

Visual Signaling in a High-Search Virtual World-based Assessment: A SAVE Science Design Study

Brian C. Nelson¹ · Younsu Kim¹ · Kent Slack¹

Published online: 15 April 2016
© Springer Science+Business Media Dordrecht 2016

Abstract Education policy in the United States centers K-12 assessment efforts primarily on standardized tests. However, such tests may not provide an accurate and reliable representation of what students understand about the complexity of science. Research indicates that students tend to pass science tests, even if they do not understand the concepts being assessed. On standardized tests, such concepts are typically assessed via multiple-choice questions, which may check student receptive understanding of science-related vocabulary terms rather than their ability to develop hypotheses and design experiments to test those hypotheses. In an attempt to address these assessment issues, our SAVE Science project has been exploring the use of immersive virtual environments as platforms for both learning and assessment. SAVE Science (Situating Assessment in Virtual Environments for Science Content and Inquiry) is an NSF-funded study exploring the use of virtual world-based tests to assess the science knowledge and skill of middle school students. The main goal of SAVE Science is to explore the value of virtual world-based assessments as supplements or alternatives to more traditional forms of assessment. In pursuit of that goal, we are examining design frameworks designed to help students manage the high cognitive load they may experience while completing the tests. In this paper, we present results from a study exploring the use of visual signaling techniques in virtual world-based assessments, with a particular focus on their use and impact in visually complex, high visual search environments. The study focused on the use of visual signaling to reduce perceived student cognitive load, while simultaneously increasing the number of interactions students perform with assessment-relevant objects in a virtual world (assessment efficiency).

✉ Brian C. Nelson
brian.nelson@asu.edu

Younsus Kim
yskim1@asu.edu

Kent Slack
kjslack@asu.edu

¹ Arizona State University, Tempe, AZ, USA

Keywords Virtual worlds · Assessment · STEM education · Educational games · Instructional design

1 Introduction

Education policy in the United States centers K-12 assessment efforts primarily on standardized tests. However, such tests may not provide an accurate and reliable representation of what students understand about the complexity of science (Ketelhut et al. 2013; Songer et al. 2003). Research indicates that students tend to pass science tests, even if they do not understand the concepts being assessed (Michael 2007). On standardized tests, such concepts are typically assessed via multiple-choice questions, which may check student receptive understanding of science-related vocabulary terms such as “inquiry” rather than their ability to develop hypotheses and design experiments to test those hypotheses (NRC 2005).

In an attempt to address these assessment issues, researchers (including our team) have, over the past decade, been exploring the use of immersive virtual environments (IVEs) as platforms for both learning and assessment. These game-like digital spaces enable the situating of science practices and content in realistic scenarios. Such contextualized experiences have been shown to be engaging for students and beneficial for learning—particularly for students who do not do well with more traditional science instruction (e.g. Barab et al. 2005; Nelson and Ketelhut 2007). Using IVE-based science curriculum, real-world problems can be embedded in contextualized scenarios for students to solve while providing meaningful information on patterns of learning over time to both students and teachers (Ketelhut et al. 2012). Research indicates that using IVEs for learning and assessment offers a wealth of information about student’s knowledge and problem solving strategies in addition to assessing their solutions (Ketelhut 2007). For example, Shute et al. (2009) explore what they label “stealth” assessments embedded in virtual world-based games. Shute and her colleagues argue that player interactions and movements in a game world can be assessed in real-time using probability analysis techniques, without interrupting the flow of their experience. In virtual worlds and other immersive games, students can be continuously and invisibly assessed as they work through series of challenging tasks situated seamlessly into game play and narrative (Clark et al. 2009). The cumulative record of these interactions constitutes meaningful evidentiary trail of student understanding of the material they encounter in the virtual world. By following a systematic, theory-based approach to designing curricula and the activities of learning within those curricula, virtual worlds can produce data from students that demonstrate their evolving levels of competency around science concepts in a way difficult to achieve through multiple-choice questions alone (Nelson et al. 2011).

