Education Tech Research Dev (2022) 70:1867–1891
https://doi.org/10.1007/s11423-022-10145-5

DEVELOPMENT ARTICLE

Connecting learning and playing: the effects of in-game cognitive supports on the development and transfer of computational thinking skills

Zhichun Liu[1] · **Allan C. Jeong**[1]

Accepted: 6 July 2022 / Published online: 1 August 2022

Abstract
Prior studies on game-based learning provide limited and mixed results in the transfer of skills learned during game play to contexts outside of the game. This study tested the effects of playing a blocked-based programming educational game implemented with in-game cognitive supports on students' ability to learn and apply computational thinking (CT) skills in near and far transfer tasks. With 79 students randomly assigned to one of two conditions, the control group received basic game supports and the treatment group received cognitive supports in addition to the basic game supports. After two hours of total gameplay over the course of four days, both groups performed equally well, and students' CT skills were improved significantly at the near transfer level but not at the far transfer level. Students in the control condition performed significantly better on far transfer compared to the students in the treatment condition. Regression analyses indicated that the overall use of the cognitive supports was infrequent, but the amount of time spent voluntarily using cognitive supports with help on goal setting and worked examples predicted far transfer performance. How students use the cognitive supports (subverting the use of cognitive support to conscientiously learn the computational skill by using them more as game cheat sheets) might explain these findings. Design implications and directions for future research on facilitating learning transfer with in-game supports are discussed.

Keywords Computational thinking · Game-based learning · Learning transfer · Cognitive supports · K-12

Game-based learning (GBL) is often designed as an engaging, interactive, and experiential learning experience within which learners solve various problems (Ke, 2016; Kiili, 2005). Many studies have provided empirical evidence to show that using educational games can support knowledge acquisition, skill development, motivation, attitude change, metacognition, and epistemological development (Clark et al., 2016; Wouters & van Oostendorp, 2013). Although the research reveals mean effect sizes that show game-based

✉ Zhichun Liu
liulukas91@gmail.com

[1] Department of Educational Psychology and Learning Systems, College of Education, Florida State University, 1114 W. Call Street, Tallahassee, FL 32306, USA

learning (GBL) can be effective (e.g., Clark et al., 2016), the findings in game-based learning research have been mixed (Prensky, 2012), and this can be partly attributed to the high variance in game designs (e.g., with versus without explicit instruction embedded into the games) and variance in how and which learning outcomes are measured (Clark et al., 2016).

Specifically, prior studies show mixed results on students' ability to transfer in-game performance to contexts outside of the game. To achieve far transfer, students must be able to apply newly acquired knowledge and skills in a context that is *different* from the learning context (Royer, 1979). Although studies show that students can acquire conceptual knowledge through gameplay without explicit or direct instruction (Kim, 2014; Shute et al., 2013), studies also show that students who excel in in-game performance do not learn the instructional content and do not perform well on tests and transfer tasks (Arena & Schwartz, 2014; Barzilai & Blau, 2014; Clark et al., 2011; Masson et al., 2011, Wang, 2017). For example, Masson et al. (2011) found that playing a game that involving realistic projectile motion helped students become proficient in applying Newtonian laws of physics to predict and manipulation of the trajectories of falling objects, and yet, these students performed poorly when given test questions that explicitly tested their understanding of the Newtonian laws. Similarly, Barzilai and Blau (2014) found that students enjoyed and successfully applied math skills to calculate costs and profits while playing a business-simulation game but were not able to perform the same calculations when given real-world scenarios on posttests. Games have been effective in helping students gain "naïve intuitions (Masson et al., 2011)", while transfer tasks often require formal understanding and articulated acquired knowledge (Chase et al., 2019).

Learning transfer and problem-solving in GBL

Game-based learning experiences are often designed to be highly context specific and integral to the game experience (c.f. Kiili, 2005). As a result, students may "succeed" in the game context, but not necessarily attend to and commit to acquiring and learning the target knowledge and skills that are used to successfully complete the game. For example, Shute et al. (2013) conducted a study with the game *Physics Playground* where students drew on levers and used properties of torque to successfully direct the trajectory of a ball through a combination of obstacles. The students were able to move the ball through the obstacle courses successfully, but they achieved only a shallow level of understanding of the properties of torque based on the marginal gains observed in their performance on a transfer task performed on paper-and-pencil tests. Similarly, Ke et al. (2019) developed an educational game where play and learning were integrated in various architectural tasks that prioritize cognitive and content engagement. Their early user-testing revealed that students primarily used trial-and-error to test and find which random combinations worked successfully as opposed to investing mental effort into learning and strategically applying the target math skills to find the correct solutions when the sufficient cognitive supports is absence. Previous self-regulated learning literature reports similar findings when students are found to use shallow and ineffective learning strategies to solve problems on their own (Azevedo & Aleven, 2013; Lajoie, 2008). One explanation as why students engage in shallow learning is that GBL requires high levels of mental effort (e.g., problem presentation, solution construction and implementation, result evaluation, and reflection; Shute et al., 2016). Cognitive load theory suggests that GBL often imposes high intrinsic and extraneous cognitive load on students (Low et al., 2010; Morrison et al., 2015). This inhibits students' ability to manage cognitive resources to effectively engage in strategic and deliberate problem-solving that give students cause to attend more closely to and more effectively distill the underlying knowledge and skills to be used in the game (Kirschner et al., 2006; Paas, 1992).

To address this issue, Hammer et al. (2004) proposed a cognitive structure of learning transfer from a resource-based view. In their theory, learning transfer is a process that relies on the coordination and stabilization of resource activation (i.e., numerous metacognitive and epistemological resources, including ones for understanding the source of knowledge, forms of knowledge, knowledge-related activities, and stances toward knowledge). According to this theory, students are not able to achieve learning transfer when students are not able to stabilize the metacognitive and epistemological resources. From the cognitive load theory perspective, Pass (1992) argued, given the limited cognitive capacity of learners, the more mental effort is needed, the less likely they are able to abstract appropriate schemata. This in turn results in the failure to achieve learning transfer. Therefore, it is important to provide learners sufficient cognitive supports in their gameplay to reduce the cognitive challenge, and thus promote learning transfer.

Supporting problem-solving and learning transfer in GBL

One approach to reduce cognitive load to increase the likelihood of achieving learning transfer is to incorporate specific types of in-game instructional supports. However, despite of the promises of supports in GBL (Shute et al., 2019), students sometimes can perceive supports as obtrusive, distracting, while at the same time reducing their sense of learner autonomy and the enjoyment in gameplay (Leemkuil & de Jong, 2012). Researchers have debated the efficacy of implementing such in-game supports, much like how researchers have debated over the efficacy of using guided discovery learning versus pure discovery learning (Kirschner et al., 2006). While Black and Deci (2000) argue for greater emphasis on increasing learner autonomy to maintain higher levels of engagement during the learning process, a review of prior studies (Mayer, 2004) reveal that pure discovery learning (where learner autonomy is maximized) was not as effective as guided discovery learning when direct guidance was systematically presented around learning objectives.

The advocacy of fun in gameplay and learners' autonomy should not suggest the absence of well-designed instructional supports. The major criticism of external supports is that it is extraneous to the learning activities and reduces the autonomy and the enjoyment (e.g., Clark et al., 2011). Therefore, it highlights the need for well-integrated supports for GBL experience. Learning supports should be part of the game design and experience that guide learners' in-game problem solving (Abdul Jabbar & Felicia, 2015; Ke, 2016). Although we know learners may need cognitive supports in their game-based learning experience, achieving the balance between supporting the gameplay without hurting the fun in gameplay is still an ongoing task for GBL researchers (Bainbridge et al., 2022; Shute et al., 2021).

As discussed, the design of the supports should (a) be integrated to the in-game problem solving experience, and (b) address learners' limited cognitive resources. Wouters and van Oostendorp's (2013) meta-analysis of game-based research reviewed 24 in-game instructional supports in the literature. The meta-analysis revealed a total of 11 types of in-game supports used and tested to facilitate the cognitive process of selection (focusing attention on the relevant concepts and skills), 9 types of supports to facilitate knowledge organization and integration, and 4 types of supports to facilitate other or non-specifiable cognitive processes. Overall, they found that instructional supports can improve learning of concepts ($d=0.34$) and skills ($d=0.62$), particularly when supports are used to facilitate the cognitive process of selection ($d=0.42$). The use of modeling and presenting advice are two types of instructional supports used to facilitate the cognitive process of *selection* to reducing cognitive load per Wouters and van Oostendorp's synthesis of game-based learning literature.

