"Anything That Can Be Streamlined Would Be Great": Validating Elementary School Teachers' Preferences for a Block-Based Programming Teaching Augmentation System

Minji Kong
Computer and Information Sciences
University of Delaware
USA
mkong@udel.edu

Owen He
Computer and Information Sciences
University of Delaware
USA
heowen@udel.edu

Matthew Louis Mauriello
Computer and Information Sciences
University of Delaware
USA
mlm@udel.edu

Lori Pollock
Computer and Information Sciences
University of Delaware
USA
pollock@udel.edu

Abstract

Conversations around teaching augmentation (TA) system designs that benefit teachers' pedagogical capabilities while teaching using block-based programming environments (BBPEs) are up and coming. Despite the growing interest in this area, past findings and design recommendations still need to be validated in additional contexts beyond what is represented in the literature. In addition, there yet exist formal investigations that ground themselves in a theoretical model of student reflection and explore how a TA system design might support students' reflective learning processes as they program using BBPEs. In this paper, we aim to address such gaps with a concept validation study, where we conducted design activities with a targeted audience of seven teachers who have used Scratch in their teaching for grades three to five. Grounded in dialogues with our teacher participants during the activities, we reveal themes of interest identified via thematic analysis that strengthen past findings on teachers' preferences for a BBPE TA system and reveal additional factors for future researchers to consider as they explore design opportunities.

CCS Concepts

• **Social and professional topics → K-12 education**; • **Human-centered computing → Empirical studies in HCI**.

Keywords

teaching augmentation tools, block-based programming environments, teaching pedagogy, concept validation

Koli Calling '24, November 12–17, 2024, Koli, Finland

ACM ISBN 979-8-4007-1038-4/24/11
https://doi.org/10.1145/3699538.3699539

ACM Reference Format:
Minji Kong, Owen He, Matthew Louis Mauriello, and Lori Pollock. 2024. "Anything That Can Be Streamlined Would Be Great": Validating Elementary School Teachers' Preferences for a Block-Based Programming Teaching Augmentation System. In *24th Koli Calling International Conference on Computing Education Research (Koli Calling '24), November 12–17, 2024, Koli, Finland*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3699538.3699539

1 Introduction

With technology becoming more integral within our lives, K-12 educators are encouraged to integrate computing pedagogies into curricula and help students develop skills with technologies for "heightened disciplinary learning, critical thinking, and self-expression" [30]. Integration of computing pedagogies can benefit from including programming activities conducted in block-based programming environments (BBPEs) such as Scratch [30, 37], iSnap [35], and Blockly [17], which are designed to make programming and broadly the computing field more approachable [30]. Since their introduction, they have grown in popularity and are widely used, especially due to proven benefits of BBPEs for novice programmers in computing pedagogies [8].

Computing pedagogies can also benefit from reflective activities (e.g., students answer a prompt that aim for students to reflect on their learning experience as they craft a response) for an authentic learning experience and formative assessments of student progress [47, 49]. However, to the best of our knowledge, there have not been formal investigations with a theoretical perspective of what potential impacts student reflections have on BBPE-based, in-class learning experiences. This misses the opportunity for students to reflect on their *progress* towards successfully completing a Scratch activity [11]. Grounded in the model of reflection [29], we argue that teachers should also be guided to hold reflective moments throughout a Scratch activity, especially since reflection is crucial for "inference making and reasoning while searching for information to inform a solution" [11, 29].

Teachers can be provided guidance on student-centered teaching practices such as reflective activities through *teaching augmentation (TA) systems*. TA systems, which include learning analytics dashboards, wearables, and ambient lighting displays, are designed to augment and complement teachers' in-class pedagogical practices. While TA system designs have been introduced and explored by researchers across multiple disciplines, such conversations are new and up-and-coming in the context of BBPEs such as Scratch, where user-centered conversations with K-12 teachers have been held to identify their needs and preferences for a BBPE-based TA system (e.g., [21]). As teacher-centered conversations around BBPE-based TA system designs emerge, there also exists room for exploration of how such systems may support student reflections on their BBPE learning progress, especially given the aforementioned gap in theoretically grounded investigations of BBPE-based reflective experiences for students.

In this paper, we aim to validate past literature's reports of teachers' needs and preferences for a BBPE TA system to the context of our project's targeted audience of teachers, all the while taking a theoretical perspective to how augmentation could be presented to support student reflections. Toward this latter goal, our research questions are formative: *What preferences (i.e., form, interaction) do teachers have for BBPE TA systems that allows for reflective practice? And what challenges do they envision with the implementation of such systems?*

To address these research questions, we conducted a concept validation study with seven teachers who have experience teaching Scratch to grades 3 to 5; we focus on this grade band as it utilizes BBPEs to integrate computing to other subjects [30]. The study involved a *speed dating* design activity [13], during which teachers were presented storyboards of imagined BBPE-based classroom scenarios that involve the use of potential TA systems designed to support student reflections based on past literature about teachers' needs and preferences for one. The intended goal of the activity was to probe gut reactions from targeted users about the scenarios and designs, while also leaving the doors open to other unexpected design opportunities. In addition, teachers completed a stack ranking exercise of main points conveyed by each storyboard to further organize teachers' points of views.

Themes revealed from thematic analysis of conversations shared from the design activity revealed that while the participants agreed on some of the design recommendations made by past literature for a BBPE TA system, they shared additional sentiments that may be beneficial for researchers to consider in future works. For example, teachers described various systemic challenges they face during their teaching (e.g., lacking time and financial resources to be innovative with their teaching practices) that future design solutions should consider and aim to not perpetuate further. We elaborate on the study design, analysis methodology, and findings in the remaining sections of our paper.

