# Impact of Digital Game-based Learning Experiences on Reasoning Skills and Science Engagement

**Spreeha Debchaudhury**
Columbia University, New York City, USA

sd3150@tc.columbia.edu

**Abstract**: This research seeks to use games to promote scientific reasoning and inquiry skills among middle-school children. Particularly, it focuses on the Control of Variables Strategy (CVS), a fundamental component of scientific thinking which enables individuals to test hypotheses and make valid causal inferences based on evidence. The present study draws on and combines existing streams of research in the fields of science education, embodied cognition, and game-based learning to investigate the efficacy of an intervention using the Preparation for Future Learning paradigm in which students play an online text-based adventure simulation game about solving escape rooms in order to prime them for later formal instruction in CVS via a randomized-controlled trial, as well as whether structured or unstructured access to the game has a differential impact on immediate learning vs. retention after a delay. Students were given tests of scientific reasoning at baseline, immediately following the training phase, and a week thereafter. The impact of playing this game on motivation and engagement with science is also analyzed in this trial through the use of survey data. The study found significant effects of study condition on all measures, with students assigned to the game groups outperforming control students, and the unstructured play group showing the strongest performance in the post-study test while the structured play group performed the most poorly in the retention test. The unstructured play group also showed the highest level of intrinsic motivation, as well as higher self-determination and self-efficacy than the structured playing group in the science engagement survey.

**Keywords:** Embodied Cognition, Educational Technology, Game-based Learning, Scientific Reasoning

## 1. Introduction

The ability to design controlled, unconfounded experiments to test hypotheses via the Control of Variables Strategy (CVS) is fundamental to all scientific reasoning and inquiry, considered a cornerstone of critical thinking as a whole, which enables individuals to make valid causal inferences (Kuhn, 2005a). CVS is considered so crucial to science and science education, in fact, that various scientific and governmental agencies a have begun including it in student curricula (National Research Council, 2012; NGSS Lead States, 2013). However, despite its unique power and flexibility as a cognitive tool and its centrality to the entire architecture of scientific inquiry, most children do not spontaneously develop use of CVS without some form of instruction or scaffolding, and students of various ages still have difficulty manipulating variables and conceptualizing controlled experiments. For example, Kuhn et al. (1995) showed that even after sustained training, less than a quarter of 4th graders, and less than 75% of adults produced valid inferences when confronted with scientific discovery tasks.

A significant amount of research effort thus has gone into the examination of the circumstances under which the learning and transfer of CVS is best supported. In a meta-analysis by Schwichow et al. (2016), for example, two factors were found that positively impacted students' ability to master CVS: cognitive conflict and demonstration. Cognitive conflict is the Piagetian notion of challenging a student's established framework of beliefs to create a new understanding. In the context of CVS instruction, cognitive conflict involves introducing "anomalous or discrepant" information in the form of a confounded comparison such that students might be able to parse that any conclusions or interpretations made based upon it is invalid. The other factor Schwichow highlighted was demonstration, with interventions that showed demonstrations of valid vs. invalid designs having a significantly larger effect size than those which don't.

Ross (1988a), meanwhile, identified two other factors that were associated with significantly higher effect sizes in intervention studies: feedback and practice. Intervention studies in which students were given feedback about their performance on training tasks showed larger effect sizes in terms of the outcome measures, as compared to those studies where students received no feedback. Similarly, intervention studies had higher effect sizes when students were afforded the opportunity to practice the CVS skills they were learning both within a school context and outside of it, as opposed to merely one or the other.

The computerization of instructional and assessment materials in this area has also met with some success in increasing the flexibility, efficiency, and accessibility of CVS training. Klahr, Triona, and Williams (2007), for example, found virtual training tasks in CVS to be equally effective as training tasks with real physical equipment, a result replicated by various studies. Nonetheless, even in these studies, the virtual tasks undertaken by students appear to largely be an extension of a classroom lecture, merely replicating the experience of a physical task without taking advantage of the more unique qualities of the medium. One avenue of research which has attempted to do so has been within the realm of embodied cognition. Embodied cognition is a concept in Cognitive Science which suggests people create mental perceptual simulations of concepts in order to understand them (Barsalou, 2008), and work such as Triona and Klahr (2003) and Glenberg (2011) has offered evidence that the positive effects of embodiment are reproducible with imagined and surrogate embodiment, determining that computer simulations of actions engage the same mechanisms as in embodied cognition.