Modeling is a commonly used cognitive support where students are provided explication or indication of how a problem is solved and then students follow the strategies on the same or a similar task (Jonassen, 1999). Modeling can be in many different forms, such as simple imitation, demonstration, and fully or partially worked examples. Modeling solutions with worked examples have been found to improve students' problem-solving in GBL, the quality in students' knowledge maps, and scores on problem-solving retention questions (Shen & O'Neil, 2006). Most of all, the use of worked examples have been found to reduce cognitive load and help students construct schemas that are generalized and applied to solving new problems, to improve learning transfer (Gick & Holyoak, 1983; Renkl et al., 2004). Overall, Wouters and van Oostendorp's (2013) meta-analysis found that the average effect size was 0.46 when modeling were used in GBL. When presenting students with advice, prompts are presented to direct attention to the relevant information. Leemkuil (2006) compared students' performance of transfer tasks between two groups (with vs. without advice) after playing a business management game that develops students' knowledge management skills. The results showed that the students receiving advice performed better on transfer tasks. Wouters and van Oostendorp's (2013) meta-analysis also reported an effect size of 0.13 when advices were used in GBL.

At the same time, prior studies that examine the use of modeling and advice have focused almost exclusively on their effects on in-game performance and performance on near transfer tasks (not far transfer tasks). Among the two studies that have examined it effects of far transfer, Mayer et al's (2002) study examined the effects of *modeling* on far transfer to find that students that received modeling at fixed times in the game improved scores on in-game performance ($d=0.56$) and on far transfer ($d=0.99$). Leemkuil's (2006) dissertation study tested the effects of *advice* on in-game performance and far transfer and found that students that received adaptive advice had similar game performance and slightly higher far transfer performance ($d=0.14$).

Fostering computational thinking skills through GBL

Due to the growth and prevalence of artificial intelligence and computing technologies, CT skills have become an essential aspect of digital literacy (Kafai & Proctor, 2021), and both educators and researchers have shown increasing interest in introducing CT and Computer Science into K-12 education (Zhang & Nouri, 2019). Various proposed conceptual frameworks of CT (Atmatzidou & Demetriadis, 2016; Brennan & Resnick, 2012; Shute et al., 2017) are composed of multiple cognitive and behavioral operations. These CT skills can be used by students to solve various computational problems across different contexts and can be viewed as a form of problem-solving (Grover & Pea, 2013).

Therefore, games can be a salient vehicle for students to develop CT skills because problem-solving (e.g., instant feedback, opportunity to test solutions multiple times, opportunity to make multiple iterative revisions, etc.) is an inherent element in games (Dondlinger, 2007). Various empirical evidence has shown that GBL can help foster CT

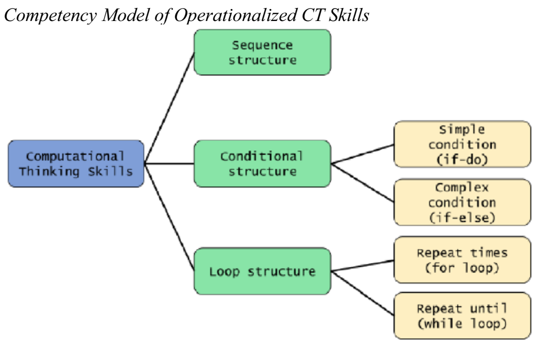


Fig. 1 Competency model of operationalized CT skills

(Hooshyar et al., 2020; Hsu et al., 2018; Israel-Fishelson & Hershkovitz, 2020). Particularly, GBL provides an engaging and accessible experience for students and learn CT (see Israel-Fishelson & Hershkovitz, 2020), which addresses the challenges in negative attitudes towards learning CT among younger students and underrepresented population (Leonard et al., 2016; Sun et al., 2021).

In this study, CT skills were operationalized, observed, and measured in terms of how students use target computational concepts (e.g., basic sequence, conditional structure, and loop structure). The competency model is shown below in Fig. 1. In the prior study (Zhao & Shute, 2019) that used GBL to teach students these CT skills, constraints were placed on solutions to increase the level of challenge and to require students to produce more abstract and efficient solutions. Although the results provided some indication that GBL can improve students' CT skills, applying constraints on the gameplay was overall not an effective strategy. As a result, this study presents a version of game without solution constraints and with cognitive supports designed to improve CT learning.

Given that some prior GBL studies have achieved learning transfer but often and/or potentially at the expense of reducing learner engagement, it remains inconclusive as to how best to design well-integrated cognitive supports that produce both engaging GBL experiences and promote far transfer in CT learning. As a result, we focus on examining the effect of in-game supports on the CT skills and the students' ability to transfer CT skills used during in-game play to a different problem-solving context outside of gameplay. Specifically, we aim to determine the added-value (Mayer, 2019) and impact of supports across multiple GBL outcomes: (a) in-game performance, (b) near transfer (i.e., game environment like), (c) far transfer (i.e., knowledge/skill assessment like), and (d) enjoyment to gauge how much learners are distilling the target knowledge from the learning/game experience and how much fun in the gameplay is preserved.

Purpose of the study

The CT educational game, *Penguin Go* (see details in the Method section), was used in this study to test the effects of cognitive supports on students' ability to transfer CT skills learned during in-game play to computational problems beyond the context of the game. *Penguin Go* is designed to leverage the popularity of video games to introduce CT skills to middle and elementary school students. To shed light on how to increase student interest and improve their CT skills to prepare them for future work and studies in STEM-related fields, the purpose of this study was to address the following research questions:

1. Does playing *Penguin Go* with additional cognitive supports in the form of modeling and advice lead to better in-game performance and attitudes compared to a control group without the cognitive supports?
2. Does playing *Penguin Go* with additional cognitive supports improve CT skills on near and far transfer than a control group without the cognitive supports?
3. What is the relationship between the level of students' use of supports and learning transfer?

Based on the reviewed literature, this study hypothesized that students achieve higher in-game performance, near, and far transfer tasks in GBL while experiencing similar levels of enjoyment and perceived learning when they are provided with modeling and advice than those who don't. This study also hypothesized that those who engage more with the supports outperform those who engage less with the supports.

Method

Participants

The sample consisted of 79 4th–6th students in two large charter schools in the southeastern United States (63 from School A, 16 from School B) with 21 4th graders, 17 5th graders, and 31 6th graders. Twenty-seven students self-identified as male and 43 self-identified as female. The sample included a wide range of ethnicities: Asian ($n=3$), Black or African American ($n=22$), American Indian or Alaska Native ($n=2$), Native Hawaiian or Pacific Islander ($n=1$), Hispanic ($n=7$), White ($n=40$), and Other ($n=5$). Students participated in the study during their math or science classes. All participants were randomly assigned to one of the two conditions—40 to the cognitive support (CS) and 39 to the no support (NS) group. Penguin Go was designed as an introductory CT-focused GBL experience with blocked-based programming for K-9 students who have no or limited coding experience. A large percentage of the students (91.1%) had prior experience with coding, with 13.9% of participants reporting that they had no prior knowledge of coding, while 84.8% of the students reporting that they had limited experience (under 20 h) with at least one of two types of block-based programming (e.g., *code.org* or *Scratch)*. Most of the students (74.7%) play games regularly, from a few times every month to more than 1 h every day.

Procedure

The experiment spanned four 50-min class sessions per classroom in total. After participants were randomly assigned to a condition, each participant completed a 15-min pretest, followed by 35 min of gameplay in Session 1. In Sessions 2—3, the participants played the game individually. In Session 4, the participants played the game for 10–15 min, completed a near transfer test followed with a far transfer test in roughly 30 min, then answered a few demographics and questions on perceived learning and game enjoyment in a paper–pencil survey. The total time spent on gameplay was approximately 135 min—similar to the treatment times used in prior studies that report the positive effects of GBL (e.g., Zhao & Shute, 2019).