2 Background and Related Work

2.1 Teaching Augmentation

Teaching augmentation (TA) systems are designed with the aim of benefiting teachers' in-class teaching and awareness and broadly benefiting student learning. Introduced and studied across various disciplines such as Human-Computer Interaction [4, 12], learning sciences [45], and learning analytics [27], TA designs take various forms. For example, there exist *learning analytics dashboards* that are designed to provide teachers information students' progress and performance in real time (e.g., [27, 42]). Some researchers have also experimented with *ubiquitous computing* in classrooms by introducing *distributed* TA systems [48]. Distributed digital lamps have been previously explored as a TA form, where ambient lamps were placed at student desks to depict work progress and help requests through color and pulse signals (e.g., [2, 4]). In addition, others have explored the use of *wearables* (e.g., [36]), such as smartwatches and earpieces, to deliver synchronous coaching and recommendations to novice teachers via a remote observer.

Although there have been additional forms of augmentation previously experimented with and reported on, we focus our concept validation on two forms in this paper: *learning analytics dashboards* and *ambient lighting displays*. Recent literature that has explicitly begun to explore the benefits of TA for teachers that teach programming using BBPEs such as Scratch has found that their teacher participants' most preferred form of augmentation was dashboards and also recommended that combinations of forms (e.g., dashboards with lighting displays) be explored (e.g., [21]). In our paper, we aim to validate such past finding with our specific audience of teachers teaching Scratch at grades 3 to 5.

We also acknowledge other prior works that are related to augmenting BBPE-based teaching, the discussion around which is still in its early stages. Specific to Scratch, research on how students' Scratch programming proficiency may be predicted predominantly explored code artifact analysis [1, 32, 38, 41], which lacks information on the *process* that led students to their final code projects. As a result, some researchers have introduced approaches towards collecting data on students' learning processes in a coding environment, including Scratch, via mediums such as screen recording [25], clickstream analytics [15], and automatic logging [22]. Revealed findings and recommendations around exploring design opportunities for a Scratch TA system have not yet been validated by additional studies. Furthermore, to our knowledge, studies related to a Scratch TA system have also not explored opportunities to potentially support designs of reflective activities to provide students more enriching Scratch learning experiences. We aim to contribute to addressing both gaps through our concept validation study.

2.2 Reflective Activities

Reflection, or the "process of purposeful contemplation or focused thinking" [11], is recognized by literature as a crucial component in higher-order thinking, such as effective problem solving and metacognitive monitoring [39]. Intentionally and consciously developing reflective skills not only impacts the growth of higher-order thinking skills, but it also plays a critical role in learning [3, 5, 14, 20, 24, 34, 39]. There exist works that have investigated how emerging technologies that initiate reflective opportunities impact skill and knowledge acquisition, with most technology designs posing reflective learning opportunities with various types of prompts via written statements or pedagogical agents [10, 18, 40]. Some works have also looked into the impacts of scaffolded reflective opportunities towards achieving learning goals during game-based learning (e.g., [28, 31, 46]). However, the grounding of reflection-involving, computing education (CEd) works in a theoretical perspective is still in its early stages. For instance, Cloude et al. investigated the quality and quantity of adolescents' reflections and their correlation to their learning and problem solving experiences with Crystal Island, a game-based learning environment, scaffolding reflective opportunities with the guidance of McAlpine et al.'s model of reflection [11, 29].

Composed of six components (i.e., goals, knowledge, action, monitoring, decision making, and a corridor of tolerance), the model of reflection states that a learner's reflection is driven by defined learning goals, which also influence inquiry, inference making, and reasoning skills [29]. Our work is also grounded in the model of reflection and argue that the model of reflection is also appropriate for studying reflection when a teacher is teaching using BBPEs while supported by a TA system. We believe TA systems, when meaningfully designed, could provide guidance to teachers with learning goals to initiate purposeful reflective opportunities for students as they work towards final code solutions.

3 Methodology

At a larger scale, our research towards designing a TA system that augments Scratch teaching practices follows Holstein et al.'s adoption of the LATUX workflow for designing and deploying learning analytics tools [19, 26]. The workflow involves the following major design phases: (1) initial needs and analysis and concept generation, (2) initial concept validation, (3) iterative lower-fidelity prototyping, (4) iterative higher-fidelity prototyping, and (5) iterative classroom piloting and evaluation. This paper reports on initial concept validation efforts to further understand and validate past findings on teachers' preferences for tools that augment their Scratch teaching abilities (e.g., [21]). Concept validation included a storyboarding session and ranking exercise with each participant to gather insights about various scenarios involving a TA system.

3.1 Concept Validation

Our concept validation study was conducted online by teleconference, with each session per participant lasting an average of about 76 minutes (SD = 11.63) and recorded upon the participant's consent.

3.1.1 Participants. Using purposive sampling strategy [33], we recruited seven teachers (six cis-women and one cis-man, all white) who taught grades 3-5 in the eastern United States at the time of our study and had prior experience teaching programming using Scratch. Requests for participants were sent via email using an IRB-approved advertisement; recipients were also encouraged to share this message with their colleagues. Each teacher participant received a $25 gift card per hour for their participation.

Ranging in years of overall teaching experience ($M = 21.29, SD = 7.65$), the seven teachers used Scratch in various subjects; two teachers taught Technology, two were Librarians/Media Specialists, two taught Gifted and Talented classes, and one taught Computer Science. For greater context on their teaching environments, four teachers were in schools where the largest student population was made up by Black or African American students. Five and six of the teachers taught students with disabilities or special needs and students whose native/dominant language was not English, respectively. All teachers are also current or former partners with our or neighboring universities' CEd professional development programs.

