CHAPTER 5

Signaling in Disciplinarily-Integrated Games: Challenges in Integrating Proven Cognitive Scaffolds Within Game Mechanics to Promote Representational Competence

Satyugjit S. Virk and Douglas B. Clark

Introduction

Interpreting, translating, and manipulating across formal representations is central to scientific practice and modeling (Pickering 1995; Lehrer and Schauble 2006a, b; Duschl et al. 2007). We have developed disciplinarily-integrated games (DIGs) such that players' actions involve the iterative development and manipulation of formal representations as the core game mechanics. These formal representations are computational and mathematized representations of focal science phenomena. Through playing a DIG, students investigate key conceptual relationships in the domain while also developing facility with the representations and inscriptions themselves.

S. S. Virk (✉)
Vanderbilt University, Nashville, TN, USA

D. B. Clark
University of Calgary, Calgary, AB, Canada

O. O. Adesope, A. G. Rud (eds.), *Contemporary Technologies in Education*, https://doi.org/10.1007/978-3-319-89680-9_5

Supporting students in engaging with these practices of manipulating and transforming across representations, however, is challenging. Madsen and colleagues have demonstrated the efficacy of signaling in helping students to concentrate on key relationships in diagrams (Madsen et al. 2012, 2013). Sweller (2006) and many others have demonstrated the efficacy of worked examples in both multimedia and educational games. Following those studies, the purpose of the current study is to (a) explore the potential efficacy of a DIG about Newtonian mechanics and (b) compare the relative contributions of a version of the game that incorporates into its design with a version of the game that does not.

DISCIPLINARILY-INTEGRATED GAMES

SURGE Symbolic (see Fig. 5.1) is the prototypical DIG template that we will consider in terms of generalizability to hypothetical DIGs for other disciplinary topics. More information and playable demos of *SURGE*

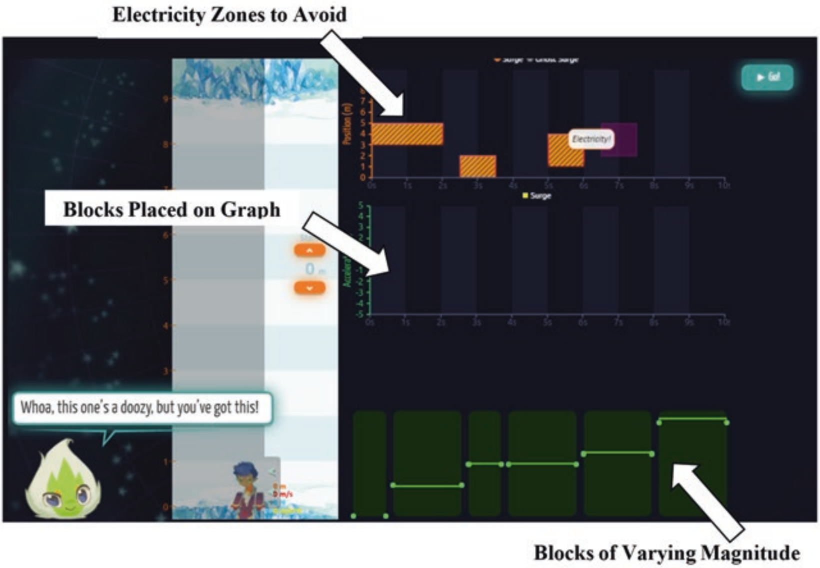


Fig. 5.1 Anatomy of an introductory block level. Blocks of varying magnitude of position, velocity, or acceleration must be placed in the correct order to create a path for SURGE to avoid the electricity zones and make it to the exit portal in the top graph

Symbolic and other *SURGE* games are available at www.surgeuniverse.com. *SURGE Symbolic* is a game that is the result of the evolution of design, research, and thinking chronicled in Clark et al. (2015, 2016d). Whereas earlier versions of *SURGE (*i.e., *SURGE Classic, SURGE Next,* and *SURGE: Fuzzy Chronicles)* focused on layering formal representations over informal representations, *SURGE Symbolic* inverts this order, layering informal representations over formal representations while organizing gameplay explicitly around navigating, translating, and coordinating across representations.

Earlier versions of *SURGE* supported reflection on the results of game play through formal representations as a means to support strategy-refinement, but the formal representations were not the medium through which players planned, implemented, and manipulated their game strategies. Earlier versions of *SURGE* provided vector representations, for example, to help students understand what was happening and how they might adjust their control strategy, but these formal representations only communicated information that a player might or might not use. The challenges and opportunities in a given game level, however, were communicated through the layout of elements in the game world, not in the formal representations. Similarly, the player's controls for executing a strategy were also independent of the formal representations. Thus, while attending to the formal representations might help a player succeed in a level, earlier *SURGE* games did not use formal representations as the medium through which challenges and opportunities were communicated to the player, nor did earlier *SURGE* games use diagrammatic formal representations as the medium of control.

As discussed in Clark et al. (2015, 2016c), *SURGE Symbolic* builds on research on teaching physics using simulations and motion sensors (e.g., Brasell 1987; diSessa et al. 1991; Mokros and Tinker 1987), research on constructing graphs based on assembling relevant "pieces" of trajectories of motion, and research on SimCalc (e.g., Hegedus and Roschelle 2013; Kaput 1992; Roschelle et al. 2010). In the work of Tinker and colleagues, students have often been provided graphs of position or velocity that they were asked to replicate using the controls of the system, which might involve a motion sensor. Similarly, students have been provided with a dot trace representation overlaid on their phenomenological view that they worked to interpret in terms of a graph (diSessa et al. 1991), and SimCalc pioneered in scaffolding students' integration and differentiation between and across Cartesian graphs of position, velocity, and acceleration over time by dynamically linking across representations (Kaput 1992; Hegedus and Roschelle 2013).

DIGs build on these bodies of research by pushing more deeply on approaches for leveraging formal representations as the means of communicating challenges to players, as well as leveraging abstract formal representations as the players' means of control within the game. Furthermore, we propose that DIGs generalize beyond time-series analyses and multiple representation systems involving Cartesian graphs of change over time (Clark et al. 2016a).

DIGs, by definition, use formal representations as the medium through which challenges and opportunities are communicated to the player (Communication Representations), and DIGs use formal representations as the medium through which the player implements strategies and exerts control over the game (Control Representations). Some DIGs might use the same representation for both control and communication, while other DIGs might use one or more formal representations for communication and one or more other representations for control. All DIGs include a phenomenological representation (which in traditional digital games would be the primary focus). Furthermore, all DIGs include an intermediate representation to support players in translating from the phenomenological representation to the formal representations and to constrain their interpretation of the formal representation. The goal in all DIGs involves interpreting, creating, modifying, and translating across these formal and phenomenological representations.