A Sample Level of Penguin Go with Completed Coding Blocks

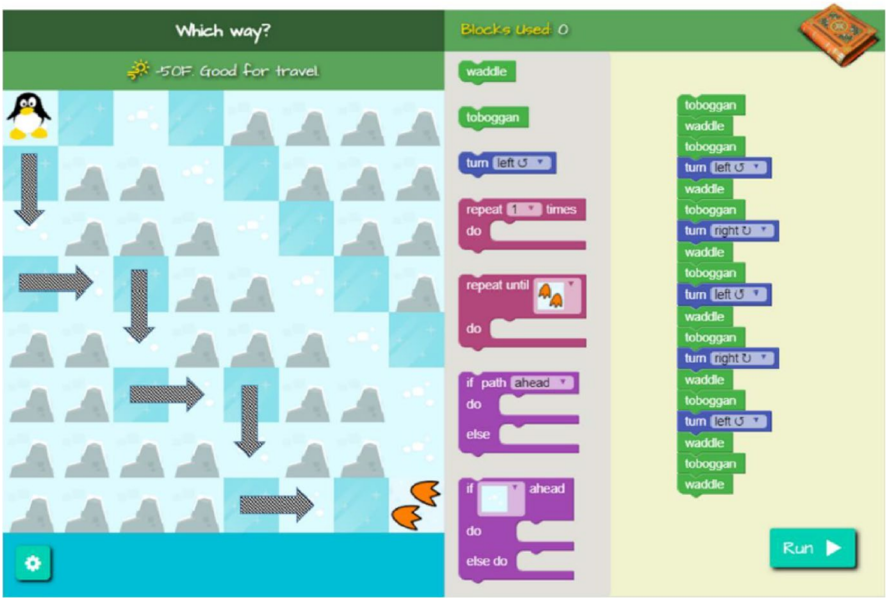


Fig. 2 A sample level of penguin go with completed coding blocks. The goal of each level is to move the penguin to its footprint. The arrows highlight a possible pathway, but they are not available in the real game. The penguin will move through different terrain. For example, the penguin can only waddle on snow and only toboggan on ice. The middle panel shows all the available blocks in one level. The almanac on the top right introduces how to use each block. The right panel is the place where players construct solutions

Intervention

Penguin Go

Using block-based programming language, students played *Penguin Go* (Fig. 2) by dragging and snapping together given blocks of executable codes to direct an on-screen penguin to a given destination without having to type and write lines of code. This design enables students to focus on computational thinking (CT) instead of syntax. In the game, students advance from one level to the next level by creating more complex blocks of code to successfully navigate their penguin on and through terrain obstacles (e.g., snow, ice, rocks) to get to the yellow footprints using different blocks. Students develop CT skills through carefully decomposing the problem scenarios, actively using the available resources, and strategically planning solutions in the in-game problem solving.

General game supports presented in CS and NS conditions

The students in both conditions were able to access detailed explanations and examples of how to play the game through a resource page called Digital Almanac. It provides guidance on how to create if–then, repeat-times, and if-else blocks. The content in the resource page covered all the necessary knowledge components of CT skills (Fig. 3A) to play the game and this content could be accessed at any time during the game.

Cognitive supports accessed in the CS condition

The cognitive supports in *Penguin Go* (Table 1) were delivered as a source of advice (concept-specific prompts that show one particular way to move the penguin to its target location) and a source of modeling (level-specific tips that direct attention to the relevant information or the goal and provide a partially worked-out example) to facilitate students' in-game problem-solving. Students received concept-specific prompts only at the start of the first level for each of the computational concepts. For example, upon starting the first level of repeat-times (Fig. 3B), students received the advice by viewing an introductory prompt describing what the repeat-times block can do (Fig. 3C) to solve the problem (moving the penguin to the target location). Students can close the prompt voluntarily. Students can choose to view the archived information in the Digital Almanac after exiting the prompt. The advice presented at each level was designed to provide scaffolding to help students find better ways to help their penguin reach its destination without revealing the answer directly.

To receive modeling during the problem-solving, students can click the help button at any time to view level-specific tips (Fig. 3D) to direct their attention to the relevant information or goal (e.g., to use as few blocks as possible to achieve the desired outcome using the prompt "*Do you know that you can solve this level with just 5 blocks?*"). Each modeling prompt revealed part of the answer by illustrating, for example, block combinations inside the loop block (i.e., waddle; turn left; waddle; and then turn right).

Instruments

Computational thinking (CT) pretest

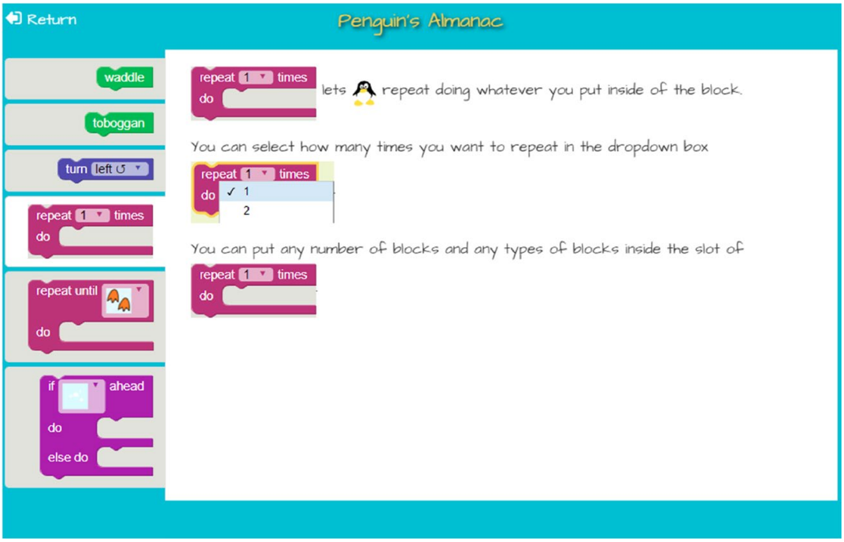
The primary dependent variable in this study (e.g., RQ1, RQ2, RQ3) was students' CT skills, measured with a pretest, a near transfer test, and a far transfer test. Both the pretest and near transfer tests were adapted from the Computational Thinking Test or CTt (Román-González et al., 2017). The original version of the instrument has 28 multiple choice questions measuring six components of CT (Cronbach's $\alpha = 0.79$). Questions targeting *while condition* and *function* were dropped because *Penguin Go* did not cover these concepts. The adapted version of CT pretest, 17 multiple choice questions, was used in the pretest, which required 20 min to complete. The average score for CTt was 8.59 ($SD = 2.53$) out of 17. The CT pretest reliability was Cronbach's $\alpha = 0.65$ [0.55, 0.76], possibly due to the reduction in the number of questions from the original CTt instrument.

Near transfer (NT) posttest

The multiple-choice questions for the NT posttest were presented in the context of *Penguin Go* so that the near transfer context was identical to the learning context (Fig. 4B). The solutions to each question in the NT posttest were also identical to those in the CT pretest to make the student scores comparable. The average score for NT posttest was 9.49 ($SD = 2.76$) out of 17. The Cronbach's α was at 0.62 [0.50, 0.74].

Screenshot Illustrations of the Levels, Almanac, and Supports in Penguin Go

A



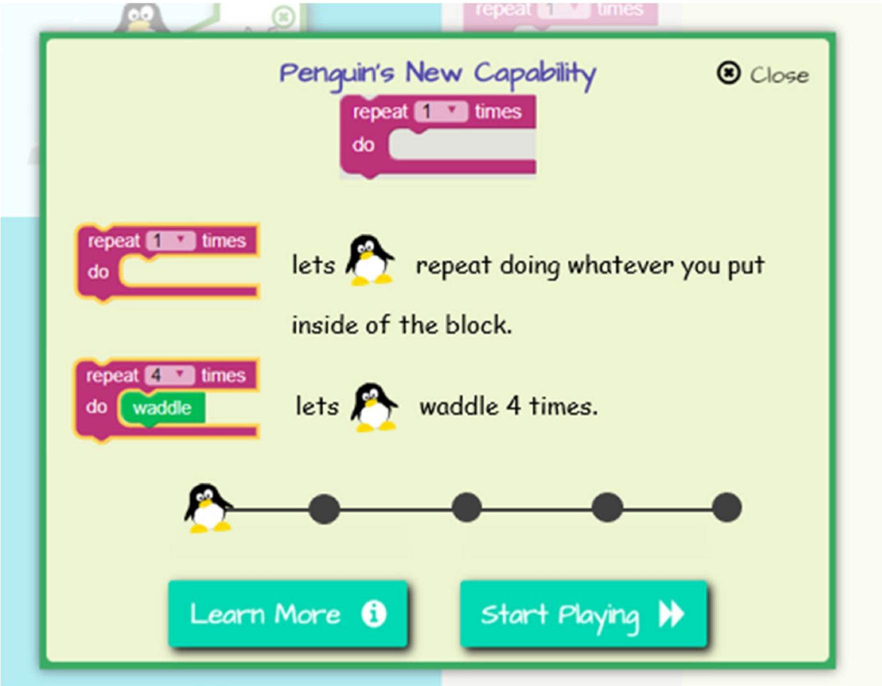
Example of the Digital Almanac.