An educational video game can bring together several of these individual threads of research; cognitive conflict via demonstration can be introduced through a game's narrative, for instance, while a game by its very nature is inherently a feedback loop designed to respond dynamically to the player's evolving competencies. This brief review will focus on the primary insight that for a gaming intervention to be successful, it needs to be engaging and motivational on top of delivering its material accurately and efficiently.

To this end, Black, Khan, and Huang (2014) adapted the Preparation for Future Learning (PFL) framework originally established for rethinking the field of transfer to game-based learning, presenting a compelling argument for video games as perceptually rich grounding environments for learning. He argued video games can serve as an embodied foundation upon which to build formal instruction on abstract concepts, thereby improving students' understanding of the material as compared to students who only received the formal instruction. For instance, they found that students who played a historical simulation game, Civilization, were then able to learn much more from reading a difficult history textbook (in terms of propositional, procedural, imagery, and system knowledge) than students who played an unrelated game (Sim City). Building on this evidence, Han and Black (2011) discovered that adding force feedback via a simple game joystick could accentuate scientific reasoning in a physics lesson about gears and Newton's laws, respectively.

The question of why people find playing games motivating, meanwhile, and how that engagement can be captured and replicated in the classroom can be addressed by flow theory (Csikszentmihalyi, 1975) and the construct of the learner hero (Rigby & Przybylski, 2009), both key elements in the development of the present study. The learner hero construct focuses on the "why" of the question. Utilizing Self-Determination Theory, Rigby & Przybylski argue that gaming fulfills three core components of intrinsic motivation: autonomy, competence, and relatedness. The common practice of setting up the user/avatar as a heroic actor allows a game world to establish its player's capacity to achieve the goals and challenges of the game and the significance of their choice to participate (autonomy). As the player progresses through the game, more and more of the virtual world's challenges and goals are met (competence), and the hero gradually becomes beloved and trusted, with their continued actions becoming ever more important to the world of the avatar and becoming self-perpetuating (relatedness). Thus, the process of playing the game itself provides intrinsic motivation to keep playing, which the authors argue has strong applications to educational technology—which aims to make learning a similarly self-sustaining process, motivation-wise.

Flow theory can be said to address the "how" of gaming motivation. In other words, what specific design features promote the experience that characterizes immersion in a game? As described by Csikszentmihalyi (1975), flow entails a sense of meaningful action, a personal locus of control, and a focus on a goal. Research discussed in Bressler & Bozdin (2013) suggests that not only does increased flow improve learning, especially in educational technology environments, but that a flow experience in enhances intrinsic motivation to participate in the game—enabling the self-perpetuating process discussed earlier.

Therefore, the present study investigates the impact of an interactive simulation game on scientific reasoning skills and science engagement, specifically effective use of CVS.

## 2. Current Study

### 2.1 The Great Escape

The game played by students was designed by the PI for the study. Its story follows five friends who participate on a reality TV gameshow in which contestants solve the puzzles of an escape room within a time limit. The primary mechanic to this end is a "choose your own adventure"-style set of branching passages: players must make individual choices through clicks in these passages to solve puzzles and gain keys to advance the story without making enough mistakes to trigger a game-over.

The game supports learning the CVS strategy through its use of narrative. Throughout the game, the characters interact with the scientific principles of the game, and the player makes choices based on their theories. In each "escape room", the characters are presented with a puzzle to solve by gathering evidence and formulating hypotheses based on the clues they find and must set up a valid test of their hypotheses to continue.