The template for *SURGE Symbolic*, for example, presents the phenomenological representation (which we refer to as "the world") on the left side of Fig. 5.1. The phenomenological representation portrays the heroine, Surge, on her hoverboard moving forward and backward along a game map. The formal Cartesian graphs on the right side are the communication and control representations. The position and velocity graphs in Fig. 5.1, for example, can present information about the specific regions of the game world that will be affected by dangerous electrical storms at given times, as well as information about locations and times where rewards or allies will rendezvous with Surge. As a result of this design approach, the Cartesian space emerges as a set of scientific instruments for the player by communicating data about the game world that are not available through other means. While Fig. 5.1 shows an example where the challenges and opportunities are communicated through the position graph and velocity graphs, any subset (or all) of the Cartesian graphs could serve this role. Simultaneously, the Cartesian graphs also play the role of an instrument panel or mission planner, offering fine-grained control over the movement of the Surge spacecraft. In Fig. 5.1, for example, the player can exert control by placing forces of various magnitudes and durations at different time points in the force graph. Alternatively, the player can exert control through the other graphs using the toggles to the right of the graphs. The author of a game level designates which graphs are visible to the player, which graphs are used for which purposes (communication or control), and what challenges and goals constitute the level.

Thus, all DIGs have the following characteristics: (a) formal representations for controlling the game, (b) formal representations for communicating challenges and opportunities, (c) a phenomenological representation presenting the phenomenon being modeled, (d) an intermediate aggregating representation, and (e) game mechanics and goals focused on engaging the player in interpreting, creating, modifying, and translating across these formal and phenomenological representations.

Signaling Theory

Effective use of visual stimuli is highly related to efficient learning (Litchfield and Ball 2011). Likewise signaling, the process of using cues to direct a learner's attention toward key events in a multimedia presentation, has shown to be an effective tool in scaffolding students' multimedia experiences. Signals can help learners to understand content presented in multimedia presentations (Mautone and Mayer 2001), select relevant information using fewer cognitive resources (Britton et al. 1982), recall relevant information and ignore irrelevant information (Mautone and Mayer), integrate information effectively in transfer problems (Loman and Mayer 1983), and focus on perceptually striking features (Lowe 1999).

More specific to this study, signals have the potential to greatly enhance physics learning environments. Madsen and colleagues found that participants who answered physics problems correctly spent more time looking at relevant areas of physics problem set diagrams, while novices spent more time looking at irrelevant areas (Madsen et al. 2012). In light of these findings, they studied cuing in physics problem- solving and found that learners that viewed selection and integration cues overlaid onto physics transfer problems spent less time looking at irrelevant areas, and more time looking at relevant, "expert" areas (Madsen et al. 2013). Furthermore, Rouinfar and colleagues found that short visual cues overlaid onto physics problems facilitated immediate problem-solving and ability to transfer problem-solving skills to novel problems and such cues applied over multiple problems causes learners to automatically extract similar features in new problems (Rouinfar et al. 2014).

Signals can be implemented in multiple forms to help learners perceive relationships among representations, including altering the luminance of objects in a display (e.g., De Koning et al. 2007), altering font style (e.g., Mautone and Mayer 2001), flashing elements (Craig et al. 2002; Jeung et al. 1997), and orienting gestures guiding learners to related elements (Lusk and Atkinson 2007). However, not all forms of signaling work equally well in all instructional contexts (Hegarty et al. 2003). Signals should be carefully designed in light of the intended function, the expertise of the learners, and the nature of the relationships highlighted. For example, maintaining consistency in labeling and color choice is an effective way of representing that objects are similar across different representations because learners can more easily perceive the relationships among them (Dufour-Janvier et al. 1987; Zhang 1996). Ainsworth (2006, 2014) advocated for the importance of matching the scale of the representation of information to the scale of the display of this same information, later generalizing this idea as a design consideration.

Research Questions

The current study explores the overall efficacy of our current approach to designing DIGs as well as the potential contributions of integrating signaling into our design of DIGs. For this study, students in the baseline condition played a version of SURGE Symbolic without signaling added. Students in the comparison condition played the same version of SURGE Symbolic with the addition of signaling functionality in a subset of the game levels. The signaling functionality adds flashing signals that visually link conceptual physics imagery to the corresponding symbolic representation in the graphical view of the game. These conditions were designed to investigate the following predictions:

1. Students in both versions of the DIG will demonstrate significant pretest-posttest learning gains.
2. Compared to students in the non-signaling condition, students in the signaling condition will demonstrate (a) increased pretest-posttest gains, (b) progress significantly further in the game, and (c) display patterns in their gameplay behavior indicating deeper conceptual sophistication.

Prediction #2 is informed by the logic that students who experienced signaling would have an enhanced understanding of how actions of the game character and relate to the corresponding graphical representations of position and velocity, resulting in a better understanding of these concepts and their connections. Arguing against prediction #2 is the possibility that increased complexity and load resulting from the addition of the signaling functionality might actually result in diminished rather than increased outcomes for students in the signaling condition. We have observed this tension in our prior work when attempting to integrate approaches to scaffolding, such as self-explanation and worked examples, into game play (Adams and Clark 2014; Adams et al. 2018). More specifically, we have found that integrating scaffolding from educational and psychological research into the context of digital games requires iterative research and learning environment design to leverage the affordances of the scaffold in a manner that does not compromise gameplay in terms of flow or complexity.

Methods

Participants

Sixty-nine seventh and eighth grade students from a diverse public middle school in Nashville of fairly high socioeconomic status participated in this study. Fourteen students were dropped from analyses because of attendance issues and/or missing the pretest/posttest. Students were randomly assigned to either the signaling or non-signaling condition. Pretests, posttests, and engagement surveys were administered to all students. A short cognitive task assessing attentional ability was also administered, but not analyzed in this paper. Interviews/screen recordings were conducted for a subset of students who provided consent forms from their parents and themselves.

SURGE Symbolic Game Design

As described in the background section on DIGs, SURGE Symbolic is a DIG designed to support student learning of Newtonian dynamics. Students play from the perspective of the space navigator, SURGE. Game play is divided into levels, each focused on a specific navigational challenge or Newtonian concept. Students must move SURGE forward or backward on her space board to find the appropriate position, velocity, or acceleration to navigate SURGE to the exit portals, represented by a purple box, while avoiding electricity zones, represented as orange boxes (see Figs. 5.2, 5.3, 5.4). SURGE's path is traced onto a graph representing the magnitude of position and velocity over time.

