Guiding Attention on Physics Problems Using Visual Cues Modeled After Experts' Eye Movements

Adrian Madsen, Adam Larson, Amy Rouinfar, Allison Coy, Lester Loschky, and N. Sanjay Rebello

Kansas State University

Abstract

To maximize learning one must ensure most of the learner's cognitive resources are spent on relevant tasks and avoid instructional environments that facilitate focusing on the irrelevant. To help novice learners focus on elements relevant for learning it may be helpful to give them insight into the way experts allocate their visual attention, for example, by using visual cues. To design appropriate cues, we must first understand how those who solve problems correctly allocate their visual attention by recording their eye movements. In Study 1, we record eye movements of introductory and graduate physics students while answering conceptual physics problems containing a diagram to determine differences in visual attention. We use the eye movements of those who answer these questions correctly to design visual cues for Study 2. In Study 2, we overlay these dynamic visual cues onto the same physics problems and find evidence of increased conceptual understanding in novices who view the problems overlaid with cues.

Guiding Attention on Physics Problems Using Visual Cues Modeled After Experts' Eye Movements

Introduction

To maximize learning one must ensure most of the learner's cognitive resources are spent on relevant tasks and avoid instructional environments that facilitate focusing on the irrelevant. Mayer's cognitive theory of multimedia learning (2001) explains that learning occurs when relevant information is successfully selected and organized into a coherent representation and integrated into the existing knowledge base. All of these processes occur in one's working memory. Often learners are faced with learning environments that impose a high cognitive load and max out the limited capacity of working memory. To help alleviate this problem, visual cues can be used. de Koning, Tabers, Rikers and Paas (2009) has devised a framework to describe three specific functions of cueing which include guiding learners' attention to essential information, emphasizing organization, and making the relations between elements more salient to foster their integration. There exists a large body of research which utilizes visual cues in a variety of contexts to increase learning in animations and static problems (de Koning, Tabers, Rikers & Paas, 2007, Grant & Spivey, 2003, Kriz & Hegarty, 2007, Mautone & Mayer, 2001, Ozcelik, Arslan-Ari & Cagiltay, 2007).

The first function of cueing is especially interesting to us as it works to ensure that cognitive resources are spent appropriately by helping the learner to focus primarily on relevant elements. This frees up working memory and allows for real learning to occur. In physics classes, students are often faced with diagrams or animations of the real world which contain elements relevant to the task at hand as well as elements which are present in the real world, but not useful for answering the given question. To help students select the relevant and ignore the irrelevant, we study the use of dynamic visual cues overlaid on static physics problems containing a diagram. But before we can design cues to guide learners' attention, we must first understand where learners who answer questions incorrectly and correctly look when viewing physics problems containing diagrams. We need this information to determine what information in the diagrams is relevant for the correct answer and what should be ignored.

Research in many disciplines has used eye-tracking technology to investigate the differences in the visual attention of experts and novices. It has been observed that experts in a domain attend to task-relevant portions of a diagram more than novices in that domain. Jarodzka, Scheiter, Gerjets, & van Gog (2009) studied the visual attention of both novices and experts when viewing videos of unfamiliar fish swimming and classifying the type of locomotion. The authors found that experts spent significantly more time fixating on relevant areas of the video than biology students, who had the necessary background knowledge for differentiating types of locomotion but little practice in this classification task. The authors also found that novices spent more time than experts fixating on areas irrelevant for determining locomotion. Similar studies have measured eye movements of experts when viewing art (Antes & Kristjanson, 1991) and playing chess (Charness, Reingold, Pomplun, & Stampe, 2001), and have shown that the increased domain knowledge in these fields affects where people fixate while performing domain-relevant visual tasks. Based on these studies and other previous research, we conclude that novice and expert learners, who differ in their domain specific knowledge, allocate visual attention differently. Experts' knowledge drives them to attend to thematically relevant areas, or those portions of the diagram relevant to the task at hand, while novices lack of knowledge leaves them to view irrelevant areas of a diagram.

Here we report on two different studies. Study 1 investigates the differences in eye movements of those who answer physics problems correctly and incorrectly, where the critical information needed to answer the problem is contained in a diagram. Study 2 uses the results of Study 1 to design visual cues based on eye movements of those who answered the questions correctly. These visual cues are overlaid on the physics problems and shown to introductory physics students. We compare the eye movements and reasoning of students who saw visual cues and those who did not.