B

The First Level of Repeat Times.

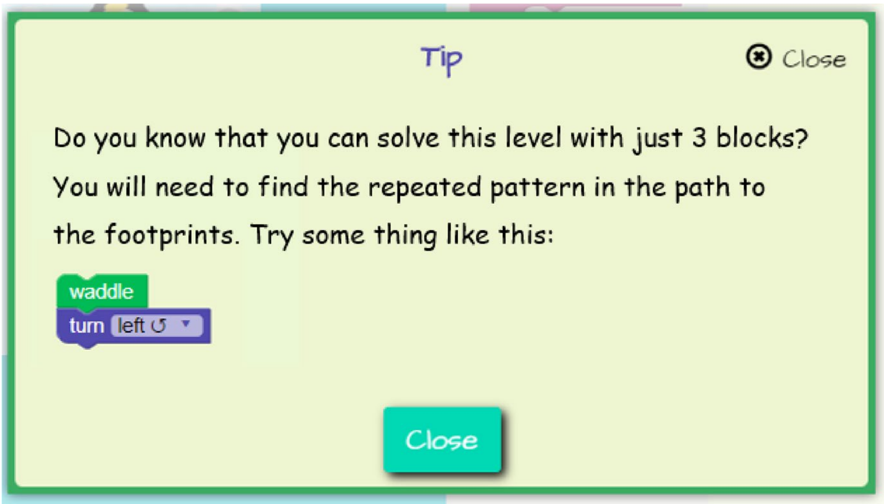
Fig. 3 Screenshot illustrations of the levels, almanac, and supports in Penguin Go. **A** Example of the digital almanac. **B** The first level of repeat times. **C** Example of concept-specific prompt for repeat times. **D** The level-specific tip in *Seeing the Pattern*

C



Example of Concept-Specific Prompt for Repeat Times

D



The Level-Specific Tip in *Seeing the Pattern.*

Fig. 3 (continued)

Table 1 Cognitive supports in Penguin Go

Cognitive support	Description	Timing	Theoretical foundation
Concept-specific prompt	A pre-concept introductory and interactive prompt that introduces the new block	Upon starting the first level of a concept	Advice (advance organizer/pre-training)
Level-specific advice	A partial worked example that (1) encourages the use of a minimum number of blocks, (2) presents the target concept, and (3) presents other blocks that nest inside the loop	Voluntary access	Modeling

Examples of Multiple-Choice Questions on the CTt, NTt, and FTt Posttest Instruments Testing the Same CT Skill

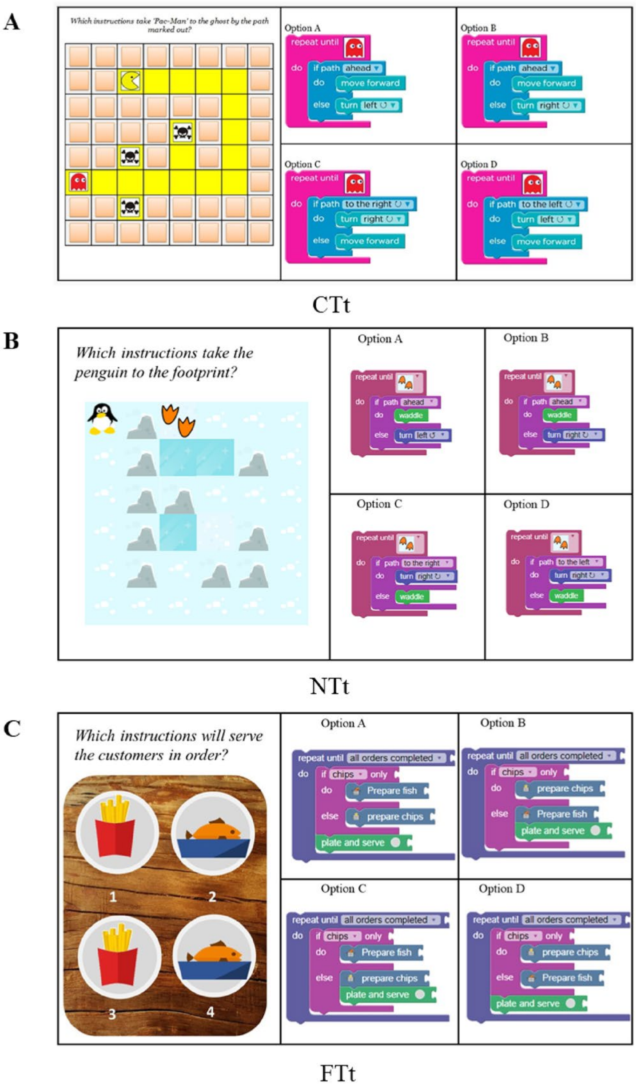


Fig. 4 Examples of multiple-choice questions on the CTt, NTt, and FTt posttest instruments testing the same CT skill

Far transfer (FT) posttest

A variant of the NT test was developed to produce the FT posttest by replacing the context and task demands of playing Pacman or Penguin Go -to with the task demands of serving fish and chips in a food truck. The task is designed to be a problem-solving context involving a series of behaviors (i.e., preparing fish, preparing chips, plating the food, and serving). These behaviors can be sequenced and automated using the same coding blocks that are used to move the penguin and Pac-Man. Therefore, the underlying computational competencies are held constant while the context of the tasks are completely different from guiding the movement of a game character (e.g., Pac-Man in Fig. 4A or Penguin in Fig. 4B).

Figure 4 shows an example of one problem and its associated solution. Figure 4A is the problem presented in the CTt, and the question stem is "Which instructions take 'Pac-Man' to the ghost via the marked path?" Fig. 4C is the problem presented in NT test and the question stem is "Which instructions take the penguin to the footprint?" Fig. 4C is the problem presented in the FT test, and the question stem is "Which instruction will serve the customers in order?" All three solutions require the use of the if-else structure nested in a repeat-until block. The average score for the FT posttest was 8.63 ($SD = 3.74$) out of 17 with Cronbach's $\alpha = 0.78$ [0.70, 0.85].

In-game performance

In-game performance was measured in reference to: (1) the number of levels completed; and (2) the level of abstraction based on the number of blocks minus the minimal number of possible blocks. The formula is shown below, where n is the number of levels completed. Using this formula, in-game performance increases as the student completes more levels and completes them by creating more elegant solutions with using the minimum possible number of blocks and minimum redundancies in the program blocks. Each additional block a student uses over the minimum possible number of blocks in one specific level decreases the student's in-game performance score on that level.

$$performance = \sum_{i=1}^{n} \frac{1}{(current\,\#block\,used - minimal\,\#\,block\,possible) + 1}$$

Demographic questionnaire

To provide a description of the participants, a demographic questionnaire was administered to collect: (1) demographic information (i.e., gender, grade, ethnicity); (2) video gaming experience; and (3) prior coding experience.

Perceived learning and enjoyment questionnaire

We used the instruments developed and validated by Barzilai and Blau's (2014) study was administered to measure: (1) Perceived learning defined as "a set of beliefs and feelings one has regarding the learning that has occurred" (Caspi & Blau, 2011, p. 285); and (2) enjoyment or level of positive reaction to the learning experience (Ritterfeld & Weber, 2006). Using 5-point Likert scale items anchored with Strongly Disagree and Strongly agree, four items were used to measure perceived learning ($\alpha = 0.90$), such as "I learned a lot form the game". Three items were used to measure enjoyment ($\alpha = 0.88$), such as "I had fun playing the game".

Results

Descriptive data

Table 2 shows the descriptive data of four cognitive (pretest, NT posttest, FT posttest, in-game performance) and two attitudinal variables (perceived enjoyment, perceived learning) and three en route variables (time viewing the almanac, concept-specific prompts, and level-specific advice). Pretest and NT posttest are normally distributed in the population while FT posttest and in-game performance are not normally distributed in the population based on Shapiro–Wilk test for normality ($W_{pretest}=0.98$, $p=0.20$; $W_{NT}=0.98$, $p=0.16$; $W_{FT}=0.95$, $p=0.005$; $W_{in\text{-}game\ performance}=0.94$, $p<0.001$). The attitudinal variables are not normally distributed in the population based on Shapiro–Wilk test for normality ($W_{enjoyment}=0.74$, $p<0.001$; $W_{learning}=0.93$, $p<0.001$) with a small standard deviation because students all reported positive experiences with the game, while the en route variables are right skewed with a large standard deviation because the durations are accumulated over time and tend to follow lognormal distribution which has a long end (Limpert et al., 2001).