Levels contain increasingly more challenging combinations of Newtonian concepts as the game progresses. Initially, students are required to manipulate SURGE's position in the worldview to guide her to the purple exit portal. The distance that students move SURGE is represented directly by the graph of SURGE's movement as seen in Fig. 5.2. When a student successfully passes all position levels, they advance to velocity based levels. On velocity levels, students must change SURGE's physical location to represent a position-over-time rate that successfully avoids the electricity zones. The change in SURGE's position over time is graphed in worldview velocity

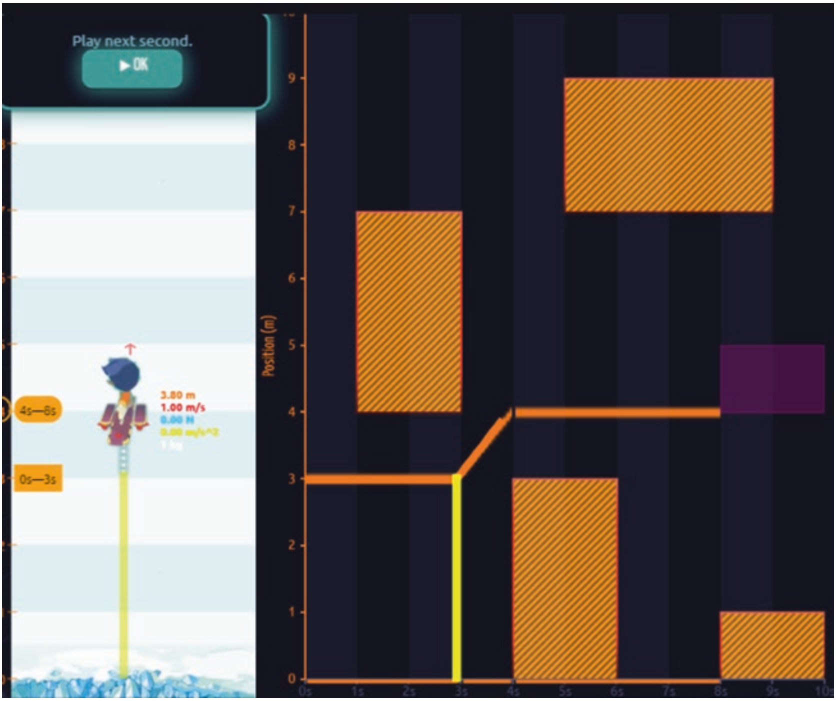


Fig. 5.2 Position WV level

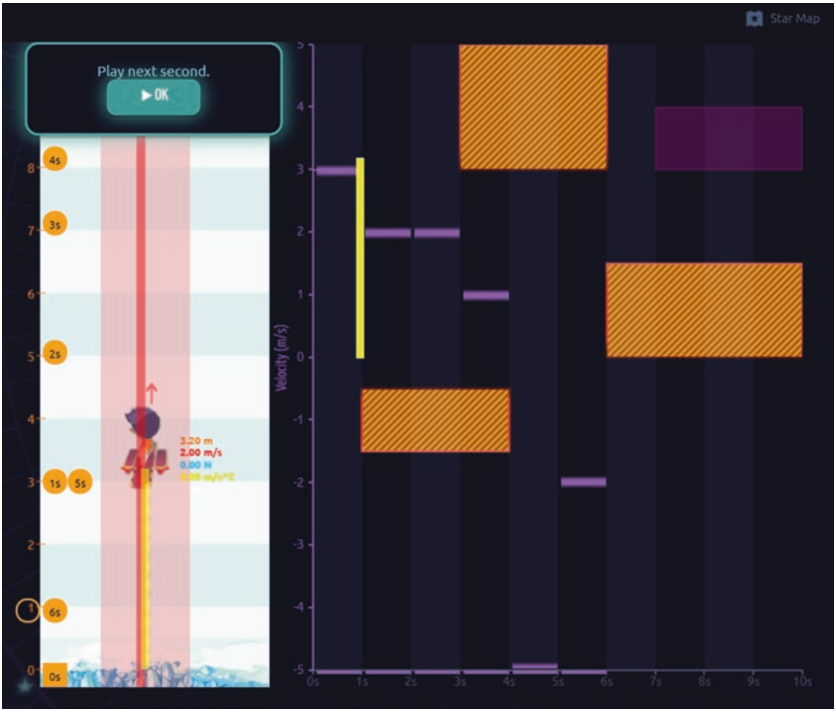
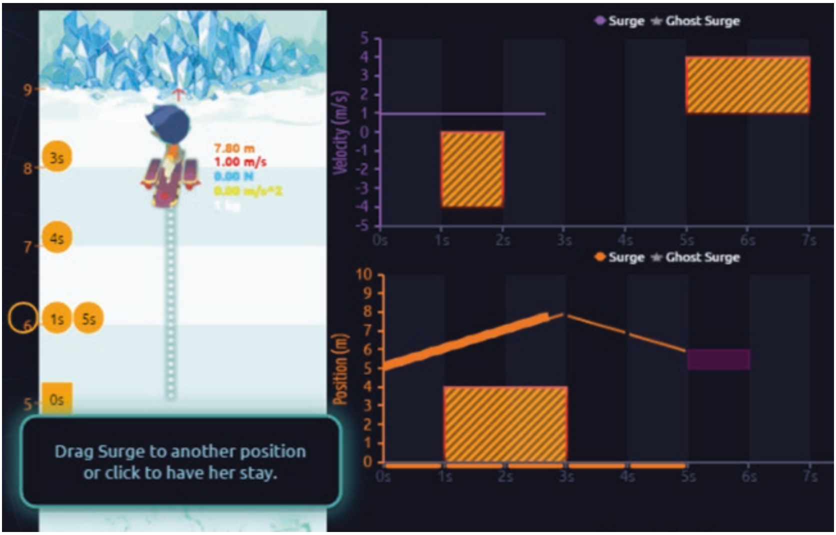


Fig. 5.3 Velocity WV level

levels (see Fig. 5.3). Students then progress to levels that require them to consider both position and velocity. On levels that combine position and velocity graphs, students are asked to manipulate SURGE's position to avoid electricity zones on the position graph as well as a worldview velocity graph (see Fig. 5.4). Finally, students apply concepts from position only, velocity only, and position/velocity combined levels to work through levels challenging their understanding of acceleration.

Depending on the level, students have two types of interfaces for setting up their strategies. In "world view" levels, the player drags the game character in the worldview to create the graph that will specify the position, velocity, or acceleration versus time for that level. In "block" levels, the player drags blocks that contain segments of a graph to create a graph to specify the position, velocity, or acceleration versus time. Block levels related to the preceding worldview topic are alternated with groups of



Drag Surge to another position or click to have her stay.

Fig. 5.4 Pos/Vel WV level

worldview levels to provide a different way to view the relationship between position, velocity, and time. Worldview levels serve as the introduction to each novel Newtonian concept. The rationale in terms of design was that the worldview interface would be more intuitive for students because students would drag the game character to the position on the map where they wanted the character to be at each time in the level. Then the block-level versions would require the player to think about what the graph should look like for the game character to be in the right place at each time in the level. We explain more about the worldview interface and the block-level interface as well as the nature of the signaling in the signaling condition in the following paragraphs.

Students have the flexibility and challenge to create any path that navigates SURGE safely through the electricity zones. Students use the toggle button to navigate SURGE and are able to test the success of their path by hitting the “Run” button. In worldview position levels, the signal in the signal condition represents the magnitude of distance SURGE is traveling during a set amount of time. The distance signaled in the game-worldview matches the distance graphed in worldview to facilitate student understanding of how the worldview graph is created. Unlike the worldview position level signal, the velocity-level signal highlights the rate at which SURGE is moving rather than the distance traveled. In worldview levels that display both position and velocity graphs, students are not given signals as the size of the graphs were smaller in these levels and no longer congruent to the height of the worldview. Hence, a signal between worldview and graphs would not make sense.