3.1.2 Speed Dating with Storyboards. The study involved a "speed dating" approach in accordance to the LATUX workflow adoption [19]. The speed dating design method is used to explore potential futures of a design concept with its users without the need for the design to be technically implemented [13, 50]. Specifically, we implemented storyboarding, one of the two main forms of speed dating, where participants (typically target users) are presented various storyboards depicting imagined scenarios that involve concepts of the researchers' proposed design based on identified user needs. Based on participants' reactions and comments after being probed with the prepared storyboards, researchers are able to validate their design concepts and/or identify unanticipated design opportunities before developing higher-fidelity prototypes.

One researcher met with the seven teacher participants individually for the concept validation study. During each study session, a participant was presented with 13 storyboards, each of which depicted a classroom scenario inspired by preferences teachers previously expressed for augmentation. The storyboards were used to probe participant reactions to design ideas and concepts derived from prior literature and discussions on such preferences.

There were three main types of scenarios that the storyboards were portraying. The first five storyboards were about how a teacher may monitor students' learning using Scratch with a TA system in the form of a learning analytics dashboard (2), ambient lighting displays (2), or a combination of both (1). In each of the five scenarios, the imagined TA system exhibited data of different granularity on student behaviors potentially exhibited during Scratch learning, such as their mouse-clicking frequency on their devices, as well as how frequently a student may specifically be interacting with block codes over other aspects of Scratch (e.g., cosmetic elements of a Scratch project such as costumes and backgrounds).

Three storyboards depicted instances where a teacher may initiate reflective moments using a TA system while students learn using Scratch. They vary in levels of balance between autonomy and automation (i.e., teacher having the freedom to craft their own written reflection prompts, teacher using one of the reflection prompts generated and recommended by a system, or a system generating a single option for a reflection prompt and requiring the teacher to utilize it). The remaining five storyboards represented different levels of granularity of data on how students reflected throughout an in-class activity and of the TA system's involvement in interpreting the data (e.g., a teacher should be responsible for reading and analyzing students' reflections word-to-word vs. the system summarizes findings from students' reflections and/or provides next-step teaching recommendations), in addition to how such data were used by a teacher (e.g., only the teacher should access the data vs. data can be shared with students as part of one's teaching). Figure 1 exhibits one of the storyboards shared with our participants. All 13 storyboards are provided in the supplementary materials.

To also translate the participants' qualitative expressed reactions to each storyboard's concept into quantitative data, participants completed three order stack ranking exercises, each based on the

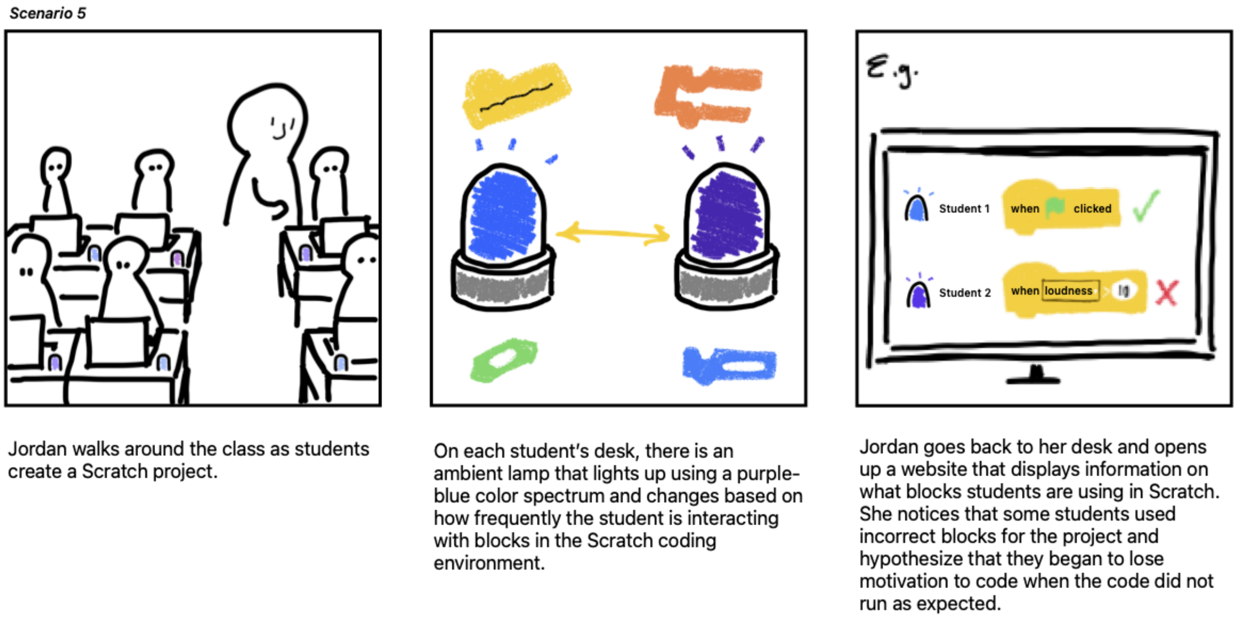

Figure 1: An example storyboard presented to teacher participants during the *speed dating* design activity.

three aforementioned types of storyboard scenarios. For each ranking exercise, on a digitally interactive whiteboard set up identically for each participant (i.e., Google Jamboard), a participant manually placed presented options, each option representing concepts that stemmed from the storyboards, in order from their favorite to least favorite concept.

