A Qualitative Study on How Students Interact with Quizzes and Estimate Confidence on Their Answers

Kartik Shah
kdshah@princeton.edu
Princeton University
Princeton, New Jersey, United States

Priscilla Lee
palee@alumni.princeton.edu
Princeton University
Princeton, New Jersey, United States

Daphne Barretto
daphnegb@princeton.edu
Princeton University
Princeton, New Jersey, United States

Soohyun Nam Liao
sooyunl@cs.princeton.edu
Princeton University
Princeton, New Jersey, United States

ABSTRACT

Many prior work from various disciplines, including computing education, investigated how students interact with quizzes and how their interactions impact their learning outcomes. However, most of their results are based on quantitative analysis which does not offer detailed nuances behind the actual behaviors of students. Some prior studies took a qualitative approach but they focused mainly on students' perception on quizzes in general or student behaviors strictly while they work on the quizzes. Thus, our work conducted a qualitative study on how students interact with quizzes in a broader context, by also including their motivation for doing quizzes and the next steps they take after completing a quiz. By investigating observed student behaviors from the interviews with respect to their performance on the midterm exam, our results revealed a variety of student interactions and how they are related to performance. Our findings also provide some suggestions on how instructors should engage students with quizzes for effective learning.

CCS CONCEPTS

• **Social and professional topics** → **Computing education**.

KEYWORDS

self-assessment, quiz, metacognition, student behaviors, follow up, motivation, confidence

ACM Reference Format:
Kartik Shah, Priscilla Lee, Daphne Barretto, and Soohyun Nam Liao. 2021. A Qualitative Study on How Students Interact with Quizzes and Estimate Confidence on Their Answers. In *26th ACM Conference on Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2021), June 26-July 1, 2021, Virtual Event, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3430665.3456377

ITiCSE 2021, June 26-July 1, 2021, Virtual Event, Germany

ACM ISBN 978-1-4503-8214-4/21/06…$15.00
https://doi.org/10.1145/3430665.3456377

1 INTRODUCTION

A quiz is one of the most common course components, and there is plenty of prior research from various disciplines about a positive correlation between doing quizzes and student learning outcomes [1–3, 7, 8, 17, 21, 23, 30, 31, 33, 39]. Additionally, some prior work focused on students' perceptions of quizzes or their problem solving behaviors in general [4, 11, 12, 19], but there has not been much work analyzing student behaviors when taking quizzes, which is surprising given how beneficial quizzes are.

Therefore, we conducted student interviews to better understand how students interact with quizzes. We looked at multiple factors during the interviews, including their question solving behaviors, their motivations of doing quizzes, when they tend to complete quizzes, and any next steps they take after completing quizzes. Additionally, the results from our own prior work [25] suggested that students' self-reported confidence estimates on their answers may not be a good proxy of student metacognition, and hence we also investigated the factors students consider when estimating confidence levels on their answers.

By interviewing 12 students who used our previously proposed quizzes [25] in a CS1 course and analyzing their performance on the midterm exam, we found that strong students tend to complete quizzes routinely, have a specific follow-up strategy, and do not use resources simply to find an answer. We also observed that most students, regardless of performance on the midterm, did not rigorously use metacognition to choose confidence levels. This could explain why our prior work showed that students who did better on exams were just as susceptible to inaccurately choosing their confidence levels.

Our results could be useful to instructors who want to use quizzes in their courses, since we have outlined a few student interactions that seem to be correlated with better exam performance. Lastly, our results about how students choose confidence levels suggest that the confidence levels on our quizzes may not be a good proxy to measure metacognitive ability. This suggests future work to investigate ways to prompt students to use metacognition on such quizzes.

2 RELATED WORK

This section first introduces prior research studies on quizzes in general. Then we describe details of our previously designed quiz named Compass, since it partly motivated our research question.

2.1 Prior Work on Quizzes

Many prior work investigated how students interact with quizzes and how their interactions impact their learning outcomes. They are found in computing education [7, 8, 11, 17, 25, 39] as well as in other STEM education [1, 3, 11, 21, 23, 31, 33]. For example, several studies have found that frequent engagement with quizzes is positively correlated with student learning outcomes [1, 3, 7, 8, 11, 17, 21, 23, 25, 30, 31, 33, 39]. In addition, Gholami and Zhang found that the best performing students often start working on the quizzes the earliest [17]. However, their findings are mostly based on quantitative analysis which often lacks detailed nuances behind the actual student behaviors.