After three levels, students are able to deselect the second-by-second option of each run through so that each review is uninterrupted. It is important to note, however, that there was no tutorial showing students how to turn off this feature, so many students may have missed this option. This may have led to differences in game play results.

Students in the signaled condition are lead through worldview levels in an incremental method. They are first prompted to click and drag SURGE to a starting position and confirm that position by clicking the “OK” button in the dialogue box. This button has to be clicked before students can make the next move. Students have the option of moving SURGE to a new location to represent a change in position, or they can keep SURGE at the same value. After the students make their second move, they are again required to confirm the change or lack of change of location by clicking “OK.” Each movement is confirmed until a marked path is created for SURGE to follow to the exit portal. For each new movement, students have the option to choose the length of time SURGE needs in order to move the distance they set. This feature scaffolds student thinking toward understanding slope, or rate of movement.

When students have finished designing their path, they must click the “Run” button to evaluate the success of their plan. During the first execution of the plan, SURGE travels down the created path second-by-second and the students are signaled to progress to the next second after each move. Once students submit their first plan, the game plays the path from start to finish in real-time with no interruptions, with the signal present as well. If SURGE successfully reaches the exit portal, students have the choice to replay the same level or advance to the next. The signal itself is highly emphasized in the second-by-second run and the real-time presentation. The signaling button helps to breakdown each plotted movement to intentionally relate the worldview to the graph. In block levels (see Figs. 5.5, 5.6, 5.7), students are required to arrange blocks of various magnitudes in the correct order to navigate SURGE to the exit portal. Each block represents a rate that dictates how much SURGE moves on the position graph. When the blocks are placed next to each other on the

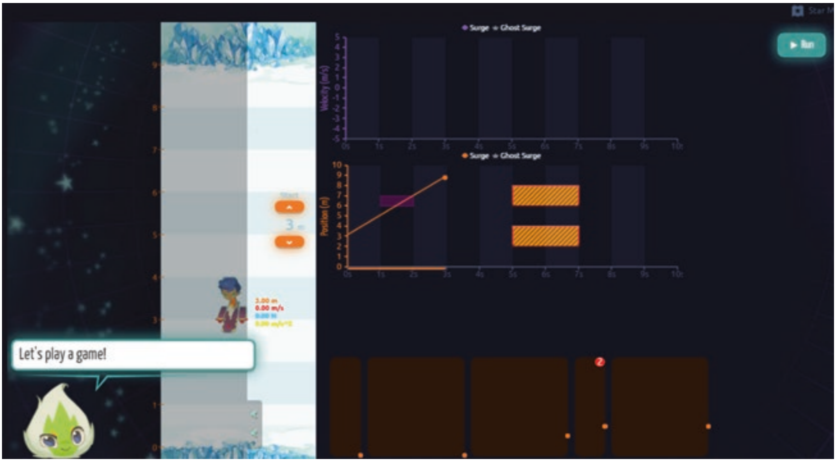


Fig. 5.5 Position block level

Fig. 5.6 Velocity block level

lower plane, they come together to create the graph of SURGE's movement. The student should now gain an understanding of the connection between real world movements and the production of a graph of movement throughout completion of worldview levels. The block levels reverse the representation, requiring students to analyze blocks of movement and

Fig. 5.7 Acceleration block level

connect them to individual movements of SURGE. Block levels also highlight the concept of slope more explicitly than worldview levels and they provide a different perspective for a student to approach Newtonian concepts. If a student can successfully navigate SURGE to the exit portal with the correct arrangement of blocks and without running into an electricity zone, the next level of the game is unlocked and the student can proceed. If the student is not successful, they can replay the level and change their initial actions. Therefore, players cannot skip ahead. Blocks can be dragged onto and off of the graph, as well as rearranged within the graph. The amount of total block movements and the individual types of block behavior were measured to evaluate how students interacted with block-level manipulatives.

Block levels are further categorized by the Newtonian concept they represented. Position block levels (see Fig. 5.5) represent a constant velocity between a start and an end point. Students organize position blocks onto a graph of SURGE's position to test whether or not their pattern of blocks navigate SURGE to the exit portal. Velocity block levels (see Fig. 5.6) requires students to place blocks signifying a constant velocity onto a graph of SURGE's velocity. Students have to avoid the electricity zones on both the velocity graph and the graph of SURGE's position over time. Acceleration block levels, represented in Fig. 5.7, require students to think about how the changing acceleration affects SURGE's position over time. This type of level was beyond the scope of expectations for students and was not tested directly.

Procedure

The study spanned six consecutive class periods over the course of a week. During the first session, all students took the pretest and completed a short cognitive task of attention that was not analyzed in this study. In all classes, students were instructed on how to navigate through initial steps in the game and received help as needed. Students were also encouraged to talk about the game, share strategies with their peers, and ask each other for help if they got stuck. Some students were interviewed about their thoughts and experiences in the game, and their screens were recorded under informed consent. Thirty minutes before the end of the last class on the final day, students were guided through the posttest and two short engagement surveys.

Assessment

A twenty-one question, multiple choice pre-posttest was created to assess physics understanding. The pretest and posttest were identical. Test questions assessed students' understanding of displacement and velocity graphically, mathematically, and verbally. Students often had to answer mathematical questions based on position and velocity graphs, or relate position and velocity concepts and/or graphs in order to answer questions correctly. Three questions assessed position-only concepts, two assessed velocity-only concepts, eleven questions assessed students' ability to link graphs or verbal descriptions of velocity and position together, five questions required students to link a graph of position to a verbal description of position or a graph of velocity to verbal description velocity, fourteen questions were modeled more after gameplay in the block levels, and seven questions were modeled more after gameplay in the worldview type levels. However, it should be noted that the block and worldview-based questions still utilized concepts that could be gleaned from either level type.

Students also completed two engagement surveys after they finished their posttest. The first was a standardized game experience survey (GEQ), which has demonstrated high reliability and validity. This survey had nineteen questions, with three possible responses ("yes," "no," "sort of") assessing students' positive affect, competence, immersion, flow, level of challenge, and other metrics of engagement. The second survey also used a three-item evaluation, assessing students' engagement in the SURGE Symbolic game both overall and for specific game features.

Results

Results are first presented within and across experimental conditions in terms of the pretest, posttest, engagement, overall gameplay metrics, and worldview level specific gameplay metrics.

Assessment Performance and Overall Gameplay Metrics

Pretest

A one-way ANOVA of pretest scores was conducted for the two conditions (see Table 5.1). As expected, there were no significant differences between the two conditions on the pretest $F(1,53) = .72$, $p = 0.40$).