Research question 1: in-game performance, perceived enjoyment, perceived learning

RQ1 is examined the Mann–Whitney's U test instead of independent *t*-test due to the violation of normality assumption in most of the variables (i.e., FT posttest, in-game performance, perceived enjoyment, and perceived learning). The Mann–Whitney's *U* test reveal no significant difference in in-game performance ($W=647$, $p=0.90$) between the CS group ($Mdn=7.44$) and the NS group ($Mdn=8.64$). The mean scores were $M=15.92$ in the CS group, and $M=16.40$ in the NS group, with effect size $d=-0.27$. No significant difference was found in perceived enjoyment ($W=757$, $p=0.81$) between the CS group ($Mdn=14.5$) and the NS group ($Mdn=15$), with effects size of $d=0.27$. Both groups reported high levels of enjoyment playing the game. No significant difference was found in perceived learning ($W=687$, $\mathrm{p}=0.36$) between the CS group ($Mdn=16$) and the NS group ($Mdn=15$) with effect size of $d=0.05$. Students in both groups felt that they learned something from the game.

Within the CS group, the average number of times and the average amount of time students viewed the *concept-specific support* were 5.5 times ($SD=1.85$) and 100.82 s ($SD=150.00$), respectively. The average number of times and the average amount of time students viewed *level-specific advice* were 2.6 times ($SD=3.65$) and 24.34 s ($SD=37.58$), respectively.

Research question 2: the difference between near and far transfer between groups

Although the FT posttest potentially do not distribute normally in the population, when the sample size is sufficiently large (in this study's case, $n=79$), it can be assumed that the mean of samples follows normal distribution per *Central Limit Theorem*. In

Table 2 Means, standard deviations, and correlations with confidence intervals

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. Pretest	8.59	2.53								
2. Near transfer	9.49	2.76	0.43**							
3. Far transfer	8.63	3.74	0.41**	0.52**						
4. In-game performance	8.77	3.78	0.03	− 0.15	− 0.18					
5. Perceived enjoyment	13.18	2.63	0.17	0.29*	0.15	0.03				
6. Perceived learning	15.34	3.69	− 0.04	0.16	− 0.05	0.14	0.43**			
7. Almanac access	16.84	28.73	0.20	0.27*	0.24*	− 0.11	0.18	0.18		
8. Concept-specific prompt access duration	100.82	150.00	− 0.01	− 0.00	− 0.05	− 0.15	− 0.08	− 0.00	0.02	
9. Level-specific advice access duration	24.34	37.58	0.30**	0.05	0.22*	− 0.08	0.16	0.01	0.16	0.07

*Indicates $p < 0.05$, **indicates $p < 0.01$

Group Means of Students' Performance on Pretest, Near Transfer, and Far Transfer with Standard Error by Condition.

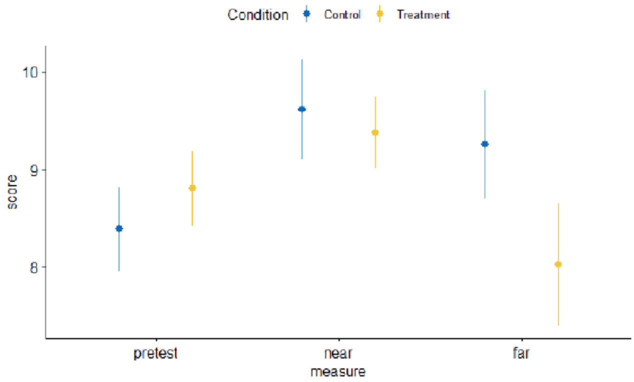


Fig. 5 Group means of students' performance on pretest, near transfer, and far transfer with standard error by condition

addition, Mauchly's test of Sphericity showed that the variance across different groups is not significantly different (*Mauchly's* $W = 0.962$, $p = 0.24$). Therefore, repeated measure ANOVA was used to compare students' CT skill change from pretest to near transfer and far transfer between CS group and NS group. The performance on the pretest ($p = 0.47$, Cohen's $d = 0.17$) and near transfer test ($p = 0.70$, Cohen's $d = -0.09$) between the two groups was not statistically significant. CS group is moderately lower on far transfer test with significance ($p = 0.05$, Cohen's $d = -0.33$) compared to the NS group. The overall effect of adding additional support to the game did not have a significant effect across measures ($F(1, 75) = 0.01$, $p = 0.99$, partial $\eta^2 < 0.001$). Overall, the results from the ANOVA model showed a significant difference among pretest, near transfer, and far transfer ($F(1, 150) = 3.67$, $p = 0.03$, Fig. 5).

The post hoc comparisons between three measures with Bonferroni correction revealed that the near transfer score was significantly higher than the pretest score ($p = 0.02$, $d = 0.34$) and the far transfer test score of marginal significance ($p = 0.07$, $d = 0.26$) across conditions and schools. The far transfer score showed no difference compared to pretest score and not significantly different ($p = 1.00$, $d = 0.01$).

Research question 3: the effect of support access on near and far transfer posttests

With multiple linear regression, the near transfer model was significant ($F(5, 73) = 4.40$, $p = 0.001$), with an R^2 of 23.17%. The results of each predictor are shown in Table 3. Neither duration of access to concept specific ($t = 0.34$, $p = 0.73$) nor level-specific support ($t = -0.09$, $p = 0.49$) in the CS group were significant predictors for near transfer performance. Pretest performance was a significant predictor of near transfer performance. Pretest score increases of 1 point resulted in near transfer score increases by 0.45 point ($t = 3.83$, $p = 0.0003$). Second, the two groups performed equally well on the near transfer test ($t = -0.42$, $p = 0.53$). The NS condition performed slightly better.

Table 3 Multiple regression analyses predicting near and far transfer performance

Predictor	Near transfer model			Far transfer model		
	β	t	p-value	β	t	p-value
(Intercept)	5.51	5.19	< 0.001***	4.66	3.34	0.001**
Pretest	0.45	3.83	< 0.001***	0.51	3.27	0.002**
Condition (treatment)	− 0.42	− 0.62	0.54	− 2.42	− 2.69	0.009**
Almanac access duration (in s)	0.02	1.94	0.05*	0.02	1.60	0.11
Concept specific prompt access duration (in s)	< 0.001	0.35	0.73	< 0.001	0.65	0.51
Level specific advice access duration (in s)	− 0.01	− 0.689	0.50	0.03	1.92	0.06
Model fit	$R^2 = 0.232$**			$R^2 = 0.281$**		

*Indicates $p < 0.05$, **indicates $p < 0.01$, ***indicates $p < 0.001$

Third, the duration of almanac access (which was available for both groups) significantly predicted the near transfer test ($t = 1.93$, $p = 0.05$). Specifically, engaging with the almanac for one more second increases near transfer scores by 0.02 points.

With multiple linear regression, the far transfer model was significant ($F(5, 73) = 5.72$, $p = 0.0002$), with an R^2 of 28.14%. The results of each predictor are shown in Table 3. The duration of access to level-specific advice predicted scores on the far transfer test ($t = 1.92$, $p = 0.06$). The effect was marginally significant. Specifically, engaging with the advice for one more second results in the far transfer score increase of 0.03 points. Similar to the near transfer model, pretest performance was a significant predictor of far transfer performance ($t = 3.28$, $p = 0.0002$). Pretest score increases of 1 point resulted in near transfer score increases by 0.51 point. Second, the NS group outperformed the CS group significantly controlling other factors constant ($t = -2.42$, $p = 0.009$). Neither accessing to almanac ($t = 1.60$, $p = 0.11$) nor level-specific support ($t = 0.65$, $p = 0.51$) were significant predictors to far transfer performance.