Some prior studies qualitatively investigated student interaction with quizzes. For example, Zingaro et al. extracted a variety of sentiments from student reflections on pre-class reading quizzes [39]. Additionally, some prior work found that quizzes help students to judge their strengths and weaknesses more accurately (i.e. better metacognition) [4, 11]. Moreover, a big part of student interaction with quizzes involves how they approach solving the quiz questions themselves. This is generalized as students' problem solving behaviors [9, 12, 19, 28]. According to Harper, experts view problem-solving as a process while novices consider it a recall task [19]. Moreover Chi et al. ran a study where they asked students to study worked examples and then asked them to solve similar questions [12]. All students had similar scores on a sample test before the study, but they found that students who did better on the questions took more time initially to study the worked examples, but once they moved to solving the questions, they only referred back to worked examples for specific clarifications. On the other hand, students who did not do as well on the questions did not spend as much time studying the worked examples, and tended to go back and re-read large chunks of the worked examples when solving questions [12]. This suggests that when doing self-assessment quizzes, stronger students should not refer to resources, or if they do, they should take time to understand it. In summary, these qualitative studies focused either on how students perceive quizzes as a course component or on how students approach questions in general. Hence, a contribution of our work is to explore a variety of student interactions with quizzes - including why they do them, when they do them, how they approach questions, and the next steps they take after completing a quiz - through a qualitative study to gain insights on effective ways of utilizing quizzes as a course component.

2.2 Compass

Compass [25] is a fully-automated optional online quiz aimed at improving students' metacognition. It was motivated by prior work which found that metacognition is critical for student success in computing [16, 26, 27, 29]. It aims to do so by asking students to self-report their confidence estimate for each answer, based on a prior study which found that confidence estimates reflect student metacognition [27]. Specifically, after each question, students are asked 'How confident are you in your response', and they respond on a 5 point Likert Scale from 'Very Unsure' to 'Very Confident'. These confidence-estimating questions are also used to generate a personalized resource guide for each student that categorizes concepts in order of study priority, and includes links to relevant resources for each concept. The priority is decided based on a student's quiz performance and confidence estimates for each learning objective. Examples of resources in the resource guide include online lectures with specific timestamps, specific textbook pages, worksheets introduced in class meetings, past exam questions and additional online practice questions.

Table 1: List of Interviewees and Their Performance on The Midterm Exam (H-: High-performers, M-: Middle-performers, L-: Low-performers)

Student ID	Midterm Percentile (%)
H1	96
H2	88
H3-Outlier	73
H4	72
M1	62
M2	48
M3	47
L1	31
L2-Outlier	20
L3	20
L4	14
L5	9

From the student survey administered at the end of the semester, it seemed like students found our quizzes helpful to identify what they knew and did not know. However, our prior work did not find evidence of improved metacognitive accuracy (defined as the relationship between confidence and correctness) over time [25]. Hence, in this work we also investigate how students choose their confidence levels for our quiz questions and whether confidence estimates seems to be a good proxy to improve student metacognition.

3 METHODS

In this section, we list our research questions and describe how we administered and analyzed the student interviews.

3.1 Research Questions

Our research question is described below. Under this RQ, we observed student interaction based on the five listed themes below.

- **RQ:** How do students interact with quizzes that include self-reported confidence estimates?
  - Why do students decide to attempt quizzes? *(Motivation)*
  - When do students usually attempt quizzes? *(Attempt date)*
  - How do students approach solving quiz questions? *(Question solving behaviors)*
  - What are the next steps taken by students once they complete a quiz? *(Follow-up)*
  - Do students utilize metacognition to indicate their confidence level on a quiz question? *(Metacognition)*

3.2 Student Interviews

The interviews were run at a North American research-intensive university in summer 2020 upon the approval from the Institutional Research Board. We recruited 12 participants from those who had completed the university's CS1 course in Spring 2020, which is when we administered our quizzes for the course. Our participants were a random sample and we interviewed every student who expressed their interest, and included them in the analysis.

All interviews were conducted virtually over Zoom [38]. The interview started with a student demo where they shared their computer screen and completed a sample quiz that was in the format of our quizzes described in Section 2.2. During the demo, the students were also asked to think aloud so that interviewers could better understand their thought process. The demo was included so we could observe student behaviors when taking our quiz live, which would offer another layer on analysis in addition to relying on their answers to our follow-up questions, especially to analyze *Question Solving Behaviors* and *Follow-up*. After the demo, the interview had a semi-structured session where the interviewers first asked students any questions or clarifications about any of their actions during the demo. After, the interviewers also went over the list of pre-selected questions about the students' experience with our quizzes during the semester. These questions were related to the themes we aimed to investigate, as detailed in Section 3.1. The pre-selected questions are listed below.

- Did you find yourself using quizzes for a specific purpose?
- How seriously did you take the quizzes?
- Did you use any of the resources in the email after taking a quiz? If yes, how often and how useful was it?
- Did you learn something new from these quizzes?
- Out of all the components in the quizzes, what was the most useful for you?
- How difficult were these quizzes compared to the midterm and final?
- Did you read the feedback on the quiz after getting a question wrong? If yes, how often and how useful was it?
- Was there a particular confidence level you tended to gravitate towards?
- What factors affected your choice of confidence levels?