Study 1: Comparison of Eye Movements of Correct and Incorrect Solvers

Method

There were 24 participants in the study (3 females) with two different levels of experience in physics. Ten participants were first-year through fifth-year PhD students in physics who had taught an introductory physics course or been a teaching assistant for an introductory physics lab. One participant was a postdoctoral candidate in physics who had received his PhD within the last two years and had teaching experience. Thirteen participants were enrolled in an introductory psychology course and had taken at least one physics course in high school, though some had also taken an introductory physics course at the university level. The PhD students and post-doc participated as volunteers and the psychology students received course credit for their participation. As we were looking to compare those who answered the physics problems correctly to those who answered incorrectly, we selected participants with a broad range of experience. We expected that the PhD students would answer correctly, while the psychology students might answer incorrectly, though we knew that this might not always be the case since there is a wide distribution of expertise among introductory physics students and physics graduate students (Mason & Singh, 2011). The participants viewed 10 multiple-choice introductory physics problems. Participants were presented with physics problems on a computer screen viewed at a distance of 24 inches using a chin and forehead rest to minimize participants' extraneous head movements. The resolution of the computer screen was set to 1024 x 768 pixels with a refresh rate of 85 Hz. Each physics problem subtended 33.3° x 25.5° of visual angle. Eye movements were recorded with an EyeLink 1000 desktop mounted eye-tracking system (http://www.sr-research.com), which had an accuracy of less than 0.50° of visual angle. An eye movement was classified as a *saccade* (i.e., in motion) if the eye's acceleration exceeded 8,500°/s^2 and the velocity exceeded 30°/s. Otherwise, the eye was considered to be in a *fixation* (i.e., stationary at a specific spatial location). A nine-point calibration and validation procedure was used at the beginning of the experiment. Participants' verbal explanations and gestures were recorded with a Flip video camcorder.

Each participant took part in an individual session lasting 20-40 minutes. At the beginning of the session, participants were given a short explanation of what to expect in the study. After calibrating the eye tracking system, if the validation's mean error was ≤ 0.50° of visual angle, the experiment began—otherwise the calibration and validation was repeated until successful. Next, the participant was instructed to silently answer 10 multiple-choice questions while their eye movements were recorded. Between questions, a calibration drift correction procedure was done to ensure proper calibration throughout the experiment. Participants indicated their answer to each question using number keys on the keyboard. Finally, each participant was asked to provide a cued verbal retrospective report (van Gog, Paas, van Merriënboer & Witte, 2005) for which they were shown a replay of their eye movements on each problem and asked to explain their thought processes (either after watching the replay of their eye movements or concurrently while watching them). This method has been found to produce more in-depth explanations than without viewing one's eye movements. If a participant's explanation was unclear, they were asked follow up questions. Participants were given unlimited time to answer the questions and provide retrospective verbal reports.

Analysis and Results

To analyze participants' eye movements, we created areas of interest (AOIs) which specified areas of the diagram that were used to determine the fixation time i.e., the total amount of time the participant spent looking at a given region (see Figure 1). There were two types of AOIs defined for a subset of the problems, these types being novice-like AOIs and thematically relevant AOIs. The thematically relevant AOIs are those which one would need to attend to in order to correctly answer the physics problem. The thematically relevant AOIs were defined by three independent raters, one physics professor and two physics graduate students. The AOI definitions were compared and any discrepancies were resolved through discussion. The definition for the novice-like AOIs comes from a previous interview study (Madsen, Larson, Loschky & Rebello, 2012) where 13 students (eight females) enrolled in an introductory psychology course participated in individual think aloud interviews. All of the participants had taken at least one physics course in high school, though some had taken an introductory physics course at the university level as well. The students were shown a set of 10 conceptual physics problems with a diagram and asked to explain the reasoning, which led them to their answer. The answers and reasoning of those who answered incorrectly were coded and analyzed to produce the definitions for the novice-like AOI. On four of the problems, there was not a consistent area those who answered incorrectly cited for their answer, so these four problems were not included in this analysis.