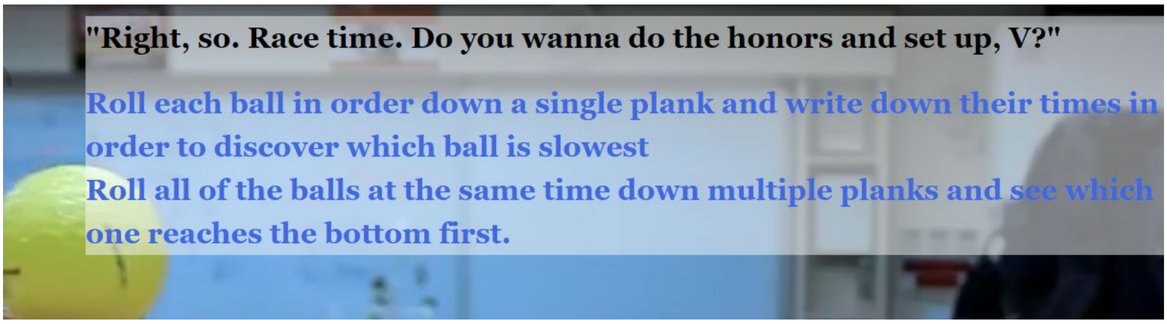


**Figure 1: Example of a choice between valid and invalid experimental setups**

As the player works through each puzzle, the characters discuss the evidence discovered via one's choices of clicks, and if the player makes a mistake, there will be a screen of further narrative with another character "correcting" them and/or realizing their own error. This shows the player why their answer is incorrect, as well as reinforcing the CVS principles at work in the game narrative—thus fulfilling the two factors identified by Ross (1988a) as significant to effect size in interventions: feedback and practice. Further, having the player choose between a valid and an invalid experimental design before observing the outcome of their choice also hits the instructional features identified by Schwichow et al (2016) as having a significant impact on effect sizes in interventions: cognitive conflict (created by introducing anomalous information, such as a confounded experiment), and demonstration.

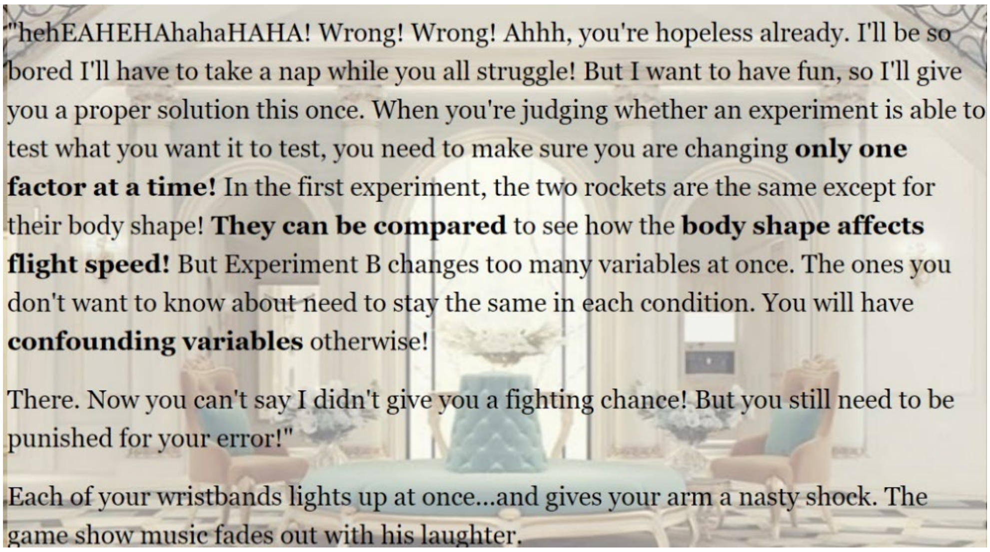


**Figure 2: Example of game feedback after an incorrect answer**

More broadly, this design of gameplay, wherein players explore puzzles while gathering clues and evaluating evidence, is inspired by the notion of educational video games as "objects to think with" (Holbert and Wilensky, 2019). Although the environments players explore are somewhat more structured and the construction and testing of hypotheses occurs under the constraints of the medium, The Great Escape is intended to reflect the idea of a game serving as a playful experience to engage in knowledge construction with, which can be used to expand one's thinking and reasoning outside the confines of the game.

## 2.2 Participants

Participants were 143 7th grade students (70 female, 73 male) from a public middle school in New Jersey. Recruitment was done through email correspondence with the superintendent and principal, who directed the

PI to a teacher willing to participate, at which point participating students had parents fill out consent forms and signed assent forms.

## 2.3 Method

Participants were given an initial test on Control of Variables Strategy (CVS) knowledge, then randomized into three groups: 1) unlimited access to the game for three consecutive days, 2) three structured play sessions over a total of three consecutive days, and 3) a control group with access to an alternative online experience—i.e. a document formally explaining the concept and application of the control of variables strategy (CVS) available for three consecutive days. After the three-day training period, all students in the first two conditions will also receive access to the document the third group received. All students will then take another CVS test following the study. One week afterward, all students will take a retention test. The assessment instruments are from the instrument developed by Chen and Klahr (1999). The baseline and post-study tests have twelve questions each, and the retention test has 24 questions, with Cronbach's alpha for each test at 0.71, 0.729, and 0.819 respectively. Each day during the training phase, the teacher gave the unstructured playing and control groups a free class period during which they could work on whatever they wished for the class, including playing the game or reading the CVS document if they wanted. Students could also view the document or play the game at their leisure at home. The structured group, meanwhile, played the game for 30 minutes at a set time for all three days.