3.2 Analysis

The analysis of data collected from our concept validation study was twofold. First, we used thematic analysis to analyze 9.3 hours total of audio recording transcripts of the teacher participants' comments in response to the speed dating storyboards. We randomly selected a transcript for two of the authors to code together for themes of interest using both inductive and deductive codes, with the code manual inspired by our study protocol and literature. Using Krippendorff's alpha (α) to measure inter-rater reliability, the two coders achieved an $\alpha = 0.82$ ($SE = 0.03$), which met the $\alpha \geq .800$ threshold of good reliability. After resolving disagreements and updating the code manual accordingly, the two coders divided the task of coding the remaining study transcripts.

Our analysis also involved quantifying our teacher participants' preferences for concepts that each storyboard presented based on their completion of the order rank exercises. Specifically, we scored each ranked concept using the Dowdall method to estimate which Scratch TA system concepts the participants favored the most. For each participant, the Dowdall method weighs the scores of n concept options being ranked with a vector of $(1, 1/2, 1/3, ..., 1/n)$, which is based on the participant's preference of options [16]. For example, if $n = 3$, a participant's first preference is given a score of 1, the second preference $1/2$, and the third preference $1/3$. We used the Dowdall method over the conventional Borda method because the Dowdall method has been shown to be more stable with small subsets of data that may contain anomalies (e.g., [23]).

4 Results

Here, we present the themes that were revealed in the teacher participants' responses during our storyboarding study about potential BBPE TA system designs that support reflective learning, in addition to how teachers ranked discussed concepts during the order ranking exercises.

4.1 Form of Augmentation

We presented to our participants five storyboards that would bring out their immediate reactions on imagined classroom scenarios that involve a dashboard interface (2), distributed displays (2), and a combination of both forms (1), each displaying different kinds of data on student behavior during a Scratch learning session.

Opinions amongst the seven participants on what augmentation form they preferred the most were mixed. Three teachers preferred dashboard interfaces explicitly while three other teachers were intrigued to explore the combination of both a dashboard and distributed displays; one participant remained unsure.

A synonymous concern among the three participants who preferred a dashboard interface over ambient lighting displays was the potential for distraction from the goals of the lesson. For example, Teacher 1 (T1) states, *"you may actually get kids focusing on that silly little light instead of their program [because] they're attempting to explore what kinds of actions would make the light change."* One teacher noted the judgemental nature of students towards their peers, posing a concern over the lighting displays as a prominently visible measurement of their progress. As T5 explains, *"they want to be fitting in with their peers. So I don't really think it would be fair*

Rank	Scenario	Score
1	*Scenario 5*: Teacher uses both a dashboard and lighting displays that display information on students' Scratch code block interaction activities	5.833
2	*Scenario 2*: Teacher uses a dashboard that displays information on what blocks students are interacting with in Scratch	4.167
3	*Scenario 4*: Teacher uses lighting displays that changes colors based on frequency of students' interactions with Scratch code blocks	2.583
4	*Scenario 1*: Teacher uses a dashboard that displays data on students' mouse-clicking frequency within the Scratch environment	1.800
5	*Scenario 3*: Teacher uses lighting displays that changes colors based on students' mouse-clicking frequency within the Scratch environment	1.600

Table 1: Monitoring students learning using Scratch scenarios and ranking.

for a student who doesn't understand a concept for other people to know that student doesn't understand the concept... I think it's just not appropriate for the kids to be able to judge each other."

The positives of the lighting display as a TA form were mentioned by teachers who preferred having access to the dashboard simultaneously (i.e., access to the dashboard on teacher's device with lighting displays throughout the classroom). T6 indicated that using both TA forms could potentially *"streamline [her] interactions with the students to be more efficient, and effective."* T4 elaborates on this further, *"I probably want both—it would give me more detail of why something's happening."* The ranking exercise further supports this preference (see Table 1), with four participants ranking the storyboard illustrating the combination as their most favorite out of the five storyboards. It also received the highest score of 5.83 via the Dowdall method, while a storyboard illustrating access to only the dashboard providing data on students' interactions with blocks placed second, with a score of 4.167.

Specific to the types of data each augmentation example provided, all seven teachers wanted a TA system to provide data of greater granularity about behaviors exhibited amongst their students when learning using Scratch (i.e., seeing how frequently each student is interacting with block codes in Scratch over observing mouse-clicking frequency on their device). Teachers noted the ability for a more catered data type such as block interactions to guide their lesson plans, with T5 explaining, *"So when we do assign something and we want to make sure you're using a 'Repeat' block and a 'Forever' block within your code and they're not using that, their code is so long and ineffective. So a quick glance with the [blocks'] color coding would be very helpful."* Four participants directly highlighted the lack of correlation between straightforward data such as mouse-clicking and students' comprehension levels of lesson concepts, with T2 stating, *"So I could see where a student who still doesn't know what they're doing is still clicking their mouse. Like, they might be clicking on whatever, they don't know what they're doing, but they're still clicking around and doing things."* T3 suggested that they would like to have the option of interchanging between the two data types, as *"seeing all of that at once on a screen might be a little much for a teacher"* and mouse clicks as a data type might be beneficial for a classroom with more easily distracted students.

4.2 Initiating Reflective Moments

Teachers were also shown three storyboards that illustrated how a teacher may initiate reflective moments using a TA system at varying degrees of autonomy. The boards reflected three levels of automation: generating a prompt freely, selecting one of the recommended prompts, and initiating reflection with a given prompt.

While all but one teacher (T7) did not like the system to have full automation over the process of composing and introducing written reflective prompts to students, there were mixed preferences amongst five participants on whether they explicitly preferred to either have the freedom to compose reflective prompts or to receive recommendations for prompts from the system. T5 did not have a preference for any of the reflection prompt generation methods.