The fixation time in the AOI for each participant on each problem was determined from eye tracking data. To account for any differences in the total viewing time on each problem, the fixation time in each AOI was divided by the total viewing time for the diagram on a given problem to get the percentage of time an individual spent in a particular AOI. For each problem, the percentage of time spent in each type of interest area was compared between participants who answered the problem correctly and those who did not using a one-way ANOVA with *percentage of time* for all three AOI types as the dependent variable and *correctness of answer* as the independent variable. Mean percentage of fixation duration and standard error for the correct and incorrect responders for each question are shown in Table 1. We found that on five out of six problems analyzed, participants who correctly answered the question spent significantly more time looking at the thematically relevant areas of the diagrams than those who answered the same question incorrectly. There were also five problems where there was a significant difference between the correct and incorrect responders in the percentage of fixation time in the novice-like portions of the diagrams. Thus, we have identified several problems which are good candidates for visual cueing, as these problem diagrams contain areas which correct solvers spend more time looking at as well as areas that incorrect solvers spend more time looking at. These findings are also consistent with previous findings (Jarodzka et. al, 2009; Antes & Kristjanson, 1991; Charness et. al, 2001) that experts spend more time looking at relevant visual elements while novices spend more time looking at irrelevant visual elements. It should be noted that in our study we compared the visual attention of those who answered the problems correctly to those who answered incorrectly, where previous studies compared participants' eye movements based on expertise.

Study 2: Using Dynamic Visual Cues to Influence Reasoning

Method

Participants in the study were 55 individuals concurrently enrolled in an introductory algebra-based physics course. To ensure sufficient prerequisite knowledge, each completed a pre-test, which consisted of four open-ended questions gauging their understanding of speed and potential energy. The pre-tests were scored as correct or incorrect by one of the researchers. When a participant's answer was unclear, two researchers discussed the answer and agreed on a conclusion.

Participants took part in individual sessions lasting between 30 and 60 minutes. They were first given an explanation of what to expect and the eye tracker was calibrated. Next, participants were instructed to spend as much time as needed on each question and answer with a verbal explanation of their reasoning when ready. Participants in the cued condition were told that colored shapes may appear on some of the problems and when these appeared, they should follow them with their eyes.

Participants were presented with physics problems on a computer screen viewed at a distance of 24 inches using a chin and forehead rest to minimize participants' extraneous head movements. The resolution of the computer screen was set to 1024 x 768 pixels with a refresh rate of 85 Hz. Each physics problem subtended 33.3° x 25.5° of visual angle. Eye movements were recorded with an EyeLink 1000 desktop mounted eye-tracking system which had an accuracy of less than 0.50° of visual angle. A nine-point calibration and validation procedure was used at the beginning of the experiment. Participants' verbal explanations and gestures were recorded with a Flip video camcorder.

The materials consisted of four sets of conceptual physics problems covering energy and kinematics, which were found to have significant differences in the way correct and incorrect solvers answered them in Study 1 (See Figure 1). It should be noted that these four specific problems were chosen from the six problems analyzed in Study 1 because they tested distinct concepts in physics. There were three problems in Study 1 that tested a similar physics concept, so only one of these problems was included in Study 2.

Within each problem set, there was an "initial" problem, four "similar" problems and a "transfer" problem. All problems were open-ended and contained a diagram with the necessary information needed to answer the problem. First, students answered the initial problem to demonstrate their current level of understanding. If they answered incorrectly, they saw a series of "similar" problems, which contained the same problem statement as the initial problem and tested the same concept and contained a diagram with similar surface features. When the student answered a similar problem correctly, they saw the transfer problem. The surface features of the transfer problems were different than the initial and similar problems, though the concept tested was the same. All participants viewed the four sets of problems in the same order.

Whenever a student was ready to answer a problem, they indicated this by pressing any key on a keyboard, at which point the problem displayed on the computer would become slightly smaller in size (this was so that the student knew they had successfully pressed a key). The participant then explained their answer and reasoning to the experimenter and were able to point to areas on the computer screen if necessary. The experimenter used a pre-defined rubric to determine if the given answer and explanation were correct or incorrect. If the answer and/or reasoning were vague, the experimenter would ask for clarification. Once the experimenter had sufficient information to determine the correctness of the answer, the experiment would proceed.