SAVE Science (Situating Assessment in Virtual Environments for Science Content and Inquiry) is an NSF-funded study exploring the use of virtual world-based tests to assess the science knowledge and skill of middle school students. We record and analyze all student actions in the game-like tests, letting us uncover patterns of scientific understanding revealed as students complete narrative-based quests. The main goal of *SAVE Science* is to explore the value of virtual world-based assessments as supplements or alternatives to more traditional forms of assessment. In pursuit of that goal, we are examining design frameworks structured to help students manage the high cognitive load they may

experience while completing the tests. By reducing the perceived complexity of the virtual worlds in which we place students, we hypothesize that they will more easily focus on the assessment activities themselves, rather than on extraneous elements such as interface and controls, leading to more accurate evidentiary data—i.e., a better test. To explore this hypothesis, we are conducting studies into the applicability of various multimedia principles in designing virtual worlds that have been shown to support learning in more traditional educational environments (Nelson et al. 2014).

In this paper, we present results from a study exploring the use of *visual signaling* techniques in virtual world-based assessments, with a particular focus on their use and impact in visually complex, high visual search environments. Visual signaling techniques in the SAVE Science assessments include adding 3D graphical arrows and colored highlights to the environment designed to direct student's attention to the salient elements in the virtual world (Fig. 1). The study explores whether the use of visual signaling to reduce perceived student cognitive load, while simultaneously increasing the number of interactions students perform with assessment-relevant objects in a virtual world (assessment efficiency).

The current study is a follow-up to one we conducted previously (Nelson et al. 2014). In the earlier study, 193 seventh grade students were randomly assigned to a virtual world-based assessment module called Sheep Trouble either with visual signaling or a version without signaling. That study found that students in the signaled version of the module, which used the signaling techniques shown in Fig. 1, reported significantly lower overall cognitive load, $F(1, 175.97) = 4.27, p = .04, d = .29$, as well as lower levels of cognitive load related to communicating with non-player characters, $F(1, 151.81) = 5.97, p = .02, d = .31$, navigating the virtual world, $F(1, 191) = 3.37, p = .07$, and finding objects in the module, $F(1, 176.94) = 2.92, p = .09$. For assessment efficiency, students in the signaled version of the virtual world interacted with more objects overall, and interacted with more sheep (the main interactive object type). Further, the signaling group took more

Fig. 1 Visual signaling techniques in SAVE science



measurements from the sheep (they could measure sheep weight, age, gender, and the length of various body parts), and recorded more information in an electronic in-world notepad.

In analyzing the earlier study, we realized the Sheep Trouble contained relatively few objects on the farm and was therefore visually simple. As such, students could easily identify objects to interact with without needing to search and select from a large number of non-interactive objects. We hypothesized that in a more visually complex environment, the impact of visual signaling on reducing cognitive load and raising object interaction rates should be more pronounced, given the increased challenges users would experience in locating interactive objects among a large number of non-interactive objects in a high-search environment.

To investigate this hypothesis in our current study, we examined these questions:

1. In terms of perceived student cognitive load, how does the use of visual signaling incorporated into a virtual world compare to a version not incorporating signaling? Does the impact of signaling differ between low visual search and high visual search versions of the virtual world?
2. In terms of assessment efficiency, how does the use of visual signaling incorporated into a virtual world compare to a version not incorporating signaling? Does the impact of signaling differ between low-search and high-search versions of the virtual world?

2 Theoretical Framework

2.1 Cognitive Load

Interacting with instructional materials in any settings causes learners to experience some level of cognitive load. In essence, cognitive load is considered the amount of mental effort associated with a learning task, such as an assessment, material, or lesson. Sweller et al. (1994) describe three types of cognitive load: intrinsic, extrinsic, and germane. Intrinsic load is the cognitive demand inherent in the task itself—the mental effort required to interact with and comprehend some body of material. Levels of intrinsic load vary as a function of the inherent difficulty of the material being studied. This in turn is related to the experience or expertise of the learner who encounters the material. So, for example, intrinsic load of conducting a scientific investigation will be high for a complete novice, but lower for students who have conducted such investigations previously. Meanwhile, extrinsic cognitive load is the mental effort imposed by extraneous or irrelevant information presented along with relevant material. Finally, germane cognitive load (Sweller et al. 1994) is associated with processing information, building mental models to understand information, and developing automation of skills. Germane cognitive load facilitates the achievement of an instructional goal by enhancing the processing of information or aiding in mental model construction. When the intrinsic load is high (because the material challenging to the learner) and the extrinsic load is reduced (through careful design), germane load can be increased. As germane load is increased, the learner has more “mental space” or resources to focus on the task at hand. A major goal then for designers of instructional and assessment materials is to reduce extrinsic load in support of germane load.