3.3 Analysis

After the interviews, we transcribed each interview using the recorded audio files and analyzed the transcripts. We used the open-coding approach [24] for the analysis. To come up with a reliable codebook, two authors of this paper iterated the process of coding the same three transcripts (25% of all transcripts) independently and discussing the discrepancies so that tags were standardized. After multiple rounds we achieved a Cohen's kappa value of 0.86.

The resulting codebook consisted of 18 codes which represented student feelings and behaviors related to our quizzes. Once all transcripts were coded, we extracted specific student comments that were relevant to our research questions. We added one of the codes, Behavior-Refer to Other Resourcesto our initial codebook after watching the recorded videos of the demo portions. These behaviors were not illustrated on the transcripts, as they occurred nonverbally during the demo. To avoid potential inconsistency of this specific code across different coders, only one author coded this behavior. The list of codes and how they were used is described below. Four Feeling- codes were not utilized for this work as they showed weak relevance to our research questions. They were Feeling-Seriousness, Feeling-Difficulty, Feeling-Pressured, and Feeling-Patience.

- Motivation
  - Feeling-Purpose
- Attempt Date
  - Feeling-Perception of Resource
- Question solving behaviors
  - Behavior-Refer to Other Resources
  - Behavior-Refer to My Notes
  - Behavior-Give up
  - Behavior-Trace the Code
- Follow-up
  - Behavior-Go to Office Hours
  - Behavior-Make Own Notes
  - Behavior-Practice Code
  - Behavior-Practice Questions
  - Behavior-Read Course Material
  - Behavior-Read Answer Feedback
  - Behavior-Read External Resources
- Metacognition
  - Behavior-Choose Confidence Level

We explored common quiz behaviors through the student quotes for relevant codes, and we investigated whether those behaviors seem to be correlated with student performance. To do so, we grouped students based on their percentile score on the midterm exam. To avoid any bias during the coding process, we referred to the exam scores only after we completed coding the interview transcripts. The midterm exam was used because the midterm was the last in-person course assessment before the campus shutdown due to COVID-19. We categorized students into three groups, with 33% and 67% percentile cutoffs. The students in the bottom, middle, and top tercile were identified as Low-performers (L-), Middle-performers (M-), and High-performers (H-) respectively. We used these three groups to provide a more fine-grained analysis than simply splitting students into halves. Splitting students into quartiles was also infeasible since we only had 12 participants. Table 1 illustrates each interviewee's midterm exam score percentile and tercile-based ID.

Our data had two outliers, indicated as L2-Outlier and H3-Outlier. They were outliers since they had a significant difference in their performance between the midterm and final exams. The percentile of L2-Outlier increased from 20 to 80, whereas that of H3-Outlier decreased from 73 to 28. All other students had similar scores on both the midterm and final. We note that the two outliers appear to be exceptions to multiple general trends in Section 4. We discuss possible reasons for their appearances as exceptions to general trends and their notable performance differences in Section 5.2.

4 RESULTS

This section presents our findings for our research question by analyzing the demos and transcripts from the student interviews.

4.1 Motivation

Students mentioned a variety of motivations such as to receive additional help, get a sense on what an exam would look like, prepare for the programming assignments, and earn late day credits. We found that none of the High-performers were motivated by the late day credit, but 2 Middle-performers (M2 and M3) and 2 Low-performers (L1 and L3) mentioned the late day incentive as a motivating factor for doing the quizzes. For instance, M2 said that *"I was just completing them to get the late day credit, I think I wasn't putting as much thought into my answers"*. In contrast, H1 said that *"There was no question about me doing them even if I'm not actually using the incentive of the late day credit, just because it was an extra chance to practice these things."*

4.2 Attempt Date

All Middle- and Low-performers except for L2-Outlier and L5 seemed to perceive quizzes as just something to submit. They submitted the quizzes very close to the deadline, and usually after being reminded to do it through a reminder email. On the other hand, all High-performers submitted the quizzes routinely at a similar time each week and they appeared to think of the quizzes as a learning resource. For instance, H4 said that *"I did them before [the class meeting] so I knew what to ask my instructor."*

4.3 Question Solving Behaviors

During the demo section of the interviews, only H2 and H3-Outlier amongst High-performers referred to course resources while taking quizzes even though all students were permitted to; the rest of the High-performers did not. However, H2 later clarified that they actually did not refer to resources during the semester when they were actually taking the quizzes, saying that *"when I'm taking these quizzes [during the semester], I would remember better and I usually don't have to refer to these resources"*. This indicates that all High-performers except H3-Outlier did not refer to resources when completing the quizzes. Conversely, all Middle- and Low-performers except for L1 and M3 referred to resources when solving the quiz questions.