Discussion

In-game performance and enjoyment

This study found no significant differences in in-game performance between groups. Although the CS group solved more levels than the NS group, the difference in in-game performance was not significant when taking solution quality into consideration. This finding suggests that giving students control over the supports helped to keep the cognitive supports from being distractive and keep them from inhibiting in-game performance as it did in Leenkuil's (2006) study where presenting advice at fixed times negatively affected in-game performance. One possible reason as to why the cognitive supports did not produce better in-game performance than without cognitive supports is that the average time students in the CS group spent on cognitive supports were very sporadic and brief in total time: Within the CS group, the average number of times and the average amount of time students viewed the *concept-specific support* were 5.5 times ($SD = 1.85$) and 100.82 s ($SD = 150.00$), respectively. The average number of times and the average amount of time students viewed *level-specific advice* were 2.6 times ($SD = 3.65$) and 24.34 s ($SD = 37.58$), respectively. Another possible reason is that students may have used the supports merely as cognitive shortcuts (or cheat-sheets) to help them immediately advance to the next level without having to learn the underlying CT skill. Ke et al. (2016) found that students did not actively reflect on the learning content because successful completion of in-game tasks did not require active reflection and learning of the content. As a result, students do not necessarily need to engage in meaningful gameplay, which should ultimately lead to better in-game performance. Similarly, although each level of Penguin Go corresponds to a specific CT skill, completing each level in Penguin Go in this study did not force students to apply the featured skills nor did the game reward students for creating more elegant solutions.

The results also show no significant differences in perceived enjoyment and perceived learning even when the cognitive supports presented in the CS group could have negatively affected level of enjoyment. Given that learner autonomy contributes to learner enjoyment (Johnson et al., 2016; Ryan et al., 2006), finding no significant differences suggests that presenting cognitive supports with learner control helped to keep the cognitive supports from being distracting, from placing additional cognitive load during gameplay, and therefore, helping to make *Penguin Go* equally enjoyable as *Penguin Go* without cognitive supports. To our knowledge, no prior studies have tested the effects of presenting cognitive supports with learner control on student enjoyment and hence, no prior findings are available to corroborate this finding or to suggest possible reasons as to why the CS students did not report levels of enjoyment that was higher than that reported from the NS students. Nevertheless, one possible explanation as to why the level of enjoyment was not higher in the CS condition is because the cognitive supports produced no significant differences in in-game performance. If the students made better use of the cognitive supports to increase in-game performance (perhaps by designing the game to award students for producing more elegant solutions, for example), the cognitive supports might then produce higher levels of enjoyment. In addition, both groups reported a similar level of perceived learning, which suggests again that the cognitive supports did not make a substantial impact because of the limited use of the supports.

Effects on near and far transfer

The results show no significant differences in *near* transfer between groups, but the NS group scored significantly higher on *far* transfer with medium effect size. This finding potentially suggests that the cognitive supports did not improve near transfer and possibly inhibited rather than facilitated far transfer. O'Rourke et al. (2014) reported similar negative results in game performance when they implement abstract and concrete hints into the educational games. The possible explanations as to why the CS group did not perform better on near transfer are the same as those used to explain why no differences were found on in-game performance—sporadic and low use of the cognitive supports, and the use of the supports as cognitive shortcuts over the use of reflection to learn the underlying computations skills modeled in the supports. Misusing of the supports likely limited or over-rode students' motivation to learn the knowledge required to successfully complete far transfer tasks, as has been found in prior studies (Stocco et al., 2010; Taatgen, 2013) because students in GBL environments are often task-oriented rather than the learning-oriented (Ke et al., 2016). They tend to focus on immediate needs or goals rather than reflect on the knowledge underlying the game interactions. As a result, GBL experience often leads to superficial learning and understanding (Masson et al., 2011).

One explanation as to why the supports inhibited far transfer but not near transfer in the CS group is that performance on far transfer tasks rely more heavily on using the necessary knowledge and skills to successfully complete a far transfer task than a near transfer task—knowledge and skills that students in the CS group were less likely to acquire as a result of using the supports more as cheats than supports for learning the knowledge and skills. With near transfer tasks, on the other hand, students can apply the same rote behaviors used to successfully complete in-game tasks without having to apply a deep understanding of the computational skills. Conversely, the findings also suggest that the *absence* of cognitive shortcuts can help increase the likelihood that students reflect on and learn the computational skills behind the game and that their learning is more likely to be evident when they perform far transfer tasks. The supports the NS group received were mostly indirect supports (e.g., level progression arrangement and the almanac) which contain the necessary information for students to develop knowledge without giving away the solution.

Relationship between support access and transfer

Concept-specific supports

Contrary to the hypotheses, the findings showed that the amount of time the CS students used the *concept-specific* supports (where concepts are presented in the form of an advanced organizer) was not significantly associated with their performance on neither the near transfer test, nor the far transfer test. These findings contradicted the prescribed use of such supports in the instructional design literature (Mayer, 1983). However, the average number of times students accessed the prompts was so infrequent that the use of this support was not likely to produce a substantial effect. Furthermore, advanced organizers have been found overall to have a small effect on learning among middle school students because of their limited cognitive abilities (Luiten et al., 1980). With *Penguin Go*, students would have to allocate more time and attention to learning and applying the content presented in the advance organizer to their gameplay to achieve more transfer. The findings in this study suggests that the CS students chose not to use the target block to produce more elegant code and hence had little motivation to direct full attention to the concept-specific supports presented at the beginning of each level.

Level-specific advice

The findings showed that the amount of time spent voluntarily accessing the *level-specific advice* (i.e., goal setting, partial worked examples) predicted performance on far transfer, but not near transfer. One reason as to why more time spent on accessing the level-specific advice leads to more far transfer is that unlike the content-specific advice, the *level-specific advice* in this study was displayed to the CS students only when they chose to view them. As a result, the students that were more focused on finding a better solution than on completing the in-game tasks (and spent more time accessing the worked examples presented in the level-specific advice) performed better on far transfer. The second reason for this finding is that the level-specific advice focused explicitly on the problem-solving process, with each tip providing a goal for redesigning solutions and a partially worked example that models the problem-solving process. The use of goal setting and partially worked examples is intended to encourage students to distill abstract knowledge through problem-solving

(Akama, 2006; Mulder et al., 2016). These findings are consistent with prior findings that show how supports targeted on the problem-solving process have a positive impact on far transfer (Fuchs et al., 2003; Plass et al., 2012). Overall, these findings suggest that the more time students spend using *level-specific advice*, the more likely they learn the underlying skills needed to perform better on far transfer tasks. Even though the majority of students in the CS group were likely more focused on completing the game tasks and using the supports as shortcuts, the significant correlation between the amount of usage time and performance on far transfer indicates that there were students that *did* use the supports to learn the target skills. These findings show how students' use of supports is as equally if not more important than the function and content presented in the supports.

General game supports

The findings showed that the amount of time accessing the Digital Almanac across both groups did predict near transfer performance. The almanac illustrates how to play the game by presenting information on how to use and manipulate each block. Although how to play the game do not directly relate to the target skills students are tasked to learn, knowledge of gameplay is still necessary for the process of knowledge development (Habgood & Ainsworth, 2011; Ke, 2016). For this reason, students that used the general game support more often performed better on the near transfer performance (where the same mechanics apply), but not on far transfer (where the game mechanics do not apply).

Design implications

Findings of this study suggest the benefit of accessing game supports in the environment. However, simply exposing students to more supports or instructional content not only may sacrifice the joy of gameplay (Shute et al., 2021), but also jeopardize students' autonomy in learning (Proulx et al., 2016). Therefore, it may also be necessary to design the game mechanisms that encourage students to actively use the supports in ways that focus students' attention to reflect on and learn the skills described and modeled in the supports (rather than using the supports only to help find the quickest and simplest way to complete the game). One way to encourage more frequent and more reflective use of the supports is to increase the level of challenge in the game**.** Although the constraints used in the prior study to increase the level of challenge (Zhao & Shute, 2019) did not improve students' CT skills compared to the non-constraint group, such constraints could potentially improve CT skills when used in conjunction with cognitive supports that provide goal setting and worked examples. An alternative method to increasing challenge is to change the goal structures within the game by rewarding points and/or special penguin powers based on the number of trials performed to complete the task and the quality of the coded solutions. In addition, rewards can also be based on the strategic use of the supports where, for example, more points are awarded with less frequent, but more time spent viewing the supports.