2.2 Visual Signaling

Visual signaling is one approach used to reduce extraneous cognitive load by directing the learner's attention to relevant information in instructional material (Morozov 2009). A learner's attention can be directed through the use of visual signals. For example, arrows can be added to instructional materials to indicate key material. Thus facilitating learners in their selection of the most relevant material. Wouters et al. (2008) put forth a set of guidelines to optimize learning and minimize extraneous cognitive load, suggesting the use of signals such as arrows to direct attention to important parts of instructional material.

A fairly large number of studies have been conducted examining the extent to which visual signals can reduce extraneous load and/or positively impact learning. For example, Richard Mayer (2010) found a consistent positive link between learning gains and the use of signaling in an overview of six empirical studies that used eye-tracking tools to understand how students process information in learning tasks. Chen and Fauzy (2008) found that visual signaling techniques including the use of directional arrows increased learner germane load, and translated to significant positive learning effects. In a study closer in focus to use of signaling in virtual worlds, de Koning et al. (2007) examined the impact of visual signals used to direct learner attention to key parts of animations. Using eye-tracking tools, they found that learners looked longer and more frequently at material that was signaled.

In the current study, we are particularly interested in the impact of visual signaling when used in virtual worlds with varying degrees of visually complex, and that therefore may require greater amounts of "visual search" by learners to locate important instructional or assessment materials. The level of difficulty (extraneous cognitive load) associated with the visual search itself is impacted by the number of objects in the learners field of vision, how closely positioned the scattered objects are, and the extent to which a learner needs to "move" to distinguish between objects. Environments that require high visual search may require learners to view and process many different visual objects, thus requiring more cognitive resources to select, organize, integrate, and process the information. Environments that require a low amount of visual search should be easier for learners to cognitively process, require less mental resources and thus impose less extraneous cognitive load (Nelson et al. 2014).

Looking specifically at the use of visual signaling in high-search environments, Jeung et al. (1997) found that adding visual signals to still images with high visual search was beneficial to learning; while use of similar visual signals added to low-search environments had little to no effect on learning. Conversely, de Koning et al. (2010) hypothesized that use of visual cues with an animation should reduce the amount of visual search and extraneous cognitive load. However, neither hypothesis was confirmed in their study with 90 high schools students viewing animations with and without signaling. A few researchers have investigated design approaches aimed at reducing extraneous cognitive load in virtual worlds and games (Lawrence 2006; Nelson and Ketelhut 2008; Erlandson et al. 2010). It is theorized that the higher levels of interactivity involved with games and virtual worlds may impact learners' cognitive load level and their ability to select, organize, and integrate information in the learning environment. The complexity of the concepts and the learning environment may be a deciding factor when including visual signals in a learning environment. Examining this idea, in two related studies, visual signals were added to an immersive physics education computer game to determine differences in perceived mental effort between students who received the visual signals and those who did not. No significant differences in perceived mental effort were found between those who received the

visual signaling and those who did not. One reason for this finding may have been the low amount of visual search required by the game (Erlandson et al. 2010; Nelson et al. 2010, 2011).

3 Methods

In the current study, conducted in spring 2014, we investigated visual signaling designed to reduce perceived student cognitive load in the Sheep Trouble virtual world, and explored its impact on assessment efficiency defined as object interaction rates (the higher the rate of object interaction, the greater the efficiency of the module). Further, we investigated whether the impact of visual signaling differed as a function of the visual complexity of the virtual world (high visual search vs. low visual search).