Students in the game conditions were asked to keep a record their thoughts, feelings, and reactions while in the process of playing the game and submit them after the study.

## 2.4 Measures

The primary outcome measures in this study were the baseline test, post-study test and retention test scores from the instrument developed by Chen and Klahr (1999). There were two additional measures for the study, in the form of surveys filled out by the participants. The first was a science engagement measure, using a questionnaire designed by Glynn et al. (2011) to determine participants' engagement in and motivation to learn science. The scale has 25 items, including five subscales with five items each. The subscales are as follows, with corresponding reliabilities, assessed by Cronbach's alpha: career motivation (0.92), intrinsic motivation (0.89), self-determination (0.88), self-efficacy (0.83), and grade motivation (0.81), with an overall reliability for all 25 items of 0.92.

*Science Motivation Questionnaire II*

In order to better understand what you think and how you feel about your college science courses, please respond to each of the following statements from the perspective of "When I am in a college science course. . ."

[Response scale: ○ Never ○ Rarely ○ Sometimes ○ Usually ○ Always]

01. The science I learn is relevant to my life
02. I like to do better than other students on science tests
03. Learning science is interesting
04. Getting a good science grade is important to me
05. I put enough effort into learning science
06. I use strategies to learn science well
07. Learning science will help me get a good job
08. It is important that I get an "A" in science
09. I am confident I will do well on science tests
10. Knowing science will give me a career advantage
11. I spend a lot of time learning science
12. Learning science makes my life more meaningful
13. Understanding science will benefit me in my career
14. I am confident I will do well on science labs and projects
15. I believe I can master science knowledge and skills
16. I prepare well for science tests and labs
17. I am curious about discoveries in science
18. I believe I can earn a grade of "A" in science
19. I enjoy learning science
20. I think about the grade I will get in science
21. I am sure I can understand science
22. I study hard to learn science
23. My career will involve science
24. Scoring high on science tests and labs matters to me
25. I will use science problem-solving skills in my career

**Figure 3: Science engagement Questionnaire**

The second was a Game Experience Questionnaire, designed by IJsselsteijn, de Kort, & Poels (2013). This measure was included to determine how positively or negatively participants reacted to the game, and whether they felt motivated and engaged while playing it. The GEQ has 33 items and seven subscales, each with 5 or 6 items and the corresponding Cronbach's alphas: sensory and imaginative immersion (0.891), tension (0.821), competence (0.843), flow (0.883), negative affect (0.728), positive affect (0.838), and challenge (0.718). Both surveys use scales from 0-4 for each response, with 0 indicating the least amount of agreement with the survey item, and 4 indicating the most. Scores are calculated for each subscale separately; there is no composite score for either instrument.

Please indicate how you felt while playing the game for each of the items, on the following scale:

| not at all | slightly | moderately | fairly | extremely |
|---|---|---|---|---|
| 0 | 1 | 2 | 3 | 4 |
| < > | < > | < > | < > | < > |

1 I felt content
2 I felt skilful
3 I was interested in the game's story
4 I thought it was fun
5 I was fully occupied with the game
6 I felt happy
7 It gave me a bad mood
8 I thought about other things
9 I found it tiresome
10 I felt competent

**Figure 4: Sample questions from Game Experience questionnaire**

## 3. Research Questions and Hypotheses

### 3.1 Research Questions

1. What is the impact of game-based learning on scientific reasoning skills?
2. Do students who play an interactive narrative simulation game requiring the use of a scientific reasoning strategy (CVS) have better retention following formal instruction on the strategy when retested after a delay than those who only received the formal instruction on CVS?
3. Does structured vs. unstructured access to the game have a differential impact on immediate learning and retention after a delay?
4. Does playing the game increase students' engagement with science or motivation to learn science?

### 3.2 Hypotheses

- **Hypothesis 1:** The structured play group will have the strongest performance on the posttest.
- **Hypothesis 2:** The unlimited play will have the strongest performance on the retention test.
- **Hypothesis 3:** The two game groups will outperform the control group on the post-test and retention test.
- **Hypothesis 4:** Playing the game will improve science engagement.

## 4. Results

The CVS tests were analyzed via logistic regression—specifically the case of binomial response, which examines the number of successes in n trials with constant probability of success. This corresponds to an investigation of how many questions are answered correctly (successes) out of a total of 12 similar questions (trials) with two options each. This method was selected due to the bounded range of values of the data, which made the sample a better fit for the binomial distribution than the normal distribution used for ANCOVA. Survey data did not show significant deviations from normality and was thus analyzed with MANOVA.