Participants in the cued group saw colored shapes overlaid on the similar problems appear four seconds after the problem was initially seen. Colored shapes were used because color is known to capture one's attention because of its high visual salience. Each colored shape appeared for 500 ms at 12 different positions in the diagram for a total cueing time of six seconds. The visual cues were designed to mimic the eye movements of those who answered the same problems correctly in Study 1. There is a large variation in eye movements from one individual to another while viewing the diagrams in these physics problems, so the visual cues could not mimic the eye movements of correct solvers exactly. Instead, video playback of the correct solvers' eye movements was viewed repeatedly and special attention was paid to the eye movements in and around the thematically relevant area of interest. Similarities between participants were observed, and visual cues modeled after these patterns. Further, the cues could have remained static and simply drawn participants' attention to the relevant areas of the problem, but we hoped by modeling the way in which correct solvers viewed the thematically relevant areas and compared elements within these areas, the cues would give the participants more insight into how to correctly answer the problems.

Analysis and Findings

In order to determine if these visual cues had a positive influence on participants' answers and reasoning, we compared the number of students who answered the "initial" problem incorrectly and then changed to a correct answer and reasoning on a "similar" problem. We used a Mann-Whitney U test to compare the number of participants in the cue and no cue groups who changed to a correct answer on any of the four similar problems seen in a problem set. We found a significant difference on the roller coaster problem ($p = .002$) where six students in the cued group ($N = 18$) changed to the correct answer while zero students in the no cue group ($N = 14$)

made this change. There were no significant differences found between the cue and no cue groups in the number of changes to a correct answer on the "ball," "skier," or "graph" similar problems. It is interesting that only one of the four problems sets studied seemed to show a difference between the cue and no cue groups. There are many possible reasons for this difference. One is the design of the cues themselves. The rollercoaster problems used repetitive simple cues while the cue patterns on the other three problems were more complex. It is possible that in the short six second cueing period, the participants couldn't draw meaning from the more complex cue patterns, and thus they were ineffective. This hypothesis will be tested in future studies.

To determine if visual cueing is useful for learning beyond the problem being cued, we compared the correctness of the answer and reasoning on transfer problems between the cue group and no cue group for each problem set. Figure 3 shows the percentage of participants who answered the transfer problem correctly after answering the initial problem incorrectly. We compared the cue and no cue groups' performances on the transfer problems using the Mann-Whitney U test. We found that there is a nearly significant difference for the ball transfer problem ($p = .06$) and the graph transfer problem ($p = .054$). There was no difference found for the roller coaster and skier transfer problems. These results suggest the visual cues in the ball and graph problem sets positively influenced performance on some of the related transfer problems.

It is curious that there is no difference between the cue and no cue groups on the transfer problem from the rollercoaster problem set, though we did find a difference in the number of students who answered similar problems correctly after answering the initial problem incorrectly. It seems that this transfer problem may have been difficult for this level of student as only one out of 14 students in the no cue group answered correctly while two out of 18 students in the cue group answered correctly. A similar explanation could account for no difference between the cue and no cue groups on the skier problem. It is also possible that the students did not view the concept tested in the transfer problem to be similar to that tested in the similar problems, even though the researchers did see these similarities. In other words, from the point of view of our students, the transfer task can be interpreted as primarily a far transfer task, while the similar problems constituted near transfer and thus were not as challenging for the students. These interpretations from students' perspectives may be expected given that the similar problems shared the same surface features with each other, rather than the transfer task which had different surface features, and only shared the deep structure with the training problems.

Discussion and Conclusion

In this study we find some evidence that viewing a physics problem overlaid with short duration visual cues can indeed help students correctly answer and reason about problems they were previously unable to. Of the four problem sets used, we found on one of these problem sets significantly more students changed to a correct answer after seeing cues. It is not enough though, to provide visual cues to help students answer a given set of problems. In looking at transfer problem performance, we found nearly significant differences on the ball and graph transfer problems with the cue group outperforming the no cue group. Thus, we find some evidence that repeatedly showing novices visual cues on related problems may help them to properly apply the factual knowledge on similar future problems viewed without cues.

