



Eye movements in the manipulation of hands-on and computer-simulated scientific experiments: an examination of learning processes using entropy and lag sequential analyses

Yu-Cin Jian¹ · Leo Yuk Ting Cheung¹ · Yi-Jye Wu¹ · Fang-Ying Yang² · Guo-Li Chiou³

Received: 31 March 2022 / Accepted: 18 April 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

Computer-simulated experiments have been gaining popularity over hands-on experiments in science education, given the availability of technology and the trend of distance learning. Past studies have focused primarily on comparing the learning outcomes and user experiences of the two experiment modes. In this study, we used an eye tracker to investigate the learning processes involved in manipulating hands-on and computer-simulated experiments, and the effect of prior knowledge and experiment mode on eye movements. A total of 105 undergraduates completed either mode of experiment to learn about pulley mechanics. Participants were asked to read relevant concepts before conducting the experiments to ensure they had basic knowledge about the subject matter. Results showed that the learning outcome of experimentation was affected by prior knowledge but not experiment mode. As for eye movements, the two experiment workstations were divided into nine functional regions. The findings revealed that eye movements in most regions were affected by the experiment mode, but not prior knowledge. The simulation group had shorter total fixation durations and smaller pupil sizes than the hands-on group, implying a lower cognitive load in learning in computer-simulated experiments. Lag sequential analysis and entropy analysis were conducted on cross-regional fixation transitions. The results revealed that participants in hands-on experiments tended to make more diversified fixation transitions across regions, whereas those in simulated experiments showed a higher level of concentration in the spatial pattern of fixation transitions. While sequential analysis offers insights into important fixation transitions on a regional level, entropy analysis allows for a more macro perspective on the overall transition distribution and facilitates conventional statistical modeling that takes individual differences into account.

Keywords Computer-simulation · Eye movements · Prior knowledge · Scientific experiments · Lag sequential analyses

Introduction

In modern science education, experimentation is an important way to develop science literacy (Jian, 2022a; National Research Council, 2000). Through experiments, students can experience changes—often described through abstract concepts in the books—in physical quantity, enabling a deeper comprehension of science concepts (Triona & Klahr, 2003). Apart from traditional, hands-on experiments, computer technology allows for virtual forms of scientific inquiry with the benefits of increased availability, higher flexibility, lower cost, precise manipulations, and support for self-paced learning (Chen, 2010; Hofstein & Lunetta, 2004). Recently, due to the COVID-19 pandemic and the growing trend of distance learning, virtual experiments have become even more widely adopted in educational practice, and in many cases, they are gradually replacing physical manipulatives (Klein et al., 2021; Radhamani et al., 2021). However, its effects on science learning are somewhat mixed (meta-analysis: Rutten et al., 2012), as the advantages of computer simulations might be, subject to software design and hardware limitations, offset by reduced affordances of kinesthetic experiences with physical objects (Martin & Schwartz, 2005). These studies on the simulated experiments have so far focused on the learning outcome and overall learner experience, in comparison with hands-on manipulations. It is high time the research paradigm with regards to virtual inquiry should have widened its scope and considered the moment-by-moment process of learning.

Moment-by-moment learning process has so far been extensively studied by eye-tracking with regards to learning activities as reading (cf., review paper, Rayner, 1998). The technology was, however, rarely deployed to study the learning process of scientific experiments (e.g., Jian 2022a; Chien et al., 2015). In empirical studies, the learning process of experimentation tends to be either neglected or, for virtual experiments, to rely on keypress logs as data (e.g., Wen et al., 2020). For the latter, the method does not support the study of hands-on experiments. Eye-tracking enables analysis of the learning process, regardless of the nature of manipulatives. Therefore, this study aimed to use eye-tracking technology to examine the learning processes that were involved in carrying out hands-on and computer-simulated scientific experiments.

Theoretical basis of hands-on and simulated experiments in science learning

Hands-on and computer-simulated experiments are widely used to help students learn scientific theories and concepts in all levels of education. On the one hand, hands-on experiments are based on the theory of cognition (Shapiro, 2011), which states that the human body and the environment play important roles in cognitive processing, and that behaviors—such as gestures and actions—can be leveraged to improve learning. Weisberg and Newcombe (2017) have suggested that embodied cognition provides learners with an opportunity to link sensory representations—such as touching an object or visualizing data or information—with abstract scientific principles. On the other hand, virtual simulations are suggested to afford the provision of dynamic visualizations and analogies for abstract and unobservable phenomena to help learners integrate corresponding scientific knowledge, representations, and practices (Chien et al., 2015; Olympiou & Zacharia, 2012; Zacharia & Olympiou 2011).