All subjects' responses were anonymized, with the only identifier for each data point the IP address of the participant. However, the school's network did not retain unique IP addresses for each student terminal, so individual student data ended up completely de-identified, and data from the baseline test was unable to be matched with the post-study test and follow-up retention test data for analysis. Because of this, the three tests have been analyzed separately.

## 4.1 CVS Tests Descriptive Statistics

Mean score on the baseline test in the unstructured playing condition was the lowest, at 7.65. The structured playing condition was in the middle at 8.44. The control condition was the highest at 9.16.

The mean on the post-study test in the unstructured play group, however, was the highest, at 9.57. The mean on the post-study test in the structured play group was again in the middle, at 9.15. The control group scored the lowest, this time, at 8.65.

The structured play group had the lowest retention test scores, with a mean of 17.05. The unstructured play group and the control group, meanwhile, had nearly identical overall means, at 18.68 and 18.69, respectively.

## 4.2 Regression Results

**Table 1: Parameter estimates for logistic regression on baseline test scores**

| | | | 95% Wald C.I. | | Hypothesis Test | | | | 95% Wald C.I. for Exp(B) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | B | SE | Lower | Upper | Wald Chi-Square | df | *p* | Exp(B) | Lower | Upper |
| (Intercept) | 1.17 | .10 | .97 | 1.38 | 128.03 | 1 | .00 | 3.23 | 2.64 | 3.96 |
| [Group=unstructured] | -.61 | .14 | -.89 | -.33 | 18.07 | 1 | .00 | .54 | .41 | .72 |
| [Group=structured] | -.31 | .13 | -.57 | -.05 | 5.49 | 1 | .02 | .73 | .57 | .95 |
| [Group=control] | 0[a] | . | . | . | . | . | . | 1 | . | . |
| (Scale) | 1[b] | | | | | | | | | |

Events: Pretest

Trials: 12

Model: (Intercept), Group

As seen in Table 1, instead of being identical, the data from the baseline test shows a significant effect of group, where the first group (unstructured/unlimited play) did the most poorly, 46% less likely to get any particular question correct than the third (Wald chi square = 18.067, $p < 0.001$), and the second group (structured game play) 27% less likely than the third to get any particular question correct (Wald chi square = 5.490, $p = 0.019$).

**Table 2: Pairwise comparisons for baseline test scores**

| | | | | | | 95% Wald C.I. for Difference | |
|---|---|---|---|---|---|---|---|
| (I) Group | (J) Group | Mean Difference (I-J) | SE | df | *p* | Lower | Upper |
| unstructured | structured | -.07 | .03 | 1 | .06 | -.13 | .00 |
| | control | -.13[a] | .03 | 1 | <0.001 | -.20 | -.06 |
| structured | unstructured | .07 | .03 | 1 | .06 | .00 | .13 |
| | control | -.06 | .03 | 1 | .05 | -.12 | .00 |
| control | unstructured | .13[a] | .03 | 1 | < 0.001 | .06 | .20 |
| | structured | .06 | .03 | 1 | .05 | .00 | .12 |

Pairwise comparisons of estimated marginal means for groups based on the events/trials proportion. Significance measured with Bonferroni adjustment.

Events: Pretest

Trials: 12

Pairwise comparisons (Table 2) back up this effect of group, showing that the unlimited play group (Group 1) performed more poorly than the control (Group 3), significant at the 0.001 level even after Bonferroni correction for multiple comparisons. The second (structured play) group also did more poorly than the control, though that is only marginally significant in the pairwise comparisons.

**Table 3: Parameter estimates for logistic regression on post-study test scores**

| | | | 95% Wald C.I. | | Hypothesis Test | | | | 95% Wald C.I. for Exp(B) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | B | SE | Lower | Upper | Wald Chi-Square | df | *p* | Exp(B) | Lower | Upper |
| (Intercept) | .95 | .10 | .76 | 1.14 | 93.51 | 1 | <.001 | 2.58 | 2.13 | 3.13 |
| [Group=1] | .42 | .15 | .12 | .72 | 7.50 | 1 | <.01 | 1.52 | 1.13 | 2.06 |
| [Group=2] | .39 | .16 | .09 | .70 | 6.41 | 1 | .01 | 1.48 | 1.09 | 2.01 |
| [Group=3] | $0^a$ | . | . | . | . | . | . | 1 | . | . |
| (Scale) | $1^b$ | | | | | | | | | |