While we did find some results that point to the usefulness of cueing, we also saw no difference between cue and no cue groups in the number of similar problems answered correctly on three of the four problem sets tested. There are many reasons that the cueing may have failed.

First, we have previously discussed how the cue pattern may have been too complex on these three problems. In the future, we can use simple repetitive cues as well as increase the time the cues are seen. We can also change the type of cue we are using, for example instead of moving colored shapes modeling correct solvers' eye movements, we could use lines or bars to illustrate comparisons in the diagram that correct solvers make. Additionally, there may be only certain types of problems that lend themselves to improvement through visual cueing. We have only explored four problems in this study. There are a plethora of problems that can be categorized in a variety of ways and tested in future studies. It could also be that the order in which the problems are presented influences the usefulness of the cue. The roller coaster problem was presented first each time and was the only problem the cues were found to influence. In future studies, the order of cued problems will be randomized to balance out any order effects.

We also found differences between the cue and no cue groups on two of the four transfer problems tested. As mentioned earlier it may be that the two transfer problems that showed no difference were too difficult for this level of student, as very few students in either group answered these problems correctly. It is also possible that the researchers viewed the transfer problems as closely related to the similar problems, though the students did not view them this way, and thus were not able to apply what they gained from the cues to the transfer problems.

Previous studies with visual cues in several domains have also found mixed results on the effectiveness of cueing. There is much work to be done to understand the factors that lead to helpful cues. This study offers some hope that cueing can potentially serve as effective conceptual scaffolding for novice physics students, but much work is necessary to perfect this method.

References

Antes, J. R., & Kristjanson, A. F. (1991). Discriminating artists from non-artists by their eye-fixation patterns. *Perceptual and Motor Skills*, 73, 893-894.

Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition*, 29, 1146-1152.

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review, 21*(2), 113-140.

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.

Grant, E. R. & Spivey, M.J. (2003). Eye movements and problem solving. *Psychological Science, 14*(5), 462.

Jarodzka, H. Scheiter, K., Gerjets, P., & van Gog, T. (2010) In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154.

Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human-Computer Studies, 65*(11), 911-930.

Lowe, R.K., (2003). Animation and learning: Selective processing of information in dynamic graphics, *Learning and Instruction.* 13, 157–176.

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

Mason, A. & Singh, S. (2011) Assessing expertise in introductory physics using categorization tasks. *Physical Review Special Topics Physics Education Research.* 7(2), 0201101-02011017.

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology, 93*(2), 377-389.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior, 26*(1), 110-117.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Sweller, J. (1999). *Instructional design*. Melbourne: ACER Press.

Van Gog, T., Paas, F., van Merriënboer, J.J.G., Witte, P. (2005) Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. *Journal of Experimental Psychology, Applied*, 11(4), 237-244.

Table 1

Mean percentage time spent (± std err) during entire problem period for thematically-relevant and novice-like AOIs for participants who answered the question correctly/incorrectly (indicates a significant difference, $p < .05$).*

Thematically Relevant AOI		
Problem #	Answered Correctly	Answered Incorrectly
1	46.6 (± 5.5) (n = 11)	33.2 (± 5.7) (n = 11)
2*	24.4 (± 2.9) (n = 13)	11.6 (± 3.3) (n = 10)
3*	28.5 (± 4.1) (n = 18)	8.9 (± 2.3) (n = 6)
4*	49.8 (± 3.9) (n = 14)	25.5 (± 4.1) (n = 9)
7*	36.7 (± 5.5) (n = 15)	10.3 (± 2.1) (n = 9)
10*	29.0 (± 5.0) (n = 11)	15.1 (± 2.7) (n = 13)
Novice-Like AOI		
	Answered Correctly	Answered Incorrectly
1*	22.3 (± 4.5) (n = 11)	43.5 (± 7.3) (n = 11)
2*	12.7 (± 3.3) (n = 13)	27.2 (± 4.8) (n = 10)
3*	19.8 (± 3.7) (n = 18)	39.4 (± 5.4) (n = 6)
4 (p=.058)	18.1 (± 2.5) (n=14)	26.8 (±3.9) (n=9)
7*	12.6 (± 2.6) (n = 15)	25 (± 6.0) (n = 9)
10*	41.2 (± 6.6) (n = 11)	62.2 (± 5.1) (n = 13)