Posttest

Next, a one-way ANOVA of posttest scores was conducted (Table 5.2). The one-way ANOVA showed that the difference in posttest scores between non-signaled and signaled conditions was significant, with the non-signaled condition outperforming the signaled participants with a fair effect size, $F(1,53) = 4.23$, $p = .05$, $d = 0.55$. Specifically, the section requiring students to link position and velocity representations showed significant differences across conditions, $F(1,53) = 4.53$, $p = .04$, $d = 0.58$, with the non-signaled group scoring higher than the signaled group. No other sections of the posttest were significant, however, the section that was modeled most after the worldview-based levels, had marginally significant differences, $F(1,53) = 3.67$, $p = .061$, where the non-signaled condition performed better than the signaled one.

To make sure that the significantly higher posttests in the non-signaled condition were not a factor of the non-significantly higher pretest in that condition, we also compared pre-post gains for both conditions. A paired

Table 5.1 Learning and engagement results

Performance metric	*Non-signaled condition*		*Signaled condition*		F	P	*Cohen's* d
	M	SD	M	SD			
Total pretest score	8.32	3.29	7.59	3.08	F(1,53) = .72	.40	
Total posttest score	9.64	4.23	7.52	3.38	F(1,53) = 4.23	.05	0.557
Position PTQ (posttest questions)	2.32	0.77	2.04	0.81	F(1,53) = 1.78	.19	
Velocity PTQ	0.68	0.77	0.74	0.0.66	F(1,53) = 0.10	.75	
Linking Pos/Vel PTQ	4.07	2.39	2.85	1.81	F(1,53) = 4.53	.04	0.58
Linking same concept PTQ	3.00	0.98	2.41	1.45	F(1,53) = 3.18	.08	
Block level PTQ	6.18	2.72	5.00	2.18	F(1,53) = 3.12	.08	
Worldview level PTQ	3.89	1.62	3.04	1.70	F(1,53) = 3.66	.06	
Total GEQ	10.5	8.04	11.78	7.84	F(1,53) = .36	.55	
Total game specific survey	9.39	5.53	9.44	5.49	F(1,53) = .00	.97	

Table 5.2 Pre/post gains across all students and by condition

Condition	*Pretest (SD)*	*Posttest (SD)*	*Gain score (SD)*	T	P	*Cohen's* d
All students	7.96(3.17)	8.60(3.95)	.64(0.77)	1.65	.11	N/A
Non-signaled	8.32(3.28)	9.64(4.22)	1.32(2.88)	2.43	.02	0.66
Signaled	7.59(3.08)	7.52(3.38)	−.074(2.73)	.14	.89	N/A

sample t-test was also performed to compare the average difference between pretest and posttest scores on various question types (see Table 5.2). The t-test was conducted to determine how gameplay affected students' understanding of conceptual questions. Means and standard deviations for the pretest, posttest, and gains scores can be found in Table 5.1. Overall and unexpectedly, there were no significant differences in learning across all students from pretest (*M* = 7.96, *SD* = 3.17) to posttest (*M* = 8.60, *SD* = 3.95); *t*(54) = 1.65, *p* = .11. There was a significant difference between pretest (*M* = 8.32, *SD* = 3.28) and posttest (*M* = 9.64, *SD* = 4.23) scores for participants in the non-signaled condition; *t*(27) = 2.43, *p* = .02, *d* = .66. Also unexpectedly, the posttest performance scores (*M* = 7.52, *SD* = .65) of students in the signaled condition did not show a significant difference to pretest scores (*M* = 7.59, *SD* = 3.08); *t*(26) = .14, *p* = .89. Therefore, game play for students in the non-signaled condition seemed to have a large impact on conceptual learning compared to students who experienced game play in the signaled condition.

Engagement

A one-way ANOVA of two engagement surveys across the two conditions found that there was no significant difference in engagement across the two conditions for either the GEQ survey, $F(1,53) = .36$, $p = .55$, or game-specific survey, $F(1,53) = .00$, $p = .97$.

Overall Game Behavior

Highest Level Completed

ANOVAs were conducted to compare the experimental conditions for progress in the game (see Table 5.3). Students could move onward to a subsequent game level only after successfully completing the game level preceding it. For this reason, the highest game level a student completed measured how far the student progressed in the game. A one-way ANOVA of the highest level completed among the two conditions found that there were no significant differences across the two conditions, $F(1,53) = 0.01$, $p = .91$.

Worldview Behaviors (Overall)

Three worldview behavior performance metrics were examined: (a) how many times students dragged SURGE within a worldview level trial, (b) the average time spent on each worldview trial, and (c) the average number of trials per worldview level. A one-way ANOVA of each of these metrics found significant differences across conditions for these behaviors with large effect sizes, (a) $F(1,53) = 15.87$, $p = .00$, $d = 1.07$ (b) $F(1,53) = 30.12$, $p = .00$, $d = 1.49$ (c) $F(1,53) = 103.31$, $p = .00$, $d = 2.72$. We explore the nature of the significant differences in a subsequent section below.

Block-Level Behaviors

Within block levels, we examined (a) the average number of times players moved blocks per trial, (b) the average number of times players dragged blocks out of play per trial, (c) the average number of times players rearranged already-placed blocks per trial, (d) the average number of times players moved blocks per trial, (e) the average number of times players moved SURGE per trial, (f) the average total number of block movements

Table 5.3 Overall gaming behavior by level type

	Non-signaled condition		*Signaled condition*		F	P	*Cohen's* d
	M	SD	M	SD			
Highest level completed	44.00	10.51	44.33	11.69	F(1,53) = 0.01	.91	
Average drags/ worldview level trial	6.97	1.35	8.62	1.72	F(1,53) = 15.87	.00	1.07
Average time/ worldview level trial	81.09	58.90	17.84	10.86	F(1,53) = 30.12	.00	1.49
Average trials/ worldview level	1.66	0.90	7.13	2.70	F(1,53) = 103.31	.00	−2.72
Average block moves to graph/trial	4.49	1.06	4.20	1.07	F(1,50) = .93	.34	
Average block drags out of play/trial	1.33	0.97	1.36	0.60	F(1,50) = .02	.90	
Average block rearrangements/ trial	7.53	4.80	6.45	3.08	F(1,50) = .94	.34	
Average total block moves/trial	13.34	5.27	12.00	4.15	F(1,50) = 1.04	.31	
Average SURGE moves/ trial	4.84	1.90	4.36	1.92	F(1,50) = 0.82	.37	
Averages total moves/trial	18.19	7.04	16.37	5.94	F(1,50) = 1.01	.32	
Average time/ trial	51.43	21.32	52.31	24.24	F(1,50) = 0.02	.89	
Average trials/block level	3.38	1.11	3.38	1.54	F(1,50) = .000	.99	

players made per trial, (g) the average time players spent per trial, and (h) the average number of trials they took to complete a block level. None of these metrics were significant, (a) $F(1,50) = .934$, $p = .339$, (b) $F(1,50) = .02$, $p = .90$, (c) $F(1,50) = .94$, $p = .34$, (d) $F(1,50) = 1.04$, $p = .31$, (e) $F(1,50) = 0.82$, $p = .37$, (f) $F(1,50) = 1.01$, $p = .32$, (g) $F(1,50) = 0.02$, $p = .89$, (h) $F(1,50) = .00$, $p = 98$.