Events: Posttest

Trials: 12

**Table 4: Pairwise comparisons for post-study test scores**

| | | | | | | 95% Wald C.I. for Difference | |
|---|---|---|---|---|---|---|---|
| (I) Group | (J) Group | Mean Difference (I-J) | SE | df | Bonferroni *p* | Lower | Upper |
| unstructured | structured | .00 | .027 | 1 | 1.00 | -.06 | .07 |
| | control | $.08^a$ | .027 | 1 | .016 | .01 | .14 |
| structured | unstructured | .00 | .027 | 1 | 1.00 | -.07 | .06 |
| | control | $.07^a$ | .028 | 1 | .03 | .01 | .14 |
| control | unstructured | $-.08^a$ | .027 | 1 | .016 | -.14 | -.01 |
| | structured | $-.07^a$ | .028 | 1 | .03 | -.14 | -.01 |

Pairwise comparisons of estimated marginal means based on the events/trials proportion

Events: Posttest

Notably, the pattern from the baseline test completely reverses in the post-study test, and the unlimited play group outperforms both other groups, becoming 150% more likely to answer a question correctly than the control group (Wald chi square = 7.499, $p < 0.01$). As before, this is corroborated by the pairwise comparison (Table 4), which shows the unlimited play group with a higher probability of answering questions correctly than the other two groups, significant at the 0.05 level after Bonferroni correction. It's also worth mentioning that the analysis for the post-test scores was performed twice. During the study, the students in the structured play group (Group 2) were interrupted during the post-test by a fire alarm, and 27 of them did not finish the test before having to leave the building. Since this occurred on a Friday, these students could not complete the test until the following Monday, resulting in a delay of three days between the training phase and the post-test for these students. The analysis was therefore re-done with the students who were affected by the fire alarm excluded, with the results reported in Table 3. This reanalysis largely supports the results of the original, which are not reported here out of space considerations. However, while pairwise comparisons in the original analysis do not show that the structured play group was significantly different from the control group, the analysis with the fire alarm students excluded does, at the 0.05 level with Bonferroni correction.

**Table 5: Parameter estimates on logistic regression for post-study scores**

| | | | 95% Wald C.I. | | Hypothesis Test | | | | 95% Wald C.I. for Exp(B) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | B | SE | Lower | Upper | Wald Chi-Square | df | *p* | Exp(B) | Lower | Upper |
| (Intercept) | 1.26 | .08 | 1.09 | 1.42 | 228.94 | 1 | <.001 | 3.52 | 2.99 | 4.14 |
| [Group=1] | -.002 | .12 | -.23 | .23 | .00 | 1 | .985 | 1.00 | .79 | 1.26 |
| [Group=2] | -.36 | .10 | -.56 | -.16 | 12.52 | 1 | <.001 | .70 | .57 | .85 |
| [Group=3] | $0^a$ | . | . | . | . | . | . | 1 | . | . |
| (Scale) | $1^b$ | | | | | | | | | |

Events: Retention

Trials: 24

The study also found a significant effect of group in the retention test scores, as seen in Table 5. However, in this case, the difference was that the second group (structured play) performed significantly more poorly than either the unstructured play group or the control group, 31% less likely to answer any particular question correctly than the control group (Wald chi square = 12.520, $p < 0.001$). Again, this was confirmed by the pairwise comparisons (Table 6), which showed no significant difference between the unstructured play and control groups, but a highly significant difference between the unstructured play and structured play groups, as well as one between the structured play group and the control group.

**Table 6: Pairwise comparisons for retention test scores**

| (I) Group | (J) Group | Mean Difference (I-J) | SE | df | Bonferroni p | 95% Wald C.I. for Difference Lower | Upper |
|---|---|---|---|---|---|---|---|
| unstructured | structured | .07[a] | .019 | 1 | .001 | .02 | .11 |
| | control | .00 | .020 | 1 | 1.00 | -.05 | .05 |
| structured | unstructured | -.07[a] | .019 | 1 | .001 | -.11 | -.02 |
| | control | -.07[a] | .019 | 1 | .001 | -.11 | -.02 |
| control | unstructured | .00 | .020 | 1 | 1.00 | -.05 | .05 |
| | structured | .07[a] | .019 | 1 | .001 | .02 | .11 |

Pairwise comparisons of estimated marginal means based on the events/trials proportion

Events: Retention

Trials: 24

## 4.3 Survey Data: MANOVA Results

There was a significant difference between the groups on the combined dependent variables, F (6, 172) = 2.667, p = 0.017. Planned contrasts show that there is a significant difference between the unstructured play and structured play groups on self-determination and self-efficacy (p < 0.01), and between unstructured play and control on intrinsic motivation and self-efficacy (p < 0.05), though when adjusted with Tukey's HSD correction for multiple comparisons, only the effect of group on self-determination remains, with Group 1 having a higher mean than Group 2 by about 2.64, p < 0.05.