The cognitive process that is involved in hands-on and computer-simulated experiments can be described by the cognitive theory of multimedia learning (Mayer, 2009). According

to Mayer, human cognition has two information processing channels: the visual-pictorial and auditory-verbal channels. In the visual-pictorial channel, visual and textual information in the learning environment is selected based on the learner's criteria of relevance and is constrained by the working memory capacity of the learner. Both types of incoming information are then organized into coherent, but separate, representations. Lastly, by engaging relevant prior knowledge, the organized representations were further reconstructed and integrated to form new knowledge. Based on the cognitive load theory (Plass et al., 2010), we hypothesized that computer-simulated experiments have an advantage over hands-on experiments because the latter make additional demands on tactile information and kinesthetic sequences for completing a task with more dexterity. However, it has been argued that, compared to computer simulations, physical manipulatives involve both tactile and visual modalities for conducting hands-on experiments, and the cognitive load could be reduced by using the tactile modality as the third information processing channel (Zacharia & Olympiou, 2011).

In terms of the empirical studies of computer-simulated versus hands-on experiments, most studies noted the superiority of simulations in enhancing learning achievement (review: Rutten et al., 2012), even among children (Klahr et al., 2007). The benefits of computer simulations include improved conceptual understanding (Chini et al., 2012; Zacharia, 2005), higher time efficiency (Gibbons et al., 2004), better cognitive focus on the subject matter (Winberg & Berg, 2007), increased motivation (Baltzis & Koukias, 2009), and relatively easy customization to meet learners' needs (Sullivan et al., 2017). Computer simulations are particularly useful for experiments that are potentially unobservable, costly, or dangerous (Zacharia & de Jong, 2014). However, the hands-on vs. simulated comparison conducted by Sullivan and colleagues (2017) on pulley mechanics showed no notable difference regarding overall learning outcomes among eighth-grade students, although simulations resulted in a better understanding of the subtopics of physical work and mechanical advantage. However, other studies showed similar effectiveness for both learning experiences (e.g., Dalgarno et al., 2009; Zacharia & Olympiou, 2011). Chini et al. (2012) manipulated the sequence of hands-on and simulated experiments of pulley mechanics among a group of non-science majors; comparisons of learning outcome found that simulations resulted in better improvements in the overall conceptual understanding, as well as work and energy subtopics. However, a few studies had found that students performed better in their learning through manipulating hands-on experiments rather than through virtual experiments (Chang et al., 2008). Apparently, the findings in the literature were inconsistent and more empirical studies are needed to resolve this issue.

Eye-movement research in science experimentation

The eye-mind hypothesis holds that when an eye is relatively stabilized and fixated on an object in the environment, the person is very likely to deploy attention and cognitive capabilities (Rayner et al., 2006). To our knowledge, only two studies used eye-trackers to investigate the learning processes of manipulating scientific experimentation (Jian, 2022a; Chien et al., 2015). Chien and colleagues (2015) compared the eye fixations of high school students while performing Boyle's experiment either in physical laboratories or through a computer-based simulation. The results indicated that the simulation group performed more trials and paid greater attention to the experiment, while the physical-laboratories group spent more time on the worksheets. They also observed that the simulation group began by manipulating a virtual experiment before they thought about the questions in the

worksheets. However, the physical-laboratories group tended to think about the questions in the worksheets before carrying out the experiment. In another study (Jian, 2022a), an eye tracker was used to investigate whether undergraduate students demonstrated different learning processes during a problem-solving task (pulley system) by reading easy and difficult scientific articles, and then conducting the experiment. The results showed that the difficult-article group preferred to check regularly on whether their manipulations had solved the test questions, while the easy-article group seldom did so. Another notable finding was that, based on observations of online eye movements, the participants seemed to have very limited procedural knowledge on how to manipulate the pulley system even though they were taught the basic concepts to do so in middle school.