Table 2

Number of students in cued and no cue group who answered a similar problem correctly after answering the initial problem incorrectly (indicates a significant difference, p<.05).*

	Rollercoaster Problem*		Ball Problem		Skier Problem		Graph Problem	
	Cued (N = 18)	No Cue (N=14)	Cued (N=10)	No Cue (N=14)	Cued (N=11)	No Cue (N=7	Cued (N=17)	No Cue (N=22)
Changed to Correct Answer	6	0	6	4	2	2	4	1

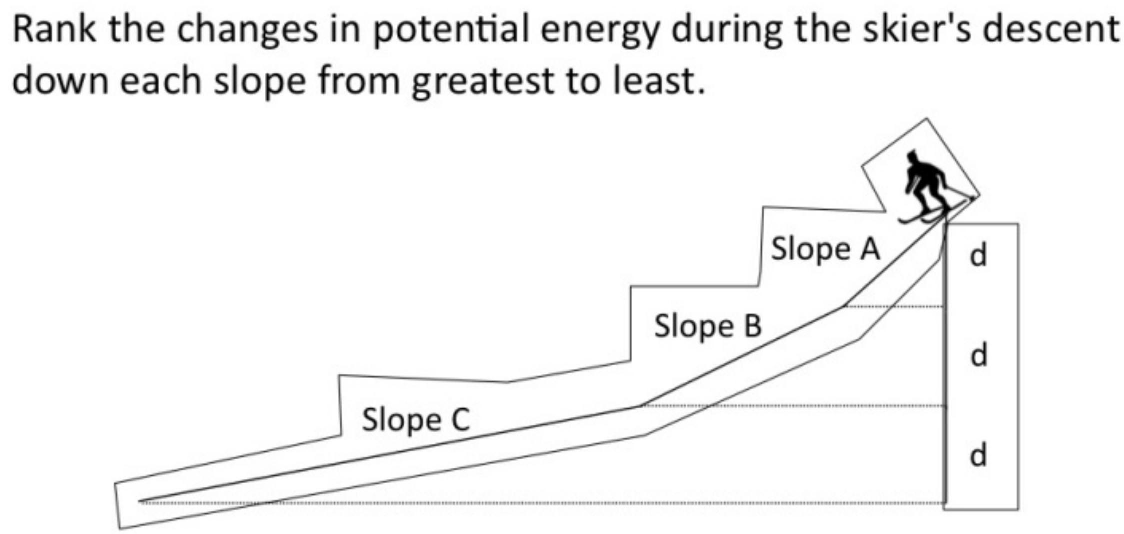


Figure 1. Problem 10 used in Study 1. Novice-like AOI along slope. Thematically relevant AOI along height of hill.