Altogether, cognitive supports can be designed and implemented to potentially increase learning transfer without reducing enjoyment by (1) presenting congratulatory pages after the completion of a level that identifies the specific skill they just learned and elicit self-explanations; (2) present eye-catching pages that "unlock" new capabilities (specific types of blocks, for example) to enable students to focus on applying the target skills and creating more elegant solutions to earn rewards; (3) impose appropriate constraints that require students to use target blocks to create more elegant solutions which can then motivate and require students to use the supports to learn and use the target skill; and (4) help students set up appropriate goals before and during game play by displaying task goals and live game scores. These design ideas can be tested and help build a systematic guideline for designing cognitive supports in GBL in future research and development.

Limitations and future directions

One of the limitations of this study was that the CS students did not use the supports extensively during gameplay. The findings highlight the importance of making cognitive supports more integral to the task demands and constraints within a game rather than extraneous to the game so that students can be more active in learning and using the instructional content to solve and work through challenges in the game. Future studies are needed to develop, apply, and more systematically and/or comprehensively test specific design guideline or principles for designing supports that can facilitate learning while minimizing extraneous load to prevent losses in perceived level of fun and autonomy in game-based learning. For example, future studies can apply the modality principle of multimedia design by replacing the text-based prompts used explain the graphic visuals with audio narration. Furthermore, future studies will need to use other methods (e.g., questionnaires, verbal protocol analysis) to determine to what extent students used the supports merely as a game cheat and mental shortcuts. Finally, the findings in this study were based on implementing the game over a limited number of classroom sessions with a limited sample of students in terms of their ethnic backgrounds, age, prior experience with computer programming, and cognitive abilities and the CT measures showed relatively low internal consistency. The level of confidence in the findings of future studies can be increased by increasing the duration of gameplay and by conducting studies to validate the instruments (including any modifications made on prior validated instruments as was done in this current study) used in this study to measure CT skills with large and diverse samples. Future studies can focus on investigating how best to integrate such game-based learning experiences into K-12 Computer Science curriculums to test its efficacy on learning transfer when implemented over longer and extended time periods and across multiple situations.

Conclusions

Overall, this study provides game designers and education researchers with empirical evidence to identify how and when an educational game can foster CT skills and improving Computer Science education in K-12. The findings in this study show how and when an educational game like *Penguin Go* can help develop students' CT skills and learning transfer, especially far transfer, through game play. This study also illustrates some of the unique challenges in developing educational games that can help students achieve substantive gains across multiple learning outcomes (enjoyment, game performance, near transfer, far transfer). Students' ability to achieve transfer is limited by their ability to formalize the knowledge and recognize the similarity between different problem contexts and scenarios. The study's findings reiterate how important it is that future GBL studies not only assess students on in-game performance and near transfer but also to assess students on far transfer. Most of all, future GBL studies can build on this study's findings to systematically identify, design, and test more effective and interactive supports that facilitate the development of transferrable skills.

Acknowledgements We thank Dr. Valerie Shute, Dr. Vanessa Dennen, Dr. James Klein, Dr. Betsy Becker, Dr. Linlin Sha, and Dr. Xufeng Niu for providing valuable consultation and precious feedback. We also thank Dr. Weinan Zhao for initially creating Penguin Go and Yongqing Zheng for the development effort. We thank Dr. Ginny Smith, Demetrius Rice, Chih-pu Dai, Curt Fulwider, and Renata Kuba for helping with the recruitment and study sessions. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Conflict of interest This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was conducted with university Institutional Review Board (IRB) approval. Being constrained by the human subject protection policies and the qualitative nature of the data, the original study data are not open. Anonymous analysis results are accessible upon request.

Consent to participate Informed consent was obtained from all individual participants included in the study.

References

Abdul Jabbar, A. I., & Felicia, P. (2015). Gameplay engagement and learning in game-based learning. *Review of Educational Research, 85*(4), 740–779. https://doi.org/10.3102/0034654315577210

Akama, K. (2006). Relations among self-efficacy, goal setting, and metacognitive experiences in problem-solving. *Psychological Reports, 98*, 895–907.

Arena, D. A., & Schwartz, D. L. (2014). Experience and explanation: Using videogames to prepare students for formal instruction in statistics. *Journal of Science Education and Technology, 23*(4), 538–548. https://doi.org/10.1007/s10956-013-9483-3

Atmatzidou, S., & Demetriadis, S. (2016). Advancing students' computational thinking skills through educational robotics: A study on age and gender relevant differences. *Robotics and Autonomous Systems, 75*, 661–670. https://doi.org/10.1016/j.robot.2015.10.008

Azevedo, R., & Aleven, V. (2013). Metacognition and learning technologies: An overview of current interdisciplinary research. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 1–16). Springer. https://doi.org/10.1007/978-1-4419-5546-3_1

Bainbridge, K., Shute, V., Rahimi, S., Liu, Z., Slater, S., Baker, R. S., & D'Mello, S. K. (2022). Does embedding learning supports enhance transfer during game-based learning? *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2021.101547

Barzilai, S., & Blau, I. (2014). Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. *Computers & Education, 70*, 65–79. https://doi.org/10.1016/j.compedu.2013.08.003

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education, 84*(6), 740–756. https://doi.org/10.1002/1098-237X(200011)84:6%3c740::AID-SCE4%3e3.0.CO;2-3

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. *Proceedings of the 2012 annual meeting of the American educational research association*, *1*. https://dam-prod.media.mit.edu/x/files/~kbrennan/files/Brennan_Resnick_AERA2012_CT.pdf

Caspi, A., & Blau, I. (2011). Collaboration and psychological ownership: How does the tension between the two influence perceived learning? *Social Psychology of Education, 14*(2), 283–298. https://doi.org/10.1007/s11218-010-9141-z

Chase, C. C., Malkiewich, L., & Kumar, A. S. (2019). Learning to notice science concepts in engineering activities and transfer situations. *Science Education, 103*(2), 440–471. https://doi.org/10.1002/SCE.21496

Clark, D. B., Nelson, B. C., Chang, H. Y., Martinez-Garza, M., Slack, K., & D'Angelo, C. M. (2011). Exploring Newtonian mechanics in a conceptually-integrated digital game: Comparison of learning and affective outcomes for students in Taiwan and the United States. *Computers & Education, 57*(3), 2178–2195. https://doi.org/10.1016/j.compedu.2011.05.007

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning. *Review of Educational Research, 86*(1), 79–122. https://doi.org/10.3102/0034654315582065

Dondlinger, M. J. (2007). Educational video game design: A review of the literature. *Journal of Applied Educational Technology, 4*(1), 21–31.

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., & Schroeter, K. (2003). Enhancing third-grade students' mathematical problem solving with self-regulated learning strategies. *Journal of Educational Psychology, 95*(2), 306–315. https://doi.org/10.1037/0022-0663.95.2.306

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher, 42*(1), 38–43. https://doi.org/10.3102/0013189X12463051

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences, 20*(2), 169–206. https://doi.org/10.1080/10508406.2010.508029

Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2004). Resources, framing, and transfer. In J. P. Mestre (Ed.), *Transfer of Learning: Research and Perspectives* (pp. 89–120). Information Age Publishing. https://s3.amazonaws.com/academia.edu.documents/43241679/Transfer_chapter_final.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1552515695&Signature=T1%2FA5h7DuFe9ShZX3oxpUttRbcY%3D&response-content-disposition=inline%3Bfilename%3DResources_framing_and

Hooshyar, D., Pedaste, M., Yang, Y., Malva, L., Hwang, G. J., Wang, M., Lim, H., & Delev, D. (2020). From gaming to computational thinking: An adaptive educational computer game-based learning approach. *Journal of Educational Computing Research, 59*(3), 383–409. https://doi.org/10.1177/0735633120965919

Hsu, T. C., Chang, S. C., & Hung, Y. T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education, 126*, 296–310. https://doi.org/10.1016/J.COMPEDU.2018.07.004

Israel-Fishelson, R., & Hershkovitz, A. (2020). Persistence in a game-based learning environment: The case of elementary school students learning computational thinking. *Journal of Educational Computing Research, 58*(5), 891–918. https://doi.org/10.1177/0735633119887187

Johnson, D., Gardner, J., & Sweetser, P. (2016). Motivations for videogame play: Predictors of time spent playing. *Computers in Human Behavior, 63*, 805–812. https://doi.org/10.1016/j.chb.2016.06.028

Jonassen, D. H. (1999). Designing constructivist learning environments. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory* (pp. 215–239). Lawrence Erlbaum.