Most students who referred to resources were Middle- and Low-performers (i.e. H3-Outlier and all Middle- and Low- performers except for L1 and M3) and they used the resources for most questions during the demo. Additionally, from the demo, it seemed like all these students, apart from L2-Outlier, often utilized resources to browse through the lecture slides in an attempt to simply find an answer to the quiz questions; they did not take the time to carefully read the information on the slides to understand it. For instance, H3-Outlier searched keywords from quiz questions in the lecture slides to look for quiz answers, but it did not seem like they understood it even after selecting the answer. L2-Outlier, on the other hand, referred to the textbook for a few questions, but when they did, they re-read large chunks of information and took the time to ensure that they understood the relevant concepts.

Also from the demo, we found that four students, H1, H2, L2-Outlier, and M2 gave up on solving a question and moved on to a next one. However, the reasons of giving up seemed different between the former three students and M2. While M2 gave up because they were tired and lost patience, the former three found it more efficient to move on at the moment and use the quiz feedback and resource guides to fully understand the content. The representative quotes are: *"I'm just going to pick Option 3 because I'm tired of the question" (M2)*; *"I'm going to go with my gut on this one. If I need refreshing, I will get resources that will help me study for it" (L2-Outlier).*

In addition, with respect to questions that included code snippets during the demo, all High- and Middle-performers traced the code completely except for H3-Outlier, whereas only L2-Outlier did this among Low-performers. High-performers also tend to refer to their own notes to refresh their memory when they are stuck on a question, as H1, H2, L2-Outlier, and L3 mentioned that they utilized their own notes.

4.4 Follow-up

There was no particular trend on the exact follow-up behaviors with respect to student performance. However, we found some differences in terms of how actively and specifically students followed up on the quiz feedback.

Middle-performers demonstrated the weakest follow-up. M1 only skimmed through the answer feedback but did not take any further steps to resolve their confusions; M3 only skimmed through the answers to check what they got right and what they got wrong, and they did not even read the answer feedback. Moreover, M2 said they would try to follow up, but they did not really put in the time or effort to ensure that all their doubts and misconceptions are cleared: *"I would read the feedback if it was something I didn't know at all or I thought I got it right... Wouldn't read feedback if it was a silly mistake. If I still don't understand the feedback, I may go to the textbook. If that still doesn't work, I would just accept that's the answer and move on"*.

All High- and Low-performers demonstrated some specific actions of follow-up, except for H3-Outlier and L1 (whose percentile score (31%) was very close to the Middle-performers). To be specific, H3-Outlier and L1 read the answer feedback for questions they got wrong, and said they *might* open the resource guide and click on a link. L1 justified this by saying *"it takes a lot more energy or effort to go and click through the links and find something that looks like super related but isn't a hit."*

However, we found that H1, H2, H4 and L2-Outlier had more specific follow-through strategies than L3, L4 and L5 (recall that L1, H3-Outlier and all the Middle-performers did not even have specific follow-up strategies). By this, we mean that the former group of students had a clear plan of what they wanted to do. For instance, H4 described their follow-through process as *"I wouldn't watch the lecture video since I have already watched it before, so I look for resources I haven't seen before. I haven't looked at the textbook, so I would go to the page [that the resource guide pointed out]. I would usually do this in the morning, and in the afternoon...go through the theory in that section... I think I got [understood] it more clearly, let's go to questions from exam so I know what kind of questions can be asked from these topics. Look at more recent ones because they are more likely in a similar format. Would also look at other ones because I would still like to understand topic better."* Additionally, H1's follow-up process consisted of reading the answer feedback then going to Office Hours to clarify doubts. H2 also read answer feedback and then asked friends in upper-level CS classes to address any confusions. They went back to the textbook if there was still something they were confused about after this. Lastly, L2-Outlier read the quiz feedback, then read the textbook, and then worked on practice questions as well.

On the other hand, students in the latter group (L3, L4, and L5) did not have as specific of a plan. For instance, L4 said *"I make flashcards sometimes, other times I go to the booksite and see if I can do some practice code to reinforce it."* Even though this student said they always read the answer feedback and the textbook first, they were unsure about what the last step of their follow-up was, and did not have a consistent follow-up process.

4.5 Metacognition

Students tended to choose confidence levels for quiz questions based either on intuition/general feeling, or whether they could reason through their answers. Also, four (M1, M3, L3 and L5) of the Middle- and Low-performers on whom we have data about how they chose their confidence levels said that they were never confident. For instance, L3 said that *"I'm never 100% confident with my answer, which is more of a "me" thing than it is anything else."*, and a similar sentiment was echoed by M1, M3 and L5 too. We had no data on M2 for this research question as their interview ran out of time.

In contrast, no High-performers said they were never confident. Still, they usually used vague indicators such as how they feel about the topic, the time they spend on a question, and even their general feeling on a day for marking down confidence levels on specific quiz questions. For instance, H2 usually relied on their perception of a topic when marking question-specific confidence levels, saying *"But on the question with the running times [one of the topics] especially, I would just...Like if you, even if you give me the right answer, I'm still unsure"*.

5 DISCUSSION

We have summarized notable observations from our results below. The rest of this section discusses the implications of our results along with suggestions for how to engage students with quizzes.