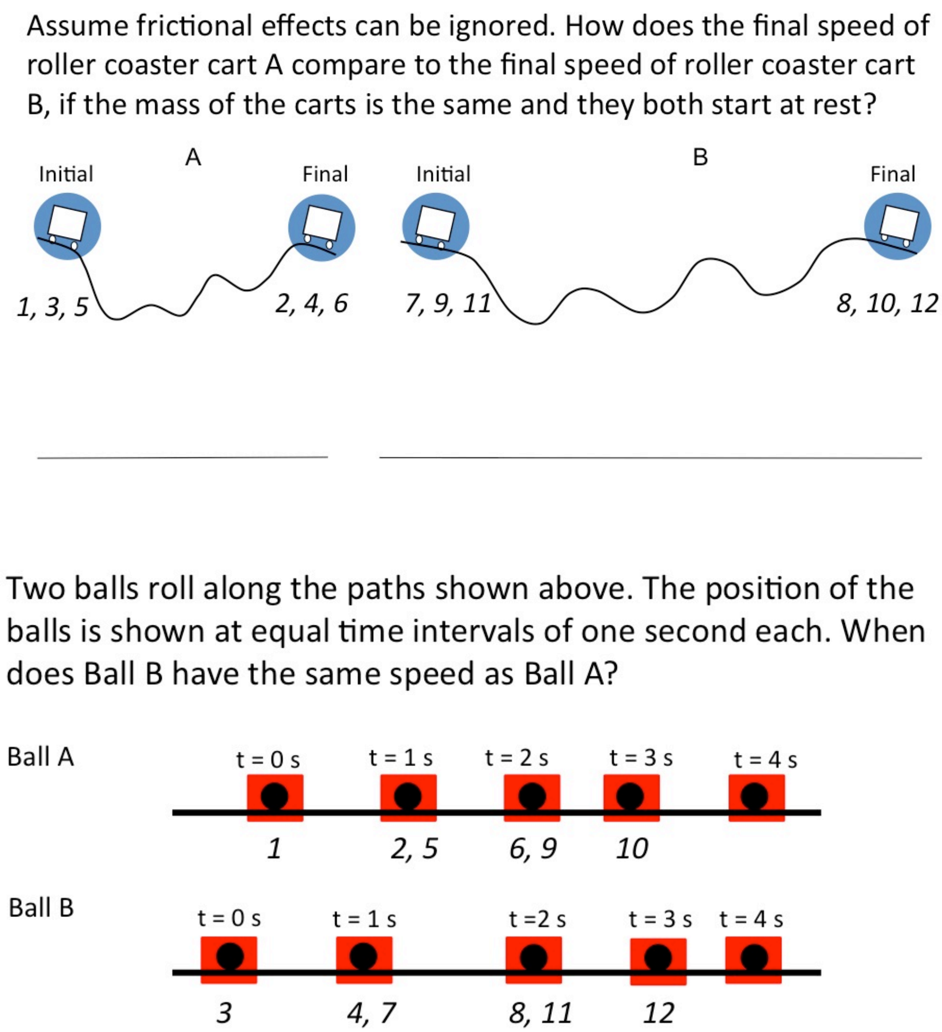
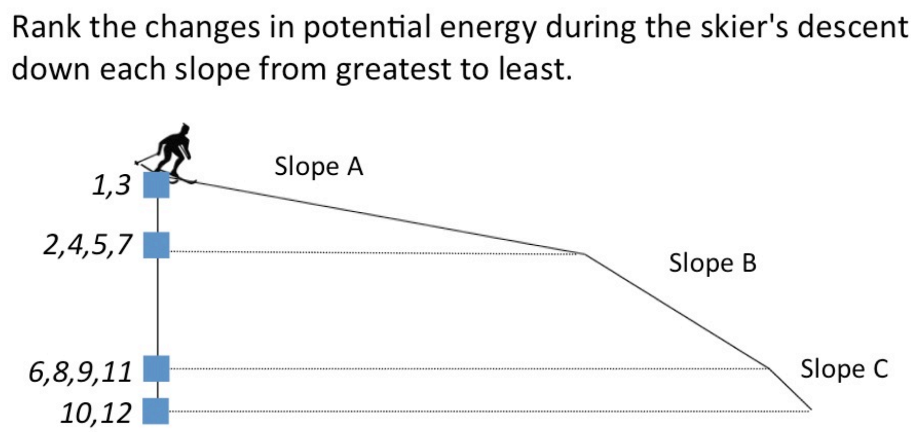


*Figure 2***.** Problems 1 and 2 used in Study 2. Colored shapes are visual cues overlaid on the problem diagram. Numbers in italics show the sequence of animated cues (the numbers were not seen by study participants).