Three teachers expressed a preference for receiving multiple suggestions over generating their own prompts. T3 highlighted the benefits in a standardized set of recommended reflection prompts, stating its ability to bridge gaps of understanding with the material on the teacher's end and allowing for *"the novice teacher and the experienced teacher both be on the same playing level."* T2 expressed interest in reducing the time spent on pushing a reflection prompt, noting that *"because we are very constrained on time... anything that can be streamlined [would be] great."* Two teachers seemed to prefer creating reflection prompts on their own, with T6 citing the differences between asking a question in class and having one student respond, or posting a reflection prompt and having the benefit of assessing the ability of all students. The storyboard rankings using the Dowdell method also were representative of the mixed results (see Table 2), with recommended reflection prompts scoring a 5.0 and free composition of reflection prompts scoring a 4.83.

While participants were asked about three forms of initiating reflective moments, many teachers suggested recommendations for additional reflection methods. T2, T4, and T5 all suggested an alternative form of response for the students. Specifically, T2 and T4 both highlighted the benefits of receiving recommended reflection questions with multiple-choice answers, while T5 proposed a scale-based response to a reflection prompt. These suggestions were rooted in the concern that many students may not have developed adequate writing ability to make good use of the time spent answering reflection prompts, with T4 stating *"they just don't have the patience to read quite a bit and really interpret it and give valued answers... this would take up a lot of time for them."* T1, T6, and T7

Rank	Scenario	Score
1	*Scenario 7*: Teacher selects one of the reflective prompts recommended by a TA system to ask students	5.000
2	*Scenario 6*: Teacher composes a reflective prompt to ask students via a TA system	4.833
3	*Scenario 8*: Teacher presses a button on a TA system to ping a recommended prompt, which is the default and only prompt available	3.000

Table 2: Initiating reflecting moments scenarios and ranking.

saw benefit in students having access to multiple different reflection prompts, allowing them to choose the one they most preferred to respond to.

4.3 Access and Presentation of Data on Student Reflections

The participants discussed preferences in how much a TA system should be involved in interpreting the collected student reflections. The TA system could present individual student responses, a summary of student responses, or a summary of responses with a recommendation on follow-up actions.

Two participants preferred to analyze individual student responses and make a decision more contingent on each student's needs. T1 and T6, both with smaller classroom sizes of about 10 to 15 students each, expressed concern over the summary's inability to discern the causes of the issues that students may have, with T6 stating *"if I couldn't see their independent—individual responses, then I wouldn't understand why it's coming up with this."* Four teachers opted for a summarized format of student responses, highlighting its time-saving quality as the primary advantage. T4 mentioned difficulties in managing multiple tasks due to time constraints, *"I will not have time to go sit down and analyze data during class time, there's just no way."* Of the four teachers who preferred the summary, three saw a benefit in having access to a recommendation on next steps as well. One teacher (T5) did not see benefit in utilizing class time for accessing or presenting data on student reflections.

Participants also expressed their opinions on whether they believe data on students' reflections should only be viewed by a teacher or shared with the class, either as a summary or anonymized individual responses. T1 and T7 emphasized the importance of solidarity between students and their peers, particularly when it pertains to individual issues they experience. T7 states the benefits of students viewing their peers' responses, *"seeing that not everybody found this easy, or that it was hard for some people... there's value in that."* The majority of teachers (5/7) preferred accessing data on their own, citing the distractions it may cause as well as negative associations that student's may feel with regards to their response being displayed in front of their peers. T5 elaborates on this further, *"I have a lot of students who have anxiety, that would throw them like, over the edge, they wouldn't want to respond... knowing that it would be on the board."* Of the teachers who preferred accessing student reflection data by themselves, all of them ranked going over the summary response above individual responses, given that both are presented to the whole class (Table 3).

5 Discussion

Through our concept validation study, participants provided their opinions on design recommendations that past works have made towards a potential Scratch TA system (e.g., [21]). The participants reached strong agreements about some of the potential design opportunities, such as: (a) teachers (six participants; one unsure) reacting favorably towards augmentation taking the form of a dashboard interface, either on its own (three) or combined with ambient lights (three); (b) all teachers preferring data of greater granularity on students' Scratch learning behaviors, and; (c) six teachers not wanting the TA system to have full automation over deciding when and how reflective opportunities should be introduced to students. Next, we discuss the dashboard and ambient light form factor in more detail, describe teachers' perspectives on student involvement in the TA system, and discuss implementation concerns, which are interesting areas for future work.

5.1 Dashboard vs. Lighting Displays

Our participants expressed varied opinions on other aspects of a potential Scratch TA system. For instance, there were participants who felt that TA systems that come in the form of ambient lighting displays may be distracting to students (3). This augmentation form also received similar mixed reactions in past Scratch augmentation studies (e.g., [21]). However, other works that explored ambient lighting displays as an augmentation form outside of the BBPE context made contradicting classroom observations, with teacher participants not considering the displays to be a distraction to teachers and students (e.g., [4, 44]). One of our own participants who felt that the lights would not be distracting addressed the "first-time enthusiasm" or "novelty effect" that a TA system may carry, where teachers and students may be extremely excited about a new technology when first introduced, which dwindles down over time as it blends into a classroom routine [6].