In Sheep Trouble, students investigate what is causing the ill-health of a flock of sheep recently imported to a country farm. As a science test, the goal of Sheep Trouble is to assess student understanding of concepts of species adaptation to a given physical environment. While interacting with computer-based human characters and two kinds of virtual sheep, students are asked to apply their classroom learning gained through traditional (book-based) lessons to complete a contextualized quest. In Sheep Trouble, students arrive on a virtual farm (see Fig. 2 below) where they meet a farmer who asks for help in finding out why his recently imported flock of sheep is in poor health. Students use a question and answer system to communicate with a farmer and his brother (see Fig. 6). They can also interact with flocks of new and local sheep wandering around a farmyard, using a set of interactive investigation tools. For example, students can measure the sheep's legs, body length, and ears with virtual rulers; can record and view their measurements of recent sheep weight loss or gain; and can view age and gender information. Once students feel they have gathered enough evidence, they explain their hypothesis to the sheep's owner. Behind the scenes, we record all student interactions and then analyze patterns in the data to understand how well students are able to collect, process, and apply their knowledge and skills to complete the quest.

For the current study, we added visual signals (glowing arrows) to interactive objects, primarily sheep and two human characters (Figs. 3, 5). Also, we designed high and low visual search versions of the module. In the high visual search version (Figs. 2, 4), we



Fig. 2 High visual search, no signaling



Fig. 3 High visual search, with signaling



Fig. 4 Low visual search, no signaling



Fig. 5 Low visual search, signaling



Fig. 6 Communicating with an NPC

placed a number of non-interactive objects around the virtual world, including buildings, trees, and additional farm animals (Fig. 6).

3.1 Data Sources

3.1.1 Audience

The study participants were 50 undergraduate students from three different computer science courses at a large public university in the southwestern United States. The participants were asked to join the study voluntarily, and they received extra credit in their respective computer science courses as an incentive for their participation. The participants' academic year ranged from freshman to senior. Among the participants, 45 students were males and 5 students were females. The participants were randomly placed in one of 4 conditions: Low-search with signaling: 11; Low-search with no signaling: 14; High-search with signaling: 9; High-search with no signaling: 16.

3.1.2 Procedure

The study lasted 1 week and each participant was asked to take part in a one-time 60-min session in a computer lab on campus sometime during the week. At the beginning of each session, the researchers gave a brief explanation of the purpose and the process of the study. Each study session consisted of two parts: the virtual world based assessment and a post assessment survey. For the virtual world-based assessment, the participants completed a PC desktop version of the Sheep Trouble module. The participants were given anonymous username and password to access and complete the assessment. Each participant was asked to spend at least 20 min working through the Sheep Trouble module. When a given participant concluded the Sheep Trouble module (by reporting their findings in the virtual world to the sheep's owner), a survey webpage was launched automatically with the post-implementation survey. Survey Monkey was used for the online survey, and participants used the same username to complete the survey as they had used in the virtual world-based assessment.

Table 1 Means and standard deviations for perceived cognitive load

	1 ^a	2 ^b	3 ^c	4 ^d	5 ^e	6 ^f	7 ^g	8 ^h	Combined
<i>Signaling</i>									
Low visual search	2.45 (1.21)	3.73 (1.62)	3.73 (1.35)	2.73 (1.56)	3.64 (1.43)	4.45 (2.12)	3.82 (2.36)	4.55 (1.97)	29.09 (7.91)
High visual search	2.67 (1.73)	2.78 (.97)	3.11 (1.36)	2.44 (1.13)	3.22 (1.99)	3.89 (2.76)	3.11 (2.42)	6.00 (2.24)	27.22 (9.04)
<i>No signaling</i>									
Low visual search	4.07 (1.94)	4.79 (2.61)	4.79 (2.19)	3.14 (2.35)	4.00 (2.48)	3.93 (2.30)	3.29 (2.37)	5.64 (2.41)	33.64 (13.82)
High visual search	2.75 (1.65)	3.94 (2.05)	3.69 (2.73)	3.19 (2.14)	3.06 (2.08)	3.56 (2.22)	3.13 (2.55)	4.94 (2.24)	28.25 (11.53)

Standard deviations shown in parentheses

- ^a How difficult was it for you to understand the content in Scientopolis?
- ^b How difficult was it for you to answer questions in Scientopolis?
- ^c How difficult was it for you to work with the Scientopolis environment (e.g. getting information, exploring information, using tools, finding things, languages, etc.)?
- ^d How difficult was it for you to collect all the information that you needed in Scientopolis?
- ^e How difficult was it for you to analyze the information that you needed in Scientopolis?
- ^f How difficult was it for you to make inferences about the information you collected in Scientopolis?
- ^g How difficult was it for you to compare information from different things you found in Scientopolis?
- ^h How hard did you concentrate during this assessment (to figure out what the problem was)?