- Most Middle- and Low-performers were motivated by extrinsic factors, such as late day credit or a reminder email
- High-performers submitted the quizzes routinely at a similar time each week
- High-performers tend to not refer to course resources while solving quiz questions. However if they do, they focus on understanding the underlying concepts behind a quiz question, while Middle- and Low-performers tend to use resources more often during a quiz and when they do they simply look for answers to quiz questions.
- High-performers actively followed up on a quiz feedback, and they also had a well-defined follow-up strategy after completing a quiz
- Most students chose their confidence level without utilizing metacognition rigorously regardless of their performance

5.1 Confidence and Student Metacognition

The student interviews gave us further insight into how students mark their confidence levels. Since our interviews were conducted after the semester was over, students should have been familiar with the material that was tested on the sample quiz students were asked to complete during the interview. Hence, we expected that students, especially High-performers, would mark their confidence levels accurately. However, our results suggest that all students mostly marked confidence levels intuitively, suggesting that the confidence level questions did not prompt students to adequately reflect on their content knowledge. Additionally, considering that 4 out of 7 Middle- and Low-performers said they are never confident is also contrary to prior work, including our own, that suggests weaker students are usually overconfident [20, 25, 27]. It is possible that this was simply an anomaly since ours was a small-scale study, but our observations of how students choose their confidence levels overall seems to suggest that at least the confidence level questions used in our quizzes are not an accurate proxy for measuring metacognitive ability. This brings up our future work of a more accurate way of gauging student metacognition. An easy alternative could be asking students to estimate an objective value such as the number of correct answers rather than a subjective value like the level of confidence itself to a question, since different students may have different interpretations of words like 'Unsure', or might have different baseline confidences.

5.2 The Two Outliers

We noticed that H3-Outlier and L2-Outlier were exceptions to most of the general trends, with H3-Outlier's behaviors resembling Middle- and Low-performers, while L2-Outlier's behaviors were similar to High-performers'. For instance, L2-Outlier had a specific, active follow-up strategy and had a routine for when they completed quizzes, not waiting until the last minute or the reminder email to submit them. Additionally, L2-Outlier never used resources while doing quizzes with the purpose of just finding an answer, but they actually took the time to read and understand the concept before choosing an answer to the quiz question. On the other hand, H3-Outlier did not exhibit an active follow-up process, and often used resources when doing quizzes with the intention of just finding an answer and moving on, without actually stopping to take the time to ensure they understand the concepts.

This motivated us to look at their exam scores in more detail, and we actually found that, in the final exam, L2-Outlier's percentile score was 80, which was a significant jump from the percentile score of 20 they received on the midterm. In contrast, H3-Outlier had a percentile score of 28 on the final, which was a drastic drop from the percentile score of 73 they received on the midterm. This prompted us to look at their interviews in more detail, and we found that H3-Outlier was a student that had prior experience in CS before taking this course, whereas L2-Outlier was a student who had no prior experience.

Given this, we hypothesize that H3-Outlier did well on the midterm due to their prior experience and not due to the fact that they had good study behaviors. This is consistent with prior work that has suggested that prior experience is correlated with student performance in CS1 [5, 10, 13, 18, 22, 32, 34, 36, 37]. Their poor behaviors when interacting with quizzes could explain why they did poorly on the final, by which time we suspect their prior experience was no longer a significant advantage. Prior work has actually shown that although students with prior experience have a significant benefit in the first course of a programming sequence, it does not have an impact in subsequent courses [22]. We suspect that the same could be true here too.

Although we are referring to two halves of the same course, our course is structured in such a way that the second half of the course mainly focuses on theory topics that are usually not covered in high school CS classes, and the final exam only tests topics from the second half of the course. Hence, their prior experience may not have helped them on the final, and their poor study behaviors when interacting with quizzes might explain why they did poorly on it.

On the other hand, L2-Outlier themselves admitted to feeling lost at the start of the course, saying it was *"very rough up until the midterm since I were a first-time coder."* They credited our quizzes for making them aware of how much they had to learn initially, which they felt they would not have realized from just watching lecture and attending classes. We believe that their quiz behaviors which resembled those of High-performers could have helped account for their improvement on the final exam. This is supported by the fact that the material tested on both exams was almost exclusive, although some topics tested on the final did build upon concepts learned in the first half of the semester. L2-Outlier's case shows how scary and daunting it can be for students with no prior experience to take CS1 that also has students with prior experience, even if their study behaviors mirror those of High-performers. Just as how unsurprising it is that this student did very well on the final exam given their study behaviors, it was surprising that they did poorly on the midterm exam. We suspect that there are two possible reasons for their midterm exam performance:

- The student felt pressured by peers who had prior experience. This is consistent with prior work by Tafliovich et al. who found that students felt stressed and intimidated by students who they perceived as "experts" [34].
- The student felt overwhelmed by the steep learning curve and pace of the class. This is also consistent with prior work by Holden and Weeden which found that the hurdle in introductory programming was seems to be learning basic concepts [22].