Worldview Levels Dissected Game Behaviors

Worldview Position Level Behaviors

For worldview position levels, we examined (a) the average number of trials per unique position worldview level played, (b) the number of times players dragged SURGE per trial on these levels, (c) the average time players spent on worldview position levels per trial for these levels (see Table 5.4).

Table 5.4 Worldview sub-level gaming behavior

Performance metric (across all levels in category)	*Non-signaled condition*		*Signaled condition*		F	P	*Cohen's* d
	M	SD	M	SD			
Average trials/ position worldview level	2.16	1.60	20.37	14.67	F(1,54) = 42.67	.00	1.75
Average drags/ position worldview level	7.13	1.88	9.51	2.12	F(1,53) = 18.84	.00	1.19
Average time/ position worldview level	123.99	102.66	17.23	8.93	F(1,53) = 28.95	.00	1.47
Average trials/ velocity worldview level	1.01	1.17	12.76	40.86	F(1,52) = 2.32	.14	N/A
Average drags/ velocity worldview level	6.00	1.53	6.71	2.07	F(1,47) = 1.827	.18	N/A
Average time/ velocity worldview level	93.26	56.19	20.29	15.70	F(1,47) = 37.55	.00	1.77
Average trials/Pos/ Vel levels combined	10.00	4.84	16.85	19.76	F(1,54) = 16.27	.00	0.47
Average drags/Pos/ Vel levels combined	6.79	1.34	9.17	1.96	F(1,53) = 23.51	.00	1.42
Average time/Pos/ Vel levels combined	47.83	18.51	17.07	10.95	F(1,53) = 31.20	.00	2.02
Average trials/ Pos+Vel levels	1.75	1.59	25.04	17.24	F(1,35) = 12.70	.00	1.90
Average drags/ Pos+Vel levels	6.79	1.63	7.98	1.81	F(1,35) = 5.07	.03	0.69
Average time/ Pos+Vel levels	119.72	94.87	22.11	13.90	F(1,35) = 22.23	.00	1.44

A one-way ANOVA revealed significant differences across conditions for all of these metrics with large effect sizes. Here, the number average number of trials per level, *F*(1,54) = 42.67, *p* = .00, *d* = 1.75, and drags per trial, *F*(1,53) = 18.84, *p* = .00, *d* = 1.19, were significantly higher in the signaled group with high effect sizes. The average time spent per trial was significantly higher in the non-signaled group with a large effect size, *F*(1,53) = 28.95, *p* = .00, *d* = 1.47.

Worldview Velocity Level Behaviors

One-way ANOVAs were conducted to investigate the average number of drags and time spent per trial and the average number of trials per velocity worldview level. The average number of drags, *F*(1,47) = 1.83, *p* = 0.18, and trials per level, *F*(1,52) = 2.32, *p* = 0.14, were not significant between conditions, while the average time per trial was, *F*(1,47) = 37.55, *p* = .00, *d* = 1.77, where the non-signaled group spent significantly more time on average per trial than the signaled group.

Worldview Velocity and Position Levels Combined Behaviors

Combining all the signaled levels together, we see that all major metrics showed behaviors similar to the worldview position levels.

Specifically, a one-way ANOVA revealed significant differences across conditions for all of these metrics. Here, the number average number of trials per level, *F*(1,54) = 16.27, *p* = .00, *d* = 0.47, and drags per trial, *F*(1,53) = 23.51, *p* = .00, *d* = 1.42, were significantly higher in the signaled group with fair and high effect sizes. The average time spent per trial was significantly higher in the non-signaled group with a large effect size, *F*(1,53) = 31.20, *p* = .00, *d* = 2.02.

Position- and Velocity-Linked Worldview Levels Behaviors

We see an identical trend for levels where position and velocity levels were linked together (here no signal was used in either condition).

Specifically, a one-way ANOVA revealed significant differences across conditions for all of these metrics. Here, the number average number of trials per level, *F*(1,35) = 12.70, *p* = .00, *d* = 1.90, and drags per trial, *F*(1,35) = 5.07, *p* = .031, *d* = 0.69, were significantly higher in the signaled group with fair and high effect sizes. The average time spent per trial was significantly higher in the non-signaled group with a large effect size, *F*(1,35) = 22.23, *p* = .00, *d* = 1.44.

DISCUSSION

The pretest and engagement were not different among conditions. Accordingly, prior knowledge and ability, overall engagement with the condition, and knowledge gleaned from completed game levels apparently were not factors for the significant posttest score differences and pre-post gain differences. The posttest and pre-post gains demonstrated significant differences in favor of the non-signaling group, where this group performed better on the overall test and also significantly better on the assessment questions which required the student to link position and velocity concepts together and marginally better on the subset of questions modeled after worldview gameplay. Overall significant pre/post differences were not found across all students, but this is because only the non-signaled group demonstrated significant gains in learning from pretest to posttest.

Signaling should have fostered stronger links between the graph and worldview, fostering students' understanding of position and velocity. While the signals used did not link between graphs in this experiment (i.e., the signals only linked the worldview with a single graph), an increase in gains for position or velocity concept-based questions that did not require linking across multiple graphs should have been present. Similarly, an increase in gains for questions that required students to link verbal and graphical representations of position concepts together or velocity concepts together should have been observed, but in fact were not. Nor were indirect effects of potentially enhanced understanding of position and velocity observed, such as higher scores on the questions where students had to link these concepts. Since signaling should have enhanced student attention to the worldview level, we might also have expected an increase in the signaling group's learning. Instead, we found the opposite.

Clearly, the signaling in between the worldview and the position and velocity graphs hinder student understanding in some way. It is odd that something which should make learning more accessible and reduce the cognitive load of merging representations would detract from students' learning, especially in light of research concerning physics problems in multimedia learning, which strongly suggests that signaling should improve learning. There is also a possibility that, in addition to disrupting game cognition, the signal may have overscaffolded students and thus prevented them from actively making key connections between the worldview and the graph themselves.

What about the design or implementation of the signaling might have resulted in these unintended and undesirable outcomes? Notably, the initial few position and velocity worldview levels had an automatic replay that worked as follows: After the student hit the “go” button, each second of their game play was enacted on the views one second at a time. Students had to click the “play next second” button to keep moving forward. Once they reached the end of the level, the game played their set course from start to finish at the normal speed so the students could see what that looked like in real time. After the first few levels, students could optionally click off the second-by-second button so that the signal would disappear, although it is doubtful that many students in the signaling group used this functionality as there was no explicit tutorial in the game highlighting it. The initial second-by-second functionality was added to ensure that students had time to truly understand the relationship between worldview and graph, and realize the importance of the signal, before toggling it off.