5.2 Student Involvement in Augmentation

With regards to allowing student involvement in viewing and assessing a class' set of reflections and utilizing it as a teaching opportunity, some teachers were concerned that their students would judge their peers' answers to reflection prompts and/or feel judged by their classmates, impacting their self-efficacy (i.e., one's belief in their own abilities to accomplish tasks [7]). To the best of our knowledge, there has not been any formal investigations conducted to explore the impacts of collaborative reflective activities on students' self-efficacy in programming and computing broadly. However, based on past studies that have revealed the potential benefits

Rank	Scenario	Score
1	*Scenario 11*: Teacher accesses and reads a summary of student's reflective responses and recommendation on next-steps	4.367
2	*Scenario 13*: Both teacher and students access and read a summary of students' reflective responses	3.583
3	*Scenario 9*: Teacher accesses and reads each student's reflective responses word-for-word	3.283
4	*Scenario 12*: Both teacher and students access and read each student's reflective responses word-for-word	2.500
5	*Scenario 10*: Teacher accesses and reads a summary of students' reflective responses	2.250

Table 3: Granularity of data and automation in TA system scenarios and ranking.

of certain designs of collaborative learning in computing for elementary students (e.g., [43]), we find that there is a chance for collaborative reflective activities to enhance one's BBPE learning experience, as opposed to our participants' expressed concerns on students potentially judging each other. We encourage future researchers to explore this further.

5.3 Challenges Faced by Teachers

We also observed that all teachers who have been teaching using Scratch in public schools (6) grounded their responses in the systemic struggles they face in a public elementary education system in the United States. For example, teachers mentioned that the potential cost of a TA system may serve as a barrier to them. The public school teacher participants were also especially hesitant to adopt new technology solutions such as a TA system as they lack the time to learn about a new technology and incorporate it into their teaching. Specifically, some were not classroom teachers (i.e., teachers who teach core subjects such as English and Mathematics) and did not allocate significant time to teach per student group (e.g., 30 to 45 minutes per week). T3, a librarian who holds a technology class section states, *"I have 45 minutes to do a lesson and a book exchange. With them getting in and out, it's really like 40 minutes that I have to activate... my problem always—I've been doing this 30 years—is figuring out how long it's going to take to do something."* T7 elaborates on this concern, *"Finding that balance of student [instruction]...opportunity for students to practice, develop something, and then the reflection piece, and then the assessment piece. I mean, just getting it all in time... there's never enough time."* In addition, hesitancy seemed to exist among teachers even with adopting Scratch in their lessons, with T5, a teacher who occasionally holds professional development (PD) for Scratch, expressing others' lack of motivation to learn, *"I know some of these teachers need PD. [My directors say] some of them don't even use [Scratch]. Even though it's in the curriculum, it's just too intimidating."*

We recommend future researchers to consider systemic challenges within the targeted teacher audience's teaching environments as potential TA system designs are explored to fuse into BBPE learning opportunities, and ensure that future designs do not perpetuate limitations teachers face in their profession.

5.4 Limitations

Our concept validation study involved a small sample of seven teacher participants, who were recruited via purposive sampling strategy [33]. As a result of focusing our recruitment on teachers who explicitly met our criteria—we were looking to hold conversations with teachers who had prior experience teaching programming using Scratch at grades 3-5 in the United States—our sample was U.S.-centric (specifically eastern U.S.) and not diverse along the lines of race or gender. However, our participants held various roles in public (6) and private (1) school settings—two teachers were librarians, two teachers taught Technology, two teachers taught Gifted and Talented program students (grouping of high-ability elementary and secondary schools in the U.S. to tailor traditional curriculum to students' skill levels [9]), and one was a Computer Science teacher—all at different schools and in urban (2) and suburban (5) neighborhoods. As with studies of qualitative nature such as ours, themes of interest that we obtained from our data are not representative of all Scratch- and BBPE-based TA system design preferences in other school setting and countries. In addition, the results of this study remain theoretical. We recommend researchers replicate our study and validate the results presented in this paper with specific emphasis on practical trials and iterative prototyping as well as further explorations of BBPE TA systems to diverse audiences and contexts.

6 Conclusion

To ensure that future design solutions are catered to the targeted user audiences to the best extent possible, it is critical to hold user-centered conversations on their design needs and preferences. Guided by up-and-coming conversations around the potentials of BBPE TA systems, we conducted a concept validation study to explore the soundness of design recommendations researchers have previously made with our targeted audience of teachers using BBPEs with grades 3 to 5. Our seven participants' preferences aligned with some of the recommendations—for instance, the teachers exhibited interest in TA taking the form of a dashboard interface that provided highly detailed data and did not want the system to have full automation over tasks such as data interpretation—while providing varied opinions and recommendations of their own. We encourage future researchers to consider the dialogues with the teacher participants presented in this paper, especially with regards to their design preferences that were grounded in systemic challenges that exist in their teaching environment, as conversations around potential BBPE TA systems continue to evolve.

Acknowledgments

This material is based on work by Minji Kong supported by the National Science Foundation Graduate Research Fellowship (NSF GRFP) under Grant No. 1940700, and any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. Owen He's contributions were supported by the University of Delaware's Undergraduate Research Program.

References

[1] Efthimia Aivaloglou and Felienne Hermans. 2016. How Kids Code and How We Know: An Exploratory Study on the Scratch Repository. In *Proc. of the 2016 ACM Conf. on Int. Comput. Educ. Res.* ACM.

[2] Hamed Alavi and Pierre Dillenbourg. 2012. An Ambient Awareness Tool for Supporting Supervised Collaborative Problem Solving. *IEEE Trans. on Learn. Techn.* 5, 3 (2012).

[3] Mohammed Alzaid and I-Han Hsiao. 2018. Effectiveness of Reflection on Programming Problem Solving Self-Assessments. In *2018 IEEE Frontiers in Education Conference (FIE)*. IEEE Computer Society, 1–5.

[4] Pengcheng An, Saskia Bakker, Sara Ordanovski, Ruurd Taconis, and Berry Eggen. 2018. ClassBeacons: Designing Distributed Visualization of Teachers' Physical Proximity in the Classroom. In *Proc. of the 12th Int. Conf. on Tangible, Embed., and Embodied Interact.* ACM.