**Table 7: Planned contrasts for survey data**

Group Simple Contrast[a]

Level 2 vs. Level 1

| Dependent Variable | Contrast Estimate | Hypothesized Value | Difference (Estimate - Hypothesized) | SE | p | 95% C.I. for Difference Lower Bound | Upper Bound |
|---|---|---|---|---|---|---|---|
| Intrinsic_motivation | -1.54 | 0 | -1.54 | 1.01 | .13 | -3.558 | .475 |
| self_determination | -2.64 | 0 | -2.64 | .96 | < .01 | -4.540 | -.736 |
| self_efficacy | -1.80 | 0 | -1.80 | .87 | .04 | -3.519 | -.072 |

[a] Reference category = 1

Level 3 vs. Level 1

| Contrast Estimate | Hypothesized Value | Difference (Estimate - Hypothesized) | SE | p | 95% C.I. for Difference Lower Bound | Upper Bound |
|---|---|---|---|---|---|---|
| -1.95 | 0 | -1.95 | 1.01 | .056 | -3.95 | .05 |
| -.72 | 0 | -.72 | .95 | .451 | -2.61 | 1.17 |
| -1.67 | 0 | -1.67 | .86 | .056 | -3.38 | .04 |

# 5. Discussion

This study sought to determine whether game-based learning could provide an embodied cognition foundation upon which formal learning of scientific reasoning principles could be enhanced. It further sought to investigate whether structured or unstructured access to this material would have a differential impact on learning, and whether the experience of playing this game would have an impact on students' science motivation. Both the game and the study itself were designed based on combining insights from the literature on transfer, game-based learning, embodied cognition, and scientific reasoning. The creation of a largely blank-slate protagonist for the player to project onto was inspired by the "learner hero". The choose-your-own-adventure format gives autonomy to the player as they control the protagonist, and a sense of competence as they advance through the puzzles. The protagonist's friends foster relatedness as well, with their enthusiasm or disappointment based on player answers. Providing an overarching goal—to solve the puzzles before the characters are "trapped forever" on game over—and the ability to meaningfully affect the chances of winning or losing based on player choices were intended to recreate elements identified as promoting a "flow" experience.

The study was also designed to avoid some of the pitfalls from previous research on game-based learning, by incorporating various insights from meta-analyses and Barnett and Ceci's (2002) transfer taxonomy. For instance, it attempts to control for transfer distance in terms of functional context by providing a tutorial within the game in a similar format to the eventual questions on the CVS tests, accounting for the potential advantage in control group students' awareness that they are learning a scientific principle. It also accounts explicitly for the impact of repeated training sessions and supplemental instructional supports, elements that were neglected in studies that claimed mixed or inconclusive results. Further, the narrative of the game is inextricable from the scientific principles to be taught, as opposed to irrelevant set dressing—a distinct weakness of a previous study claiming a disadvantage for a gaming condition over a matched Powerpoint slides condition. It also addresses a concern put forth by Young et al. (2012) about science games on the necessity for students to be prompted on the skills they learned from a game to utilize those skills in practice outside the game by explicitly drawing the connection to experiments in the tutorial section. Survey results and reflection comments speak to successfully promoting intrinsic motivation within the game. Over a third of the survey respondents in both game conditions noted finding the game enjoyably challenging, and a similar amount mentioned a deep curiosity about the plot and a strong drive to find out what was going to happen next as they played.

The CVS tests also provide interesting evidence on the impact of the game on scientific reasoning. The significant effect of group on the pretest scores is particularly surprising, as the students in each group are as similar as possible, with the entire sample from the same grade learning at the same school under the same teacher. It is conceivable, if improbable, for this pattern of baseline scores to have occurred by chance. Even if there was some manner of unobserved latent variable responsible for the group differences at baseline, however, the reversal of outcomes between groups from pretest to posttest creates an implication (albeit one that cannot be verified statistically) that the unstructured playing group improved over the course of the study period—perhaps enough so to overcome any unknown advantage the control group may have had. This notion is worth further examination, especially if it might mean that the utility of the game as an intervention has an interaction with student ability—in other words, students with the weakest scientific reasoning abilities at the outset benefit the most from the game.