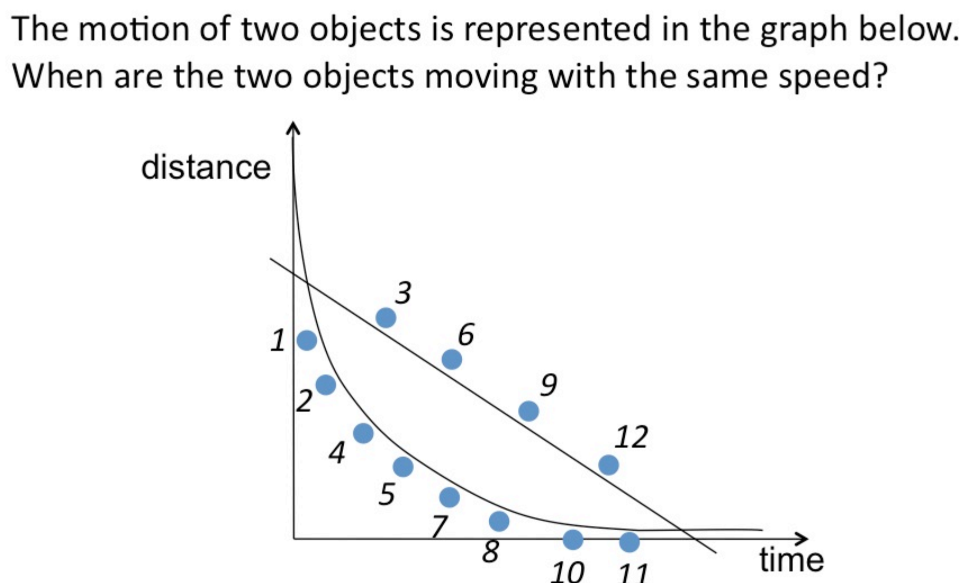


*Figure 3***.** Problems 3 and 4 used in Study 2. Colored shapes are visual cues overlaid on the problem diagram. Numbers in italics show the sequence of animated cues (the numbers were not seen by study participants).

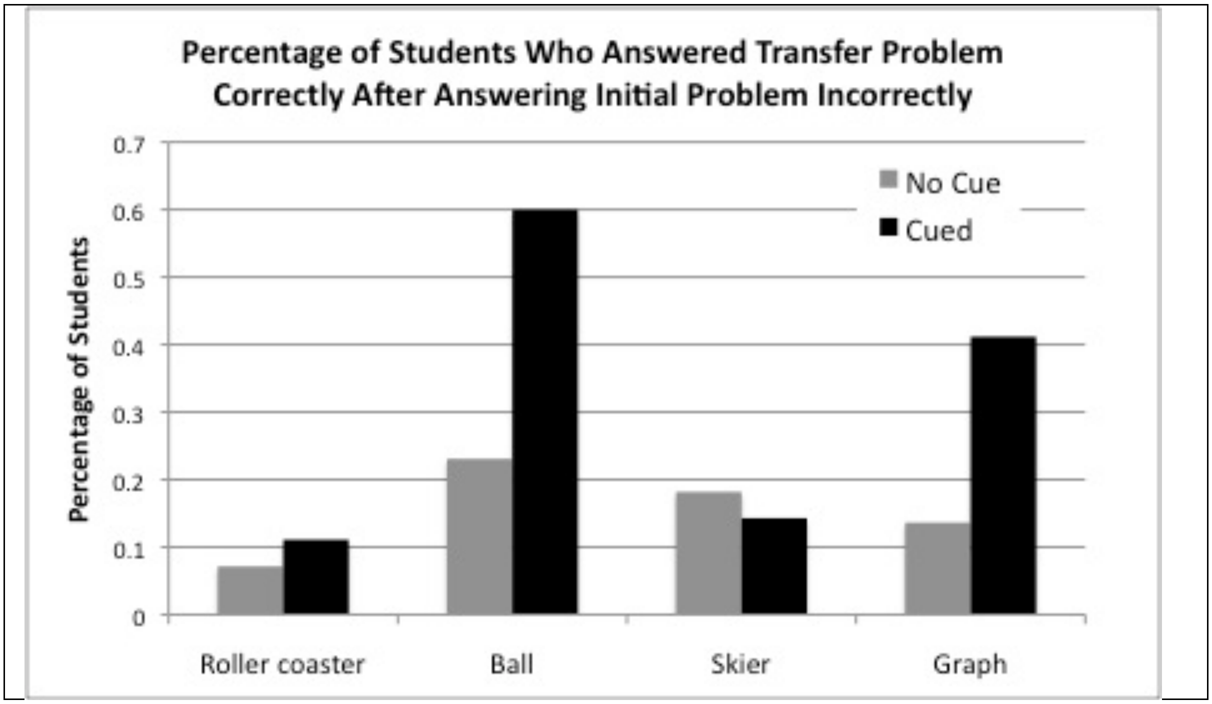


Figure 4. Percentage of students in "cued" and "no cue" conditions who answered initial problem incorrectly, but answered transfer problem correctly.