisic, M. (2014). Online experiences: Flow theory, measuring online customer experience in e-commerce and managerial implications for the lodging industry. *Information Technology & Tourism, 14*(1), 49–71

Cha, V. (2020). Asia’s COVID-19 lessons for the west: Public goods, privacy, and social tagging. *Washington Quarterly, 43*(2), 33–50

Cowley, B., Charles, D., Black, M., & Hickey, R. (2008). Toward an understanding of flow in video games. *Computers in Entertainment, 6*(2), 20:1–2027

Csikszentmihalyi, M. (2000). *Beyond Boredom and Anxiety*. Jossey-Bass

de Melo, F. A. C. S. (2017). Flow-Z: A flow-based adaptable game to maintain optimal challenge. Tecnico Lisboa

Desai, R., Tailor, A., & Bhatt, T. (2015a). Complementary Therapies in Clinical Practice Effects of yoga on brain waves and structural activation: A review. *Complementary Therapies in Clinical Practice, 21*(2), 112–118

Desai, R., Tailor, A., & Bhatt, T. (2015b). Effects of yoga on brain waves and structural activation: A review. *Complementary Therapies in Clinical Practice, 21*(2), 112–118
Linstead, E., Hughes, L., Lopes, C., & Baldi, P. (2009). Exploring Java software vocabulary: A search and mining perspective. In 2009 ICSE Workshop on Search-Driven Development-Users, Infrastructure, Tools and Evaluation (pp. 29–32). IEEE

McMahan, T., Parberry, I., & Parsons, T. D. (2015). Evaluating Player Task Engagement and Arousal Using Electroencephalography. Procedia Manufacturing, 3(Ahfe), 2303–2310. https://doi.org/10.1016/j.promfg.2015.07.376

Morschheuser, B., Hamari, J., Werder, K., & Abe, J. (2017). How to gamify? A method for designing gamification. In Proceedings of the 50th Hawaii International Conference on System Sciences (2017) (pp.1298–1307)

Nacke, L. E., & Lindley, C. A. (2008). Flow and immersion in first-person shooters: Measuring the player’s gameplay experience. Proceedings of the 2008 Conference on Future Play: Research, Play, Share, 81–88

Neurosky Inc (2009). NeuroView User ’s Guide

Novak, T. P., Hoffman, D. L., & Yung, Y. F. (2000). Measuring the customer experience in online environments: A structural modeling approach. Marketing Science, 19(1), 22–42

Pavlo, P., Davis, F., & Dimoka, A. (2007). Neuro-IS: The potential of cognitive neuroscience for information system research. In Proceedings of the 27th International Conference on Information Systems (ICIS) (p.122). Montreal, Canada

Pelet, J., Ettis, S., & Cowart, K. (2017). Optimal experience of flow enhanced by telepresence: Evidence from social media use. Information & Management, 54(1), 115–128

Peng, H., Hu, B., Liu, Q., Dong, Q., Zhao, Q., & Moore, P. (2011). User-centered depression prevention: An EEG approach to pervasive healthcare. In 2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops (pp.325–330). IEEE

Piteira, M., Costa, C. J., & Aparicio, M. (2018). Computer programming learning: How to apply gamification on online courses? Journal of Information Systems Engineering & Management, 3(2), 1–9

Plotnikov, A., Stakheika, N., De Gloria, A., Schatten, C., Bellotti, F., Berta, R., & Ansovini, F. (2012). Exploiting real-time EEG analysis for assessing flow in games. Proceedings of the 12th IEEE International Conference on Advanced Learning Technologies, ICALT 2012, 688–689. https://doi.org/10.1109/ICALT.2012.144

Riedl, R., Banker, R. D., Benbasat, I., Davis, F. D., Dennis, A., Dimoka, A., & Weber, B. (2010). On the foundations of NeuroIS: reflections on the Gmunden Retreat 2009. Communication of the Association for Information Systems, 27(15), 243–264

Rong, W. J., & Min, Y. S. (2005). The effects of learning style and flow experience on the effectiveness of e-learning. In Fifth IEEE International Conference on. Advanced Learning Technologies (ICALT 2005) (pp.802–805)

Shrestyuk, A. Y., Kasinathan, K., Karapooninott, V., Knight, R. T., & Gurumoorthy, R. (2019). Individual EEG measures of attention, memory, and motivation predict population level TV viewership and Twitter engagement. Plos One, 14(3), 1–27

Shum, S. B. (2008). Cohere: Towards web 2.0 argumentation. COMMA, 8, 97–108

Shum, S. B., & Ferguson, R. (2012). Social learning analytics. Journal of Educational Technology & Society, 15(3), 3–26

Skadberg, Y. X., & Kimmel, J. R. (2004). Visitors’ flow experience while browsing a web site: its measurement, contributing factors and consequences. Computers in Human Behavior, 20(3), 403–422

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. Journal of Communication, 42(4), 73–93

Tiago-Costa, E., Queillas-Costa, E., & Santos-Baptista, J. (2016). Changes in EEG amplitude (Alpha and Beta waves) with Thermal environment. Dyna, 83(197), 87–93

Toth, A., & Tovolgyi, S. (2017). The introduction of gamification: A review paper about the applied gamification in the smartphone applications. In 7th IEEE International Conference on Cognitive Informatics,CogInfoCom 2016 - Proceedings (pp.213–217)

Trant, J. (2009). Studying social tagging and folksonomy: A review and framework. Journal of Digital Information, 10(1), 1–44

Ülker, B., Tabakcioğlu, M. B., Çizmeci, H., & Ayberkin, D. (2017). Relations of attention and meditation level with learning in engineering education. In Proceedings of the 9th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2017 (pp.1–4)

von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. In Proceedings of the 2004 conference on Human factors in computing systems - CHI ’04 (pp.319–326)
von Ahn, L., Ruoran, L., & Manuel, B. (2006). Peekaboom: A game for locating objects in images. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’06* (pp.55–64)

Wal, T., & Vander (2005). Folksonomy coinage and definition. Retrieved August 22, 2012, from http://vanderwal.net/folksonomy.html

Wang, C. C., & Hsu, M. C. (2014). An exploratory study using inexpensive electroencephalography (EEG) to understand flow experience in computer-based instruction. *Information & Management, 51*, 912–923

Wilson, C., Sudol, L. A., Stephenson, C., & Stehlik, M. (2010). *Running on empty: The failure to teach K-12 computer science in the digital age*. ACM

Yager, S. E., Kappelman, L. A., Maples, G. A., & Prybutok, V. R. (1997). Microcomputer playfulness: Stable or dynamic trait? *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 28*(2), 43–52

Yang, X., Cheng, P. Y., Lin, L., Huang, Y. M., & Ren, Y. (2019). Can an Integrated System of Electroencephalography and Virtual Reality Further the Understanding of Relationships Between Attention, Meditation, Flow State, and Creativity? *Journal of Educational Computing Research, 57*(4), 846–876

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.