[5] Roger Azevedo, Michelle Taub, and Nicholas V Mudrick. 2017. Understanding and Reasoning about Real-Time Cognitive, Affective, and Metacognitive Processes to Foster Self-Regulation with Advanced Learning Technologies. In *Handbook of Self-Regulation of Learn. and Perform.* Routledge, 254–270.

[6] Saskia Bakker, Elise van den Hoven, Berry Eggen, and Kees Overbeeke. 2012. Exploring Peripheral Interaction Design for Primary School Teachers. In *Proc. of the 6th Int. Conf. on Tangible, Embed., and Embodied Interact.* ACM.

[7] Albert Bandura, William H Freeman, and Richard Lightsey. 1999. Self-efficacy: The exercise of control.

[8] David Bau, Jeff Gray, Caitlin Kelleher, Josh Sheldon, and Franklyn Turbak. 2017. Learnable Programming: Blocks and Beyond. *Commun. ACM* 60, 6 (2017).

[9] Rachana Bhatt. 2011. A review of gifted and talented education in the United States. *Education Finance and Policy* 6, 4 (2011), 557–582.

[10] Dan Carpenter, Michael Geden, Jonathan Rowe, Roger Azevedo, and James Lester. 2020. Automated Analysis of Middle School Students' Written Reflections During Game-Based Learning. In *Int. Conf. on Artificial Intelligence in Educ.* 67–78.

[11] Elizabeth Cloude, Dan Carpenter, Daryn A Dever, James Lester, and Roger Azevedo. 2021. Game-Based Learning Analytics for Supporting Adolescents' Reflection. *J. Learn. Analytics* 8, 2 (2021), 51–72.

[12] Bernice d'Anjou, Saskia Bakker, Pengcheng An, and Tilde Bekker. 2019. How Peripheral Data Visualisation Systems Support Secondary School Teachers during VLE-Supported Lessons. In *Proc. of the 2019 on Des. Interact. Syst. Conf.*

[13] Scott Davidoff, Min Kyung Lee, Anind K Dey, and John Zimmerman. 2007. Rapidly Exploring Application Design Through Speed Dating. *UbiComp 2007: Ubiquitous Computing* (2007), 429.

[14] John Dewey. 1923. *Democracy and Education: An Introduction to the Philosophy of Education.* Macmillan.

[15] Daniel Amo Filvà, Marc Alier Forment, Francisco José García-Peñalvo, David Fonseca Escudero, and María José Casañ. 2019. Clickstream for Learning Analytics to Assess Students' Behavior with Scratch. *Future Gener. Comput. Syst.* 93 (2019).

[16] Jon Fraenkel and Bernard Grofman. 2014. The Borda Count and its real-world alternatives: Comparing scoring rules in Nauru and Slovenia. *Australian Journal of Political Science* 49, 2 (2014), 186–205.

[17] Neil Fraser. 2015. Ten Things We've Learned from Blockly. In *2015 IEEE Blocks and Beyond Workshop (Blocks and Beyond).*

[18] Michael Geden, Andrew Emerson, Dan Carpenter, Jonathan Rowe, Roger Azevedo, and James Lester. 2021. Predictive Student Modeling in Game-Based Learning Environments with Word Embedding Representations of Reflection. *Int. J. Artificial Intelligence in Educ.* 31, 1 (2021), 1–23.

[19] Kenneth Holstein, Bruce McLaren, and Vincent Aleven. 2019. Co-Designing a Real-Time Classroom Orchestration Tool to Support Teacher–AI Complementarity. *J. Learn. Anal.* 6, 2 (2019).

[20] Cruz Izu and Brad Alexander. 2018. Using unstructured practice plus reflection to develop programming/problem-solving fluency. In *Proc. of the 20th Australasian Comput. Educ. Conf.* 25–34.

[21] Minji Kong, Matthew Louis Mauriello, and Lori Pollock. 2022. Exploring K-8 Teachers' Preferences in a Teaching Augmentation System for Block-Based Programming Environments. In *Proc. of the 22nd Koli Calling Int. Conf. on Comput. Educ. Res.* ACM.

[22] Minji Kong and Lori Pollock. 2020. Semi-Automatically Mining Students' Common Scratch Programming Behaviors. In *Koli Calling '20: Proc. of the 20th Koli Calling Int. Conf. on Comput. Educ. Res.* ACM.

[23] Victor Le Pochat, Tom Van Goethem, and Wouter Joosen. 2019. Evaluating the long-term effects of parameters on the characteristics of the tranco top sites ranking. In *12th USENIX Workshop on Cyber Security Experimentation and Test.*

[24] Tian Luo and John Baaki. 2018. Scaffolding Problem-Solving and Instructional Design Processes: Engaging Students in Reflection-in. *Student-Centered Virtual Learn. Environments in Higher Educ.* (2018), 40.

[25] Sze Yee Lye and Joyce Hwee Ling Koh. 2018. Case Studies of Elementary Children's Engagement in Computational Thinking Through Scratch Programming. In *Comput. Think. in the STEM Discipl.* Springer.

[26] Roberto Martinez-Maldonado, Abelardo Pardo, Negin Mirriahi, Kalina Yacef, Judy Kay, and Andrew Clayphan. 2015. LATUX: An Iterative Workflow for Designing, Validating, and Deploying Learning Analytics Visualizations. *J. Learn. Analytics* 2, 3 (2015), 9–39.

[27] Manolis Mavrikis, Sergio Gutierrez-Santos, and Alex Poulovassilis. 2016. Design and Evaluation of Teacher Assistance Tools for Exploratory Learning Environments. In *Proc. of the 6th Int. Conf. on Learn. Anal. & Knowl.* ACM.

[28] Richard E Mayer. 2014. *Computer Games for Learning: An Evidence-Based Approach.* MIT Press.