Kafai, Y. B., & Proctor, C. (2021). A revaluation of computational thinking in K–12 education: Moving toward computational literacies. *Educational Researcher*. https://doi.org/10.3102/0013189X211057904

Ke, F. (2016). Designing and integrating purposeful learning in game play: A systematic review. *Educational Technology Research and Development, 64*(2), 219–244. https://doi.org/10.1007/s11423-015-9418-1

Ke, F., Shute, V. J., Clark, K. M., & Erlebacher, G. (2019). *Designing dynamic support for game-based learning* (pp. 119–140). https://doi.org/10.1007/978-3-030-04339-1_6

Ke, F., Xie, K., & Xie, Y. (2016). Game-based learning engagement: A theory- and data-driven exploration. *British Journal of Educational Technology, 47*(6), 1183–1201. https://doi.org/10.1111/bjet.12314

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education, 8*(1), 13–24. https://doi.org/10.1016/J.IHEDUC.2004.12.001

Kim, Y. J. (2014). *Search for the optimal balance among learning, psychometric qualities, and enjoyment in game-based assessment* [Florida State University]. https://doi.org/10.4236/ojo.2014.48035

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Lajoie, S. P. (2008). Metacognition, self regulation, and self-regulated learning: A rose by any other name? *Educational Psychology Review, 20*(4), 469–475. https://doi.org/10.1007/s10648-008-9088-1

Leemkuil, H. (2006). *Is it all in the game? Learner support in an educational knowledge management simulation game*. Universiteit Twente, The Netherlands.

Leemkuil, H., & de Jong, T. (2012). Adaptive advice in learning with a computer-based knowledge management simulation game. *Academy of Management Learning and Education, 11*(4), 653–665. https://doi.org/10.5465/amle.2010.0141

Leonard, J., Buss, A., Gamboa, R., Mitchell, M., Fashola, O. S., Hubert, T., & Almughyirah, S. (2016). Using robotics and game design to enhance children's self-efficacy, STEM attitudes, and computational thinking skills. *Journal of Science Education and Technology, 25*(6), 860–876. https://doi.org/10.1007/S10956-016-9628-2/TABLES/7

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *Bioscience, 51*(5), 341–352. https://doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2

Low, R., Jin, P., & Sweller, J. (2010). Learners' cognitive load when using educational technology. In *Gaming and cognition* (pp. 169–188). IGI Global. https://doi.org/10.4018/978-1-61520-717-6.ch008

Luiten, J., Ames, W., & Ackerson, G. (1980). A meta-analysis of the effects of advance organizers on learning and retention. *American Educational Research Journal Summer, 17*(2), 211–218.

Masson, M. E. J., Bub, D. N., & Lalonde, C. E. (2011). Video-game training and naive reasoning about object motion. *Applied Cognitive Psychology, 25*(1), 166–173. https://doi.org/10.1002/acp.1658

Mayer, R. E. (1983). Can you repeat that? Qualitative effects of repetition and advance organizers on learning from science prose. *Journal of Educational Psychology, 75*(1), 40–49. https://doi.org/10.1037/0022-0663.75.1.40

Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology, 70*(1), 531–549. https://doi.org/10.1146/annurev-psych-010418-102744

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, *59*(1), 14–19. https://doi.org/10.1037/0003-066X.59.1.14

Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94*(1), 171–185. https://doi.org/10.1037/0022-0663.94.1.171

Morrison, J. R., Bol, L., Ross, S. M., & Watson, G. S. (2015). Paraphrasing and prediction with self-explanation as generative strategies for learning science principles in a simulation. *Educational Technology Research and Development, 63*(6), 861–882. https://doi.org/10.1007/s11423-015-9397-2

Mulder, Y. G., Bollen, L., de Jong, T., & Lazonder, A. W. (2016). Scaffolding learning by modelling: The effects of partially worked-out models. *Journal of Research in Science Teaching, 53*(3), 502–523. https://doi.org/10.1002/tea.21260

O'rourke, E., Ballweber, C., & Popovi´c, Z. P. (2014). Hint systems may negatively impact performance in educational games. In *Proceedings of the first ACM conference on learning@ scale conference*, 51–60. https://doi.org/10.1145/2556325.2566248

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Plass, J. L., Homer, B. D., Hayward, E. O., Frye, J., Huang, T.-T., Biles, M., Stein, M., & Perlin, K. (2012). The effect of learning mechanics design on learning outcomes in a computer-based geometry game. In S. Göbel, W. Müller, B. Urban, & J. Wiemeyer (Eds.), *E-learning and games for training, education, health and sports* (Vol. 7516, pp. 65–71). Springer. https://doi.org/10.1007/978-3-642-33466-5_7

Prensky, M. R. (2012). *From digital natives to digital wisdom: Hopeful essays for 21st century learning*. Corwin Press.

Proulx, J. N., Romero, M., & Arnab, S. (2016). *Learning mechanics and game mechanics under the perspective of self-determination theory to foster motivation in digital game based learning*. https://doi.org/10.1177/1046878116674399,48(1),81-97.10.1177/1046878116674399

Renkl, A., Atkinson, R. K., & Große, C. S. (2004). How fading worked solution steps works—A cognitive load perspective. *Instructional Science, 32*(1/2), 59–82. https://doi.org/10.1023/B:TRUC.0000021815.74806.f6

Ritterfeld, U., & Weber, R. (2006). Video games for entertainment and education. In P. V. J. Bryan (Ed.), *Playing video games: Motives, responses, and consequences* (pp. 399–413). Lawrence Erlbaum Associate Publishers. http://www.seriousgamessummit.com/

Román-González, M., Pérez-González, J. C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the computational thinking test. *Computers in Human Behavior, 72*, 678–691.

Royer, J. M. (1979). Theories of the transfer of learning. *Educational Psychologist*, *14*(1), 53–69. https://doi.org/10.1080/00461527909529207

Ryan, R. M., Rigby, C. S., & Przybylski, A. K. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion, 30*(4), 347–363. https://doi.org/10.1007/s11031-006-9051-8

Shen, C. -Y., & O'Neil, H. (2006). *The effectiveness of worked examples in a game-based learning environment*. Annual Conference of American Educational Research Association. https://files.eric.ed.gov/fulltext/ED491961.pdf
Shute, V. J., Ke, F., Almond, R. G., Rahimi, S., Smith, G., & Lu, X. (2019). How to increase learning while not decreasing the fun in educational games. In R. Feldman (Ed.), *Learning science: Theory, research, and practice* (pp. 327–357). McGraw Hill.
Shute, V. J., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C. P., Kuba, R., Liu, Z., Yang, X., & Sun, C. (2021). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning, 37*(1), 127–141. https://doi.org/10.1111/JCAL.12473
Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review, 22*, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003
Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in newton's playground. *The Journal of Educational Research, 106*, 423–430. https://doi.org/10.1080/00220671.2013.832970
Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior, 63*, 106–117. https://doi.org/10.1016/j.chb.2016.05.047
Stocco, A., Lebiere, C., & Anderson, J. R. (2010). Conditional routing of information to the cortex: A model of the basal ganglia's role in cognitive coordination. *Psychological Review*, *117*(2), 541–574. https://doi.org/10.1037/a0019077
Sun, L., Hu, L., Yang, W., Zhou, D., & Wang, X. (2021). STEM learning attitude predicts computational thinking skills among primary school students. *Journal of Computer Assisted Learning, 37*(2), 346–358. https://doi.org/10.1111/JCAL.12493
Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, *120*(3), 439–471. https://doi.org/10.1037/a0033138
Wang, L. (2017). *To help , or not to help? An investigation of the effects of learning support in physics playground*. Florida State University.
Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education, 60*(1), 412–425. https://doi.org/10.1016/j.compedu.2012.07.018
Zhang, L. C., & Nouri, J. (2019). A systematic review of learning computational thinking through Scratch in K-9. *Computers & Education*. https://doi.org/10.1016/J.COMPEDU.2019.103607
Zhao, W., & Shute, V. J. (2019). Can playing a video game foster computational thinking skills? *Computers & Education*. https://doi.org/10.1016/J.COMPEDU.2019.103633

Zhichun Liu is an Assistant Professor of Learning Design and Implementation Sciences at Hong Kong University and a research scientist at Kaput Center of Research and Innovation in STEM Education of University of Massachusetts Dartmouth. His areas of research are problem solving in game-based learning, development of computational thinking skills, learning analytics, and learning transfer.

Allan C. Jeong is an Associate Professor of Instructional Systems at Florida State University. His research is in the field of computer-mediated communication and computer-supported collaborative work and collaborative learning.