What does the gameplay data suggest? The highest level completed was not significantly different across conditions. This suggests that differences were not due to signaled students being held up by the presence of the signal. Similarly, the number of block levels completed, trials, or any behavior involving blocks were not different among groups, so it seems that there was no kickback effect of better learning from the non-signaled worldview levels transferring to non-signaled groups’ understanding/cognition for the block levels. Looking at the worldview level data overall, we see that the signaled group had significantly more average trials per worldview level and also more average “mouse drags” of the game character per trial in setting up a plan for the level. The signaled group also spent significantly less time on average for each worldview trial. Hence, we see the group with the signaling required many more attempts to complete a level and utilized more actions to complete the level, but also spent far less time per trial. This set of behaviors could indicate students using more “brute force” to solve a level as rapidly as possible in the signaling condition, as opposed to thinking about each trial and utilizing only drag actions that seemed to make mathematical and scientific sense based on what they knew. This may indicate a lack of understanding of the physics concepts, resulting in frustration and subsequent inability to thoughtfully solve each level.

Accordingly, the signaling group scored lower on the overall assessment, including the subset of questions that require linking, and performed no better on questions that required them to match verbal and graphical representations of position or velocity together. This suggests a lack of understanding, potentially stemming from a deficiency in their responses to the signal and their gaming behavior. Additionally, repetitively and passively clicking each second by second at the end of each level may have encouraged more passive, autopilot cognition for solving levels.

Examining game behavior in the three types of worldview levels separately, we see that both the position and combined position and velocity worldview levels demonstrated a pattern of behavior similar to the overall worldview results. Specifically, the signaling students utilized significantly more drags and trials on average per level and significantly less time on average. This pattern holds true when we combine data for the position and velocity combined worldview levels, so all the levels with signals are aggregated. However, when we examine just the velocity worldview levels, we see no significant differences in drags or average trials per level. Yet, we still see significant differences in average time per trial, where the signaled group spent significantly less time on average per trial.

Accordingly, the game behaviors in the initial position worldview levels seem to be driving the overall results we see coupled with the position and velocity combined levels. It is interesting that the velocity worldview levels showed few differences in these metrics, perhaps because by the time students had gotten to these levels, they were more proficient in the game. Students in the signaled group still spent significantly less time on the levels on average, so some behaviors did persist, even in these levels. The reason for the differences in gaming behavior in the combined position and velocity levels observed later may be because these levels are much more advanced and more elicit and more aberrant behavior in gameplay due to deficiencies in learning. Interestingly, the combined position and velocity levels did not utilize signals in either condition. For this reason, the students were influenced to behave differently across conditions in these levels due to the presence or absence of signals in the prior position and velocity worldview levels.

Based on these findings, the signaling most likely disrupted understanding through the second-by-second playback at the start of the worldview levels, which might have never been turned off by students and could have easily disrupted and overloaded their cognition. Specifically, it may have disrupted them from holding the goal for the level in working memory, causing poorer understanding and gameplay. Even if slowing down the signal playback had some benefit for understanding what each point on the position graph meant, the overall learning benefit was compromised because of the disruption. Specifically, since students needed to go second-by-second to see the signal, it may have been too slow for their natural flow and could have disrupted their ability to examine incorrect answers at the end of a trial. This metacognitive examination would have helped students to think critically about the correct answer and what changes they would try out in a subsequent trial. This theory would help explain the inefficient game behaviors demonstrated across many worldview levels. The signaling may have also over scaffolded the students, preventing them from actively linking the worldview and graphs for themselves and instead promoting automatic, passive cognition and gameplay.

Possibly, then, it is the way the signal is specifically programmed and operated in the game levels. Our future research will compare groups in which the signal has no replay, optional replay, and mandatory replay with and without the second-by-second viewing mode. Monitoring differences between groups could help provide more evidence that mandatory playback design choices disrupt learning rather than signaling, itself. Likewise, comparing performances between a condition where the signal does not have the one second replay and a non-signaled condition may prove informative. Regardless, the findings of the current study are interesting and demonstrative of the power that early learning in games can have on understanding and success in later, more advanced concepts and levels.

However, that a change in the worldview foundational levels, especially position levels, would have such a striking effect on post test scores is, itself, interesting and demonstrative of the power that early learning in games can have on understanding and success in later, more advanced concepts and levels.

Conclusion

The Findings from the base condition of SURGE Symbolic without the signaling demonstrate significant pre-post gains on challenging physics and graphing concepts. In a prior study, we included a null condition with only a pretest and posttest but no intervention to determine whether a test/retest phenomenon could account for gains without an intervention (Martinez-Garza and Clark, submitted). That study demonstrated that gains on the test could not be attributed simply to a test/retest effect. We, therefore, interpret the significant pre-post gains in the base condition as demonstrating the efficacy of the overarching disciplinary integrated game approach enacted in SURGE Symbolic. Newtonian concepts as well as graphing concepts are very challenging for students, and often resistant to instruction (e.g. Hestenes et al. 1992; Hestenes and Halloun 1995). We are, therefore, pleased with the findings with regard to the base version of Surge Symbolic in terms of the overarching disciplinary integrated ideas that it represents.

The findings from the signaling condition, however, are disappointing and contrary to our predictions. Students in the signaling condition demonstrated no significant pre-post gains. While disappointing, we have encountered similar patterns in our prior research as we have attempted to integrate well-documented principles about scaffolding from psychology and cognitive science into digital games for learning. Our research has demonstrated that when worked examples come at the expense of time spent in gameplay, they can detract from game cognition and STEM learning (Adams et al. 2018). This research also found no benefit to worked examples embedded into game play for students with low prior knowledge. This is also consistent with Adams and Clark's (2014) findings, in which self-explanation prompts slowed students in the prompt condition such that they completed significantly fewer levels and scored significantly lower on a learning assessment. Looking across those studies and the current study, we see that the efficacy of implementing well-documented multimedia principles of learning in STEM games, here signaling, may not enhance learning if the design interferes with students' flow, cognitive load, or engagement with the game mechanics. In particular, results demonstrated that when signaling is overemphasized in a STEM game, it can disrupt or possibly overscaffold learners, resulting in detrimental learning gains and gaming behavior. More specifically, across all three of these studies investigating science learning games, we see that it is critical that scaffolds based on multimedia research (a) do not overscaffold the student or promote passive, automatic behaviors, (b) do not excessively detract from the student's gameplay time, and (c) do not disrupt game cognition and flow, especially the pace of flow. Tutorials are also imperative for all major features the student will encounter, such as the toggle off/on the button for the signal scaffold in this study.

This does not mean that these well-documented learning and scaffolding principles have no place in the design of digital games for learning. It simply means that refining designs that carefully integrate game mechanics and the design of the scaffolding can require careful iterative refinement. The meta-analysis of games for learning by Clark et al. (2016b) certainly highlights the key role of the specifics of design over simple binary comparisons of medium or approach. In the case of the self-explanation functionality, for example, building on the findings of the Adams and Clark (2014) study, we redesigned the self-explanation functionality to adaptively adjust to students' level of sophistication in terms of the abstraction of the prompts. We also adjust the timing and frequency of the prompts such that the prompts only appeared after the player had successfully completed a level. By timing the prompts in this fashion, the prompts were less intrusive and disruptive to players' gameplay and allowed more certainty that the explanation prompts would be appropriate to the player's current progress and solution. Our research on this refined approach to self-explanation demonstrated significant pre-post learning gains compared to a condition without the self-explanation functionality (Clark et al. 2016c). Similarly, we interpret these findings as implying the need to refine our approach to signaling within gameplay rather than implying that signaling is inappropriate for application in this setting. Essentially, we consider the findings a reminder of the complexity of integrating scaffolding that has been developed and other educational contexts into the context of digital games for learning.