Our results further motivate designing CS1 in ways that do not significantly advantage students with prior experience, such as object-first CS1 [35]. An alternative, which has been used in many institutions, is simply placing students with and without prior experience in different classes, or identifying students without prior experience early and providing more support to them.

5.3 How Should Students Engage with Quizzes?

Based on our results, we make some suggestions for instructors on how to guide their students to engage with quizzes effectively. First, instructors should encourage students to not work on the quizzes close to the deadline, but to leave a sufficient amount of time to reason through questions carefully. Prior work too has shown that students who start earliest are often the strongest students [17]. Our results also suggest that High-performers seemed to have a set routine and did the quizzes at a similar time each week. However, we do not think that this is directly correlated with performance, but rather we suspect that this could just indicate stronger self-regulation, which have been shown to be correlated with performance [6, 14, 15]. Still, encouraging students to follow a routine to submit quizzes could help reinforce their self-regulation, but instructors could also consider incorporating explicit training of self-regulation in their courses, if feasible.

Second, we believe it is especially important for instructors to emphasize that quizzes are more than just quizzes; they are a resource that students can use to identify their shortcomings and confusions. Thus, students should view quizzes as a form of formative assessment and more importantly, submitting the quiz should not be the final stage of their study process. Instead, once they submit the quiz, they should read the quiz feedback and go back to the course resources to clarify any confusions that the quiz brings up. This could include revisiting the textbook, the lecture notes, or any other course material. They also should seek help to get their doubts cleared such as going to office hours.

Third, consistent with prior work in physics education [12], it seems as though students who did not refer to any resource during quizzes did better on exams. While it is possible that these were simply stronger students who did not need to refer to resources, we suspect that there are indeed benefits to taking the quiz as an assessment to get an accurate gauge of understanding on the course content. Thus, instructors should design quizzes in a way that allows students to revisit course resources until they fully understand the concepts then re-attempt the quiz. One example is to not administer these quizzes under timed conditions. We think that this approach, in combination with our previous suggestions, may encourage students to truly use such quizzes to identify and work on their conceptual gaps.

6 THREATS TO VALIDITY

Our results may not generalize to students in other universities or courses as they are based on a small-scale study and all our interview participants were from a single cohort of a single course at a single university. Some of our findings rely on their own self-reported comments, so they may not accurately reflect what they actually did when taking these quizzes during the semester. Lastly, the student interviews were administered during the summer vacation which was approximately 5 months after the midterm, so it is possible that a few behaviors may have developed or changed in this time frame. However, we believe this is not as critical concern as it may sound since most students had similar percentile scores between the midterm and final exam (which was only 2 months before the interview), except for the two outliers about whom we have discussed extensively in Section 5.2.

7 CONCLUSION AND FUTURE WORK

Through this study, we explored how students interact with quizzes in a broader context, observing motivations behind doing quizzes, when students completed quizzes, actual question solving behavior, and next steps after quiz completion. Additionally, we also explored how students estimated their confidence levels for quiz answers. By analyzing the extracted student interactions with respect to their midterm exam scores, we found relationships between certain interactions and exam performance. These relationships suggest that students should follow a routine when taking quizzes, revisit feedback to ensure content understanding, and take quizzes without resources for effective learning.

With our more nuanced analysis of student behaviors, the relationships we found between certain interactions and student performance can be validated in a larger-scale study. Additionally, since confidence levels on our quizzes do not appear to be a good proxy for metacognition, possible future work includes investigating how to encourage students to more effectively use metacognition in quizzes.

REFERENCES

[1] S. Angus and Judith Watson. 2009. Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. *Br. J. Educ. Technol.* 40 (2009), 255–272.

[2] Anonymous. 201X. Anonymized Paper 1. In *Proceedings of Anonymized.* ACM.

[3] Vasantha Aravinthan and Thiru Aravinthan. 2010. Effectiveness of self-assessment quizzes as a learning tool. In *Proceedings of Engineering Education Conference (EE 2010).* Higher Education Academy Subject Centres for Materials and Engineering.

[4] Syed Aziz. 2003. Online Quizzes for Enhancing Student Learning in a First Year Engineering Course. (08 2003).

[5] Susan Bergin and Ronan Reilly. 2005. Programming: Factors That Influence Success. In *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education* (St. Louis, Missouri, USA) *(SIGCSE '05).* Association for Computing Machinery, New York, NY, USA, 411–415. https://doi.org/10.1145/1047344.1047480

[6] Susan Bergin, Ronan Reilly, and Desmond Traynor. 2005. Examining the Role of Self-Regulated Learning on Introductory Programming Performance. In *Proceedings of the First International Workshop on Computing Education Research* (Seattle, WA, USA) *(ICER '05).* Association for Computing Machinery, New York, NY, USA, 81–86. https://doi.org/10.1145/1089786.1089794

[7] Peter Brusilovsky and Sergey Sosnovsky. 2005. Engaging students to work with self-assessment questions: A study of two approaches. In *Proceedings of the 10th annual SIGCSE conference on Innovation and technology in computer science education.* 251–255.