What's especially interesting about the retention test results is that the expected learning loss after a delay is significantly greater in the structured game play group than the other two. One potential interpretation is the unstructured playing group took full advantage of their unrestricted access to the game and played it a great deal more than the structured group's three exposures to the game. The second group, with less overall exposure, would then experience learning loss faster. If so, it would imply not only that the game is helpful in promoting learning, but that students sought it out to play more times than would've been necessary simply to have solved all the puzzles, a suggestion supported by the significant effect of group observed in the survey data for the intrinsic motivation subscale. In future post-pandemic work, a longer study period with a mechanism to record gameplay time could draw out more subtle differences between the groups.

One notable limitation of the study design was the ceiling effects observed in the CVS tests. A few factors may explain this trend—the original instrument used by Chen & Klahr (1999) was used to evaluate CVS in slightly younger children, finding robust effects in 4th grade students. The present study works with middle-school children, when experimental design and scientific reasoning begins to be formally taught for the first time. It is possible that the observed skew in the data might be a result of some older children not finding the instrument challenging enough, or a latent variable such as SES or cognitive ability in the student population at the school. Teasing apart the impact of the individual study components may shed further light on the matter. With the PFL paradigm, this study investigates the impact of exposure to the game on a child's ability to benefit from formal instruction later. However, it would also be interesting to examine the direct impact of the game itself on learning prior to any formal instruction, as well as to perform comparisons between the use of the game alone, formal instruction alone, and the game combined with formal instruction.

Perhaps the most important potential future direction of this work would be its potential to be adapted for use in communities in which students have low engagement in science or may be at risk for school failure, especially low-income and minority communities, and for students with learning disabilities for whom conventional instruction methods are often insufficient, and for whom engagement with school and learning often suffers as a result. The current study had sample expected to be varied in its demographics to be as generalizable as possible, but future work can be targeted toward these more vulnerable demographics for the development of efficient and low-cost interventions along these lines.

## References

Barnett, S.M. and Ceci, S.J., 2002. When and where do we apply what we learn?: A taxonomy for far transfer. Psychological bulletin, 128(4), p.612.

Barsalou, L. W. (2008). Grounded cognition. Annu. Rev. Psychol., 59, 617-645.

Black, J.B., Khan, S.A.and Huang,S.C.D. (2014) Video games as grounding experiences for learning. In F. C. Blumberg (Ed.) Learning by playing: Frontiers of videogaming in education. New York: Oxford University Press

Bressler, D. M., & Bodzin, A. M. (2013). A mixed methods assessment of students' flow experiences during a mobile augmented reality science game. Journal of Computer Assisted Learning, 29(6), 505-517.

Chen Z, Klahr D. All other things being equal: Acquisition and transfer of the control of variables strategy. Child development. 1999 Sep;70(5):1098-120.

Csikszentmihalyi, M. (1975). Play and intrinsic rewards. Journal of Humanistic Psychology, 15, 41–63.

Han, I., & Black, J. B. (2011). Incorporating haptic feedback in simulation for learning physics. Computers & Education, 57(4), 2281-2290.

Holbert, N., & Wilensky, U. (2019). Designing educational video games to be objects-to-think-with. Journal of the Learning Sciences, 28(1), 32-72.

Klahr, D., Triona, L.M. and Williams, C., 2007. Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. Journal of Research in Science teaching, 44(1), pp.183-203.

Kuhn, D. (2005a). Education for thinking. Cambridge, MA: Harvard University Press.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. Monographs of the Society for Research in Child Development, 60(4, Serial No. 245), 1–128.

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. National Academies Press.

NGSS Lead States (2013). Next generation science standards: For states, by states. Washington, DC: The National Academies Press.

Rigby, C.S. and Przybylski, A.K., 2009. Virtual worlds and the learner hero: How today's video games can inform tomorrow's digital learning environments. Theory and Research in Education, 7(2), pp.214-223.

Ross, J. A. (1988a). Controlling variables: a meta-analysis of training studies. Review of Educational Research, 58(4), 405–437.

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review*, *39*, 37-63.

Young, M.F., Slota, S., Cutter, A.B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M. and Yukhymenko, M., 2012. Our princess is in another castle: A review of trends in serious gaming for education. Review of educational research, 82(1), pp.61-89.