3.2 Instruments

3.2.1 Cognitive Load Survey

Subjective rating scales were used to measure overall cognitive load. Subjective rating scales are the most frequently used measure to indicate cognitive load in design studies (Sweller et al. 2011). The participants were asked 8 self-report cognitive load items related to perceived cognitive load. Each item was a 10 points Likert-style questions based on those used in previous studies (Cierniak et al. 2009; Gerjets et al. 2009). The survey questions included items rating the level of overall difficulty experienced in the virtual world-based assessment, including such items as interacting with the virtual environment (e.g. “How difficult was it for you to work with the Scientopolis environment (e.g. using tools, finding things, etc.)?”), understanding the content (e.g. “How difficult was it for you to understand the content in Scientopolis?”), and concentration during the assessment (e.g. “How hard did you concentrate during this assessment (to figure out what the problem was)?”).

3.2.2 Assessment Efficiency

We defined efficiency as the number of interactions a given participant had with assessment-related objects. We measured: (1) total “collisions”: the number of times a student walks up to an in-world object to interact with it; (2) sheep collisions; (3) collisions with NPCs or other objects; (4) number of measurements taken while interacting with sheep; and (5) number of records recorded in an electronic notebook.

4 Results

A 2×2 analysis of variance (ANOVA) was conducted to assess the effect of visual signaling and visual search on perceived cognitive load and on assessment efficiency. The signaling treatment factor included two levels: visual signaling and no visual signaling. The visual search factor included two levels: high visual search and low visual search.

4.1 Cognitive Load Survey

For combined perceived cognitive load, ANOVA indicated no significant interaction between signaling and visual search, $F(1, 46) = .30, p = .59$, partial $\eta^2 = .01$, no significant main effects for signaling, $F(1, 46) = .74, p = .39$, partial $\eta^2 = .02$, and for visual search, $F(1, 46) = 1.26, p = .27$, partial $\eta^2 = .03$. For individual aspects of cognitive load, non-significant interaction between signaling and visual search, and non-significant main effects for signaling and visual search were found. The means and standard deviations for individual aspects and combined cognitive loads are presented in Table 1.

4.2 Assessment Efficiency

For assessment efficiency, ANOVA indicated no significant main effect for signaling treatment and no significant interaction between signaling treatment and visual search:

Table 2 Means and standard deviations of assessment efficiency

	Total collision	Sheep collision	Collision with NPC or other object	Number of measurements	Records in clipboard
<i>Signaling</i>					
Low visual search	29.82 (12.06)	16.64 (4.50)	7.36 (3.53)	81.45 (29.22)	69.18 (32.76)
High visual search	58.67 (23.10)	28.00 (14.99)	9.67 (4.80)	90.22 (16.28)	80.00 (28.71)
<i>No signaling</i>					
Low visual search	37.00 (11.73)	19.14 (10.80)	7.93 (2.87)	81.64 (27.61)	59.50 (36.79)
High visual search	64.62 (29.36)	22.63 (11.28)	9.75 (4.12)	103.75 (42.67)	92.25 (43.11)

Standard deviations shown in parentheses

Total collisions, $F(1, 46) = 1.16$, $p = .29$, partial $\eta^2 = .03$ for signaling and $F(1, 46) = .01$, $p = .92$, partial $\eta^2 < .001$ for interaction; sheep collisions, $F(1, 46) = .21$, $p = .65$, partial $\eta^2 = .004$ for signaling and $F(1, 46) = 1.57$, $p = .22$, partial $\eta^2 = .03$ for interaction; collisions with NPCs or other objects, $F(1, 46) = .09$, $p = .77$, partial $\eta^2 = .002$ for signaling and $F(1, 46) = .05$, $p = .83$, partial $\eta^2 = .001$ for interaction; Total number of measurements of sheep taken, $F(1, 46) = .54$, $p = .47$, partial $\eta^2 = .01$ for signaling and $F(1, 46) = .51$, $p = .48$, partial $\eta^2 = .01$ for interaction; and Total number of records entered into an electronic clipboard, $F(1, 46) = .01$, $p = .91$, partial $\eta^2 < .001$ and $F(1, 46) = 1.05$, $p = .31$, partial $\eta^2 = .02$ for interaction. The mean values on the assessment efficiency for signaling group were not statistically different to those for non-signaling group. Also, differences in the means between high-search and low-search did not vary as a function of signaling treatment. The means and standard deviations for assessment efficiency are presented in Table 2.