[8] Peter Brusilovsky and Sergey Sosnovsky. 2005. Individualized Exercises for Self-Assessment of Programming Knowledge: An Evaluation of QuizPACK. *J. Educ. Resour. Comput.* 5, 3 (Sept. 2005), 6–es. https://doi.org/10.1145/1163405.1163411

[9] E. W. Burkholder, J. K. Miles, T. J. Layden, K. D. Wang, A. V. Fritz, and C. E. Wieman. 2020. Template for teaching and assessment of problem solving in introductory physics. *Phys. Rev. Phys. Educ. Res.* 16 (Apr 2020), 010123. Issue 1. https://doi.org/10.1103/PhysRevPhysEducRes.16.010123

[10] Pat Byrne and Gerry Lyons. 2001. The Effect of Student Attributes on Success in Programming. In *Proceedings of the 6th Annual Conference on Innovation and Technology in Computer Science Education* (Canterbury, United Kingdom) *(ITiCSE '01).* Association for Computing Machinery, New York, NY, USA, 49–52. https://doi.org/10.1145/377435.377467

[11] Olle Bälter, Emma Enström, and Bernhard Klingenberg. 2013. The effect of short formative diagnostic web quizzes with minimal feedback. *Computers & Education* 60, 1 (2013), 234 – 242. https://doi.org/10.1016/j.compedu.2012.08.014

[12] Michelene T.H. Chi, Miriam Bassok, Matthew W. Lewis, Peter Reimann, and Robert Glaser. 1989. Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science* 13, 2 (1989), 145 – 182. https://doi.org/10.1016/0364-0213(89)90002-5

[13] Donald Chinn, Judy Sheard, Angela Carbone, and Mikko-Jussi Laakso. 2010. Study Habits of CS1 Students: What Do They Do Outside the Classroom?. In *Proceedings of the Twelfth Australasian Conference on Computing Education - Volume 103* (Brisbane, Australia) *(ACE '10).* Australian Computer Society, Inc., AUS, 53–62.

[14] Tim Cleary and Peter Platten. 2013. Examining the Correspondence between Self-Regulated Learning and Academic Achievement: A Case Study Analysis. *Education Research International* 2013 (01 2013). https://doi.org/10.1155/2013/272560

[15] Popa Daniela. 2015. The Relationship Between Self-Regulation, Motivation And Performance At Secondary School Students. *Procedia - Social and Behavioral Sciences* 191 (2015), 2549 – 2553. https://doi.org/10.1016/j.sbspro.2015.04.410 The Proceedings of 6th World Conference on educational Sciences.

[16] Anthony Estey and Yvonne Coady. 2017. Study Habits, Exam Performance, and Confidence: How Do Workflow Practices and Self-Efficacy Ratings Align?. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education.* 158–163.

[17] Arash Gholami and Larry Yueli Zhang. 2018. Student Behaviour in Unsupervised Online Quizzes: A Closer Look. In *Proceedings of the 23rd Western Canadian Conference on Computing Education.* 1–6.

[18] Dianne Hagan and Selby Markham. 2000. Does It Help to Have Some Programming Experience before Beginning a Computing Degree Program?. In *Proceedings of the 5th Annual SIGCSE/SIGCUE ITiCSEconference on Innovation and Technology in Computer Science Education* (Helsinki, Finland) *(ITiCSE '00).* Association for Computing Machinery, New York, NY, USA, 25–28. https://doi.org/10.1145/343048.343063

[19] Kathleen Harper. 2006. Student Problem-Solving Behaviors. *The Physics Teacher* 44 (04 2006). https://doi.org/10.1119/1.2186244

[20] Brian Harrington, Shichong Peng, Xiaomeng Jin, and Minhaz Khan. 2018. Gender, confidence, and mark prediction in CS examinations. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education.* 230–235.

[21] Debra Henly. 2003. Use of Web-based formative assessment to support student learning in a metabolism/nutrition unit. *European journal of dental education : official journal of the Association for Dental Education in Europe* 7 (09 2003), 116–22. https://doi.org/10.1034/j.1600-0579.2003.00310.x

[22] Edward Holden and Elissa Weeden. 2004. The Experience Factor in Early Programming Education. In *Proceedings of the 5th Conference on Information Technology Education* (Salt Lake City, UT, USA) *(CITC5 '04).* Association for Computing Machinery, New York, NY, USA, 211–218. https://doi.org/10.1145/1029533.1029585

[23] Jonathan Kibble. 2007. Use of unsupervised online quizzes as formative assessment in a medical physiology course: Effects of incentives on student participation and performance. *Advances in physiology education* 31 (10 2007), 253–60. https://doi.org/10.1152/advan.00027.2007