Acknowledgments This material is based upon work supported by the National Science Foundation under Grant No. 1119290. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Adams, D. M., & Clark, D. B. (2014). Integrating self-explanation functionality into a complex game environment: Keeping gaming in motion. *Computers and Education, 73*, 149–159.

Adams, D. M., Clark, D. B., & Virk, S. S. (2018). Worked examples in physics games: Challenges in integrating proven cognitive scaffolds into game mechanics. In D. Cvetković (Ed.), *Simulations and gaming* (pp. 61–73). Rijeka: InTECH Open. https://doi.org/10.5772/intechopen.72152. http://www.intechopen.com/articles/show/title/worked-examples-in-physics-games-challenges-in-integratingproven-cognitive-scaffolds-into-game-mech

Ainsworth, S. E. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183–198.

Ainsworth, S. E. (2014). The multiple representation principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 464–486). New York: Cambridge University Press.

Brasell, H. (1987). The effect of real-time laboratory graphing on learning graphic representations of distance and velocity. *Journal of Research in Science Teaching, 24*, 385–395.

Britton, B. K., Glynn, S. M., Meyer, B. J. F., & Penland, M. J. (1982). Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology, 74*, 51–61.

Clark, D. B., Sengupta, P., Brady, C., Martinez-Garza, M., & Killingsworth, S. (2015). Disciplinary integration in digital games for science learning. *International STEM Education Journal, 2*(2), 1–21. https://doi.org/10.1186/s40594-014-0014-4. http://www.stemeducationjournal.com/content/pdf/s40594-014-0014-4.pdf

Clark, D. B., Sengupta, P., & Virk, S. S. (2016a). Disciplinarily-integrated games: Generalizing across domains and model types. In D. Russell & J. Laffey (Eds.), *Handbook of research on gaming trends in P-12 education* (pp. 178–194). Hershey: IGI Global. https://doi.org/10.4018/978-1-4666-9629-7.

Clark, D. B., Tanner-Smith, E., & Killingsworth, S. (2016b). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86*(1), 79–122. https://doi.org/10.3102/003465431558206 http://rer.sagepub.com/content/86/1/79.full.pdf+html

Clark, D. B., Virk, S. S., Barnes, J., & Adams, D. M. (2016c). Self-explanation and digital games: Adaptively increasing abstraction. *Computers & Education, 103,* 28–43. Online. https://doi.org/10.1016/j.compedu.2016.09.010.

Clark, D. B., Virk, S. S., Sengupta, P., Brady, C., Martinez-Garza, M., Krinks, K., Killingsworth, S., Kinnebrew, J., Biswas, G., Barnes, J., Minstrell, J., Nelson, B., Slack, K., & D'Angelo, C. M. (2016d). SURGE's evolution deeper into formal representations: The siren's call of popular game-play mechanics. *International Journal of Designs for Learning, 7*(1), 107–146. https://scholarworks.iu.edu/journals/index.php/ijdl/article/view/19359

Craig, S., Gholson, B., & Driscoll, D. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*(2), 428.

De Koning, B., Tabbers, H., Rikers, R., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology, 21*(6), 731–746.

diSessa, A., Hammer, D., Sherin, B., & Kolpakowski, T. (1991). Inventing graphing: Meta-representational expertise in children. *Journal of Mathematical Behavior, 10*, 117–160.

Dufour-Janvier, B., Bednarz, N., & Belanger, M. (1987). Pedagogical considerations concerning the problem of representation. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics.* Hillsdale: LEA.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.) (2007). *Taking science to school: Learning and teaching science in grades K-8.* National Research Council Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction, 21*(4), 209–249.

Hegedus, S., & Roschelle, J. (Eds.). (2013). *The SimCalc vision and contributions: Democratizing access to important mathematics.* New York: Springer.

Hestenes, D., & Halloun, I. (1995). Interpreting the FCI. *The Physics Teacher, 33*, 502–506.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*, 141–158.

Jeung, H., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17*(3), 329–345.

Kaput, J. (1992). Technology and mathematics education. In D. Grouws (Ed.), *Handbook on research in mathematics teaching and learning* (pp. 515–556). New York: Macmillan.

Lehrer, R., & Schauble, L. (2006a). Cultivating model-based reasoning in science education. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 371–388). Cambridge, UK: Cambridge University Press.

Lehrer, R., & Schauble, L. (2006b). Scientific thinking and science literacy: Supporting development in learning in contexts. In W. Damon, R. M. Lerner, K. A. Renninger, & I. E. Sigel (Eds.), *Handbook of child psychology* (Vol. 4, 6th ed.). Hoboken: John Wiley and Sons.

Litchfield, D., & Ball, L. J. (2011). Using another's gaze as an explicit aid to insight problem solving. *The Quarterly Journal of Experimental Psychology, 64*(4), 649–656.

Loman, N. L., & Mayer, R. E. (1983). Signaling techniques that increase the understandability of expository prose. *Journal of Educational Psychology, 75*, 402–412.

Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education, 14*(2), 225–244.

Lusk, M., & Atkinson, R. (2007). Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples? *Applied Cognitive Psychology, 21*(6), 747–764.

Madsen, A. M., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2012). Differences in visual attention between those who correctly and incorrectly answer physics problems. *Physical Review Special Topics – Physics Education Research, 8*(1), 010122.

Madsen, A., Rouinfar, A., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2013). Can short duration visual cues influence students' reasoning and eye movements in physics problems? *Physical Review Special Topics – Physics Education Research, 9*(2), 020104.

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology, 93*(2), 377.

Mokros, J. R., & Tinker, R. F. (1987). The impact of microcomputer-based labs on children's ability to interpret graphs. *Journal of Research in Science Teaching, 24*, 369–383. https://doi.org/10.1002/tea.3660240408.

Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago: University of Chicago Press.

Roschelle, J., Knudsen, J., & Hegedus, S. (2010). From new technological infrastructures to curricular activity systems: Advanced designs for teaching and learning. In M. J. Jacobson & P. Reimann (Eds.), *Designs for learning environments of the future: International perspectives from the learning sciences* (pp. 233–262). New York: Springer.

Rouinfar, A., Agra, E., Larson, A. M., Rebello, N. S., & Loschky, L. C. (2014). Linking attentional processes and conceptual problem solving: Visual cues facilitate the automaticity of extracting relevant information from diagrams. [Original research]. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01094.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction, 16*, 165–169. https://doi.org/10.1016/j.learninstruc.2006.02.005.

Zhang, J. J. (1996). A representational analysis of relational information displays. *International Journal of Human-Computer Studies, 45*(1), 59–74.