The above two studies on eye-movement research (Jian, 2022a; Chien et al., 2015) were attempts to pioneer the study of processes that are involved in manipulating science experiments. While Chien et al. (2015) used the two regions of doing experiment and worksheet as the areas of interest (AOIs) to investigate how learners paid attention to both regions in physical and virtual manipulations, the study by Jian (2022a) examined only the use of physical (hands-on) manipulation in science learning. In addition, both studies used fixation durations (the sum of all fixations on a region) to measure eye movements in order to understand the regions that required more attention or cognitive effort (Jian, 2019, 2021, 2022b; Mason et al., 2013; Miller, 2015; Wang et al., 2022; Wu et al., 2021). However, pupil size is another measurement that could be used to assess cognitive load and mental effort (Alemdag & Cagiltay, 2018; Appel et al., 2019; van Gog & Jarodzka 2013; Van Gog et al., 2009). Therefore, this study investigated whether learning processes, in terms of fixation durations, pupil size, and fixation transitions, were different between hands-on and simulated experiments, and whether prior knowledge, a very important variable in scientific learning, played a role in these processes. This study also divided the regions of doing experiment (a pulley system) and worksheet into smaller units (e.g., fixed-pulley lift marks area, movable-pulley lift marks area, material area, diagram area, and formula area) in order to examine the specific cognitive processes involved in a scientific problem-solving task.

The role of prior knowledge in experimentation

Prior knowledge, or pre-existing domain knowledge, has long been identified as a robust predictor of student achievement in science, even after controlling for motivation, learning goals, and learning strategies (Song et al., 2016). In guided inquiry for science learning, prior knowledge has also been identified as a pre-requisite for a basic understanding of the topic of inquiry; related concepts and theories are required in the initial phase of the inquiry learning process (Pedaste et al., 2015). For instance, learners need some basic knowledge to understand the questions central to the inquiry and to become motivated to learn about the topic, which is in turn expected to affect the learning process and outcome. Empirical studies have found that knowledgeable learners, for instance, tended to adopt theory-driven approaches to inquiry-based tasks (Lazonder et al., 2008); and that prior knowledge was a significant predictor of successful learning outcome from experimentation, especially if appropriate guidance was provided (Wen et al., 2020; van Riesen et al., 2018). As such, the design of this study was conducted in two phases to resemble a typical instructional procedure: In the first phase, learners were engaged in reading about pulley mechanics, which was aimed to provide essential scaffolding for students to reach the zone of proximal development before performing experiments in the second phase (Mestad

& Kolstø, 2014). Despite the reading materials, knowledge about pulley mechanics would still vary among participants. Such individual difference could affect the learning outcome through differential learning processes. We examined whether the eye-movement patterns during experimentation was affected by levels of prior knowledge, in conjunction with computer simulations.

Research questions

The overall objective of this study is to examine the eye movements during a computer-simulated versus hands-on experiment, from which the learning process could be inferred, and how the eye movements were affected by prior knowledge and experiment mode. Specifically, the present study aimed to answer the following research questions:

- (1) Is the learning outcome of experimentation affected by prior knowledge and experiment mode (hands-on versus simulated)?
- (2) Are eye-movement indicators (e.g., fixation durations and pupil size) affected by prior knowledge and experiment mode in the regions of manipulating experiments and work-sheets?
- (3) Are cross-regional fixation transitions affected by prior knowledge and experiment mode in the manipulation of an experiment?

Methods

Participants

In total, 105 university students in Taiwan (Female: 83; Age: 18–25) were recruited. To ensure that participants had little pre-existing knowledge of pulley mechanics, only students who were enrolled in the college of humanities or social sciences were recruited. A total of 53 participants (Age: $M=20.66$, $SD=1.73$) were assigned to the hands-on group, and the remaining 52 participants (Age: $M=20.37$, $SD=1.73$) formed the simulation group. The result of a t-test showed that the ages for the two groups did not significant differ ($p>.05$). In the hands-on group, the number of participants who majored in arts, business, education, and social science were 11, 4, 23, and 15 respectively; in the simulation group, the numbers were 12, 3, 26, and 11, respectively. The result of a chi square test showed that the distribution of majors in the two groups did not significant differ ($p>.05$). All the participants were native readers of Traditional Chinese, and had normal or correct-to-normal vision. Informed consent was obtained from all participants prior to the procedure.

A sensitivity analysis using G*Power (Faul et al., 2007) suggested that, with an alpha level of 0.05, and given our sample size, the study design was expected to yield at least a medium effect size (Cohen's $f=0.31$ or $\eta^2_p=0.09$) for main effects and interactions (Cohen, 1988).

Procedure

The procedure was carried out in two stages: a reading task, followed by a guided science inquiry (Fig. 1), which were separated by a gap of four to seven days. The reading