[29] L McAlpine, C Weston, C Beauchamp, C Wiseman, and J Beauchamp. 1999. Building a metacognitive model of reflection. *Higher Educ.* 37, 2 (1999), 105–131.

[30] Kelly Mills, Merijke Coenraad, Pati Ruiz, Quinn Burke, and Josh Weisgrau. 2021. *Computational Thinking for an Inclusive World: A Resource for Educators to Learn and Lead.* Technical Report. Digital Promise.

[31] Roxana Moreno and Richard E Mayer. 2005. Role of Guidance, Reflection, and Interactivity in an Agent-Based Multimedia Game. *J. Educ. Psychology* 97, 1 (2005).

[32] Jesús Moreno-León, Gregorio Robles, and Marcos Román-González. 2015. Dr. Scratch: Automatic Analysis of Scratch Projects to Assess and Foster Computational Thinking. *RED. Revista de Educación a Distancia* (2015).

[33] LA Palinkas, SM Horwitz, CA Green, JP Wisdom, N Duan, and K Hoagwood. 2015. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health* 42, 5 (2015), 533–544.

[34] Niraj Patel, S Glenn Baker, and Laura D Scherer. 2019. Evaluating the cognitive reflection test as a measure of intuition/reflection, numeracy, and insight problem solving, and the implications for understanding real-world judgments and beliefs. *J. Experimental Psych.: General* 148, 12 (2019), 2129.

[35] Thomas Price, Yihuan Dong, and Dragan Lipovac. 2017. iSnap: Towards Intelligent Tutoring in Novice Programming Environments. In *Proc. of the 2017 ACM SIGCSE Tech. Symp. on Comput. Sci. Educ.* ACM.

[36] Rebecca Quintana, Chris Quintana, Cheryl Madeira, and James Slotta. 2016. Keeping Watch: Exploring Wearable Technology Designs for K-12 Teachers. In *Proc. of the 2016 CHI Conf. Extended Abstr. on Hum. Factors in Comput. Syst.* ACM.

[37] Mitchel Resnick, John Maloney, Andrés Monroy-Hernández, Natalie Rusk, Evelyn Eastmond, Karen Brennan, Amon Millner, Eric Rosenbaum, Jay Silver, Brian Silverman, and Yasmin Kafai. 2009. Scratch: Programming for All. *Commun. ACM* 52, 11 (2009).

[38] Gregorio Robles, Jean Carlo Rossa Hauck, Jesús Moreno-León, Marcos Román-González, and Roberto Nombela. 2018. On Tools that Support the Development of Computational Thinking Skills: Some Thoughts and Future Vision. In *Proc. of the Int. Conf. on Comput. Think. Educ.* The Education University of Hong Kong.

[39] Pina Tarricone. 2011. *The Taxonomy of Metacognition.* Psych. Press.

[40] Michelle Taub, Roger Azevedo, Amanda E Bradbury, and Nicholas V Mudrick. 2020. Self-Regulation and Reflection during Game-Based Learning. *Handbook of Game-Based Learn.* 239 (2020).

[41] Peeratham Techapalokul. 2017. Sniffing Through Millions of Blocks for Bad Smells. In *Proc. of the 2017 ACM SIGCSE Tech. Symp. on Comput. Sci. Educ.* ACM.

[42] Mike Tissenbaum and James Slotta. 2014. Developing an Orchestrational Framework for Collective Inquiry in Smart Classrooms: SAIL Smart Space (S3). *Int. Conf. the Learn. Sci.* (2014).

[43] Jennifer Tsan, Jessica Vandenberg, Zarifa Zakaria, Joseph B. Wiggins, Alexander R. Webber, Amanda Bradbury, Collin Lynch, Eric Wiebe, and Kristy Elizabeth Boyer. 2020. A Comparison of Two Pair Programming Configurations for Upper Elementary Students. In *Proc. of the 51st ACM Tech. Symp. on Comput. Sci. Educ. (SIGCSE '20).* Association for Computing Machinery.

[44] Erik van Alphen and Saskia Bakker. 2016. Lernanto: Using an Ambient Display During Differentiated Instruction. In *Proc. of the 2016 CHI Conf. Extended Abstr. on Hum. Factors in Comput. Syst.* ACM.

[45] Anouschka van Leeuwen, Nikol Rummel, Kenneth Holstein, Bruce McLaren, Vincent Aleven, Inge Molenaar, Carolien Knoop-van Campen, Baruch Schwarz, Naomi Prusak, Osama Swidan, et al. 2018. Orchestration Tools for Teachers in the Context of Individual and Collaborative Learning: What Information Do Teachers Need and What Do They Do With It? In *Int. Conf. the Learn. Sci.* Int. Soc. the Learn. Sci.

[46] J Vrugte, T Jong, P Wouters, S Vandercruysse, J Elen, and H Oostendorp. 2015. When a game supports prevocational math education but integrated reflection does not. *J. Comput. Assisted Learn.* 31, 5 (2015), 462–480.

[47] David Weintrop, Daisy Rutstein, Marie Bienkowski, and Steven McGee. 2021. Assessment of Computational Thinking. In *Comput. Think. in Educ.* Routledge.
[48] Mark Weiser. 1993. Hot Topics - Ubiquitous Computing. *Computer* 26, 10 (1993).
[49] Ho YuekMing and Latifah Abd Manaf. 2014. Assessing Learning Outcomes through Students' Reflective Thinking. *Procedia - Social and Behavioral Sciences* 152 (2014), 973–977.
[50] John Zimmerman and Jodi Forlizzi. 2017. Speed Dating: Providing a Menu of Possible Futures. *She Ji: The J. Design, Economics, and Innovation* 3, 1 (2017).

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009