However, ANOVA indicated significant main effects for visual search in the following 3 assessment efficiency measurements: Total collisions with all possible in-world objects, $F(1, 46) = 21.36$, $p < .01$, partial $\eta^2 = .32$, Total collisions with sheep, $F(1, 46) = 5.56$, $p < .05$, partial $\eta^2 = .11$, Total number of records entered into an electronic clipboard, $F(1, 46) = 4.14$, $p < .05$, partial $\eta^2 = .08$. The visual search main effect indicated that the participants in the High-search environment tended to collide with more objects, to interact with more sheep, and to record more information into an in-world notebook than those in Low-search environment.

5 Significance of the Study

This study found no evidence that use of visual signaling in a virtual world reduces students' perceived cognitive load. These findings do not replicate those from our previous study, which found significantly lower overall levels of cognitive load related to the use of signaling, as well as lower levels of cognitive load related to finding interactive objects and navigating the virtual world. What accounts for the differences between the two studies? There are two likely explanations. The most likely source of the differences in findings is

that visual signaling may have only a small impact on cognitive load levels when used in a virtual world. If signaling has only a small effect on cognitive load, then that effect will only be uncovered in studies with larger numbers of participants. The current study had only 50 participants across for conditions, compared to nearly 200 participants in two conditions in the earlier study that showed significant results. Second, the differences in audience demographics may have resulted in differing levels of cognitive load. The prior study involved large numbers of middle school students, while the current study involved university undergraduate students, all of whom were computer science students who may be more proficient with computer systems, games, and virtual environments. This additional knowledge, experience, and cognitive development may have made it easier for the undergraduate students to interact with the Sheep Trouble module, reducing their extraneous cognitive load.

In a similar contradiction with the prior study which saw strong differences in assessment efficiency between signaled and non-signaled versions of the virtual world, no such difference was seen between groups in the current study. No differences were seen in the total number of collisions with interactive objects, nor with specific categories of interactive objects (sheep, non-player characters, posters). Also, no differences were seen in the number of measurements taken when interacting with the sheep, nor in the numbers of records entered in the provided electronic notebook. High visual search participants (with and without visual signaling) did interact with significantly more objects overall, but that is likely related to the fact that the high-search virtual world had more objects with which to collide.

The lack of significant findings related to assessment efficiency in terms of object interaction rates could relate to: (1) small sample size: The current study had fewer participants than the previous study, so effects would be difficult to detect; (2) assessment content: the assessment in Sheep Trouble was originally designed for middle school science students who recently completed a science unit in the content area. The content being assessed may have been too simple for the university students who took part in the current study, and thus may not have imposed sufficient cognitive load that could be diminished via visual signaling; (3) participant demographics: in a similar vein, it is likely that the older, more tech-savvy computer science student participants in the current study were more experienced with virtual world-based games than the middle school students from the previous study. If so, the lack of visual signals may not have increased participant cognitive load or hampered their efforts to interact with objects in the virtual world. In addition, the gender distribution in the earlier study was roughly 50/50 male/female. In the current study, 90 % of participants were males. It will be important to conduct a follow-up study with larger numbers of participants, and a participant pool that matches the earlier study.

Ultimately, the current study into the use of visual signaling with virtual world-based assessments has not echoed the relatively strong findings into the value of signaling seen in our previous study, nor the findings related to the value of signaling used with more traditional computer-based instructional environments. In addition to running much larger-scale studies, it may be fruitful for virtual worlds researchers to explore whether there is something in the experience of learning or completing assessment activities in a virtual world that sets it apart from other forms of digital material. Example factors to consider in such explorations include different users' demographic information and their pre-existing knowledge of the content and the digital materials. From this study, we conjectured that learner characteristics such as age, gender, and/or pre-existing knowledge may have influenced the results differently compared to the prior study. It will be important to

consider these types of factors in future studies to more accurately uncover and explain student experience in virtual world-based assessments. If visual signaling is not needed to help support learners in dealing with information in a virtual world, even a fairly visually complex one, an intriguing, and unanswered question, is why not?