[24] Päivi Kinnunen and Beth Simon. 2010. Building Theory about Computing Education Phenomena: A Discussion of Grounded Theory. In *Proceedings of the 10th Koli Calling International Conference on Computing Education Research* (Koli, Finland) *(Koli Calling '10).* Association for Computing Machinery, New York, NY, USA, 37–42. https://doi.org/10.1145/1930464.1930469

[25] Priscilla Lee and Soohyun Nam Liao. 2021. Targeting Metacognition by Incorporating Student-Reported Confidence Estimates on Self-Assessment Quizzes. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (Virtual Event, USA) *(SIGCSE '21).* Association for Computing Machinery, New York, NY, USA, 431–437. https://doi.org/10.1145/3408877.3432377

[26] Soohyun Nam Liao, Sander Valstar, Kevin Thai, Christine Alvarado, Daniel Zingaro, William G Griswold, and Leo Porter. 2019. Behaviors of higher and lower performing students in CS1. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education.* 196–202.

[27] Murali Mani and Quamrul Mazumder. 2013. Incorporating metacognition into learning. In *Proceeding of the 44th ACM technical symposium on Computer science education.* 53–58.

[28] Jason W. Morphew, Gary E. Gladding, and Jose P. Mestre. 2020. Effect of presentation style and problem-solving attempts on metacognition and learning from solution videos. *Phys. Rev. Phys. Educ. Res.* 16 (Jan 2020), 010104. Issue 1. https://doi.org/10.1103/PhysRevPhysEducRes.16.010104

[29] Laurie Murphy and Josh Tenenberg. 2005. Do computer science students know what they know? A calibration study of data structure knowledge. In *Proceedings of the 10th annual SIGCSE conference on Innovation and technology in computer science education.* 148–152.

[30] Yasin Ozarslan and Ozlem Ozan. 2016. Self-Assessment Quiz Taking Behaviour Analysis in an Online Course. *European Journal of Open, Distance and e-learning* 19, 2 (2016), 15–31.

[31] M. Richards-Babb, Reagan Curtis, Zornitsa Georgieva, and John Penn. 2015. Student Perceptions of Online Homework Use for Formative Assessment of Learning in Organic Chemistry. *Journal of Chemical Education* 92 (08 2015), 150827072242001. https://doi.org/10.1021/acs.jchemed.5b00294

[32] Judy Sheard, Angela Carbone, Selby Markham, A J Hurst, Des Casey, and Chris Avram. 2008. Performance and Progression of First Year ICT Students. In *Proceedings of the Tenth Conference on Australasian Computing Education - Volume 78* (Wollongong, NSW, Australia) *(ACE '08).* Australian Computer Society, Inc., AUS, 119–127.

[33] Margaret Snooks. 2004. Using practice tests on a regular basis to improve student learning. *New Directions for Teaching and Learning* 2004 (12 2004), 109 – 113. https://doi.org/10.1002/tl.178

[34] Anya Tafliovich, Jennifer Campbell, and Andrew Petersen. 2013. A Student Perspective on Prior Experience in CS1. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education* (Denver, Colorado, USA) *(SIGCSE '13).* Association for Computing Machinery, New York, NY, USA, 239–244. https://doi.org/10.1145/2445196.2445270

[35] Jr. Ventura, Philip R. 2004. *On the origins of programmers: Identifying predictors of success for an objects first CS1.* Ph.D. Dissertation. https://search.proquest.com/dissertations-theses/on-origins-programmers-identifying-predictors/docview/305082053/se-2?accountid=13314 Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2020-11-19.

[36] Chris Wilcox and Albert Lionelle. 2018. Quantifying the Benefits of Prior Programming Experience in an Introductory Computer Science Course. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (Baltimore, Maryland, USA) *(SIGCSE '18).* Association for Computing Machinery, New York, NY, USA, 80–85. https://doi.org/10.1145/3159450.3159480

[37] Brenda Cantwell Wilson and Sharon Shrock. 2001. Contributing to Success in an Introductory Computer Science Course: A Study of Twelve Factors. In *Proceedings of the Thirty-Second SIGCSE Technical Symposium on Computer Science Education* (Charlotte, North Carolina, USA) *(SIGCSE '01).* Association for Computing Machinery, New York, NY, USA, 184–188. https://doi.org/10.1145/364447.364581

[38] Eric Yuan. [n.d.]. *Zoom.* https://zoom.us/

[39] Daniel Zingaro, Cynthia Bailey Lee, and Leo Porter. 2013. Peer Instruction in Computing: The Role of Reading Quizzes. In *Proceeding of the 44th ACM Technical Symposium on Computer Science Education* (Denver, Colorado, USA) *(SIGCSE '13).* Association for Computing Machinery, New York, NY, USA, 47–52. https://doi.org/10.1145/2445196.2445216