References

- Barab, S., Arici, A., & Jackson, C. (2005). Eat your vegetables and do your homework: A design based investigation of enjoyment and meaning in learning. *Educational Technology*, 45(1), 15–20.
- Chen, C. J., & Fauzy, W. M. W. I. (2008). Guiding exploration through three-dimensional virtual environments: A cognitive load reduction approach. *Journal of Interactive Learning Research*, 19(4), 579–596.
- Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior*, 25, 315–324.
- Clark, D., Nelson, B., Sengupta, P., & D'Angelo, C. (2009). Rethinking science learning through digital games and simulations: Genres, examples, and evidence. In *Learning science: Computer games, simulations, and education workshop*. Washington, DC: National Academy of Sciences.
- de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology*, 21(6), 731–746. doi:10.1002/acp.1346.
- de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, 20(2), 111–122.
- Erlandson, B., Nelson, B., & Savenye, W. (2010). Collaboration modality, cognitive load, and science inquiry learning in virtual inquiry environments. *Educational Technology Research and Development*, 58(6):693–710.
- Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. S. (2009). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior*, 25, 360–370.
- Jeung, H. J., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology*, 17(3), 329–345.
- Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in river city, a multi-user virtual environment. *Journal of Science Education and Technology*, 16(1), 99–111.
- Ketelhut, D. J., Nelson, B., Schifter, C., & Kim, Y. (2013). Improving science assessments by situating them in a virtual environment. *Education Sciences*, 3(2), 172–192.
- Ketelhut, D. J., Yates, A., Sil, A., & Timms, M. (2012). Applying educational data mining in E-learning environments. *Section within the new measurement paradigm report* (pp. 47–52). http://cadrek12.org/sites/default/files/NMP%20Report%20041412_0.pdf
- Lawrence, C. (2006). Take a load off: Cognitive considerations for game design. *Proceedings of the 3rd Australasian conference on interactive entertainment* (pp. 91–95).
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20, 167–171.
- Michael, J. (2007). Conceptual assessment in the biological sciences: A National Science Foundation sponsored workshop. *Advances in Physiological Education*, 31, 389–391.
- Morozov, A. (2009). The effects of spatial visualization ability and graphical navigational aids on cognitive load and learning from web-based instruction. *Educational Multimedia and Hypermedia*, 18(1), 27–70.
- National Research Council. (2005). *America's lab report: Investigations in high school science*. Washington, DC: National Academies Press.
- Nelson, B. (2007). Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. *The Journal of Science Education and Technology*, 16(1):83–97.
- Nelson, B. & Ketelhut, D. J. (2007). Scientific Inquiry in Educational Multi-User Virtual Environments. *Educational Psychology Review*, 19(3):265–283.
- Nelson, B. & Ketelhut, D. J. (2008). Exploring embedded guidance and self-efficacy in educational multi-user virtual environments. *International Journal of Computer-Supported Collaborative Learning*, 3:413–427.

- Nelson, B., Ketelhut, D. J., & Schifter, C. (2010). Exploring cognitive load in immersive educational games: The SAVE Science project. *International Journal of Gaming and Computer-Mediated Simulations*, 2(1):31–39.
- Nelson, B., Erlandson, B., & Denham, A. (2011). Global channels for learning and assessment in complex game environments. *British Journal of Educational Technology*, 42(1):88–100.
- Nelson, B., Kim, Y., Foshee, C., & Slack, K. (2014). Visual signaling in virtual world-based assessments: The SAVE Science project. *Information Sciences*, 264:32–40.
- Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). Mahwah, NJ: Routledge, Taylor and Francis.
- Songer, N. B., Lee, H. S., & McDonald, S. (2003). Research towards an expanded understanding of inquiry science beyond one idealized standard. *Science Education*, 87, 490–516.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Measuring cognitive load, cognitive load theory* (pp. 71–85). New York: Springer.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1994). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.
- Wouters, P., Paas, F., & Van Merriënboer, J. J. G. (2008). How to optimize learning from animated models: a review of guidelines based on cognitive load. *Review of Educational Research*, 78(3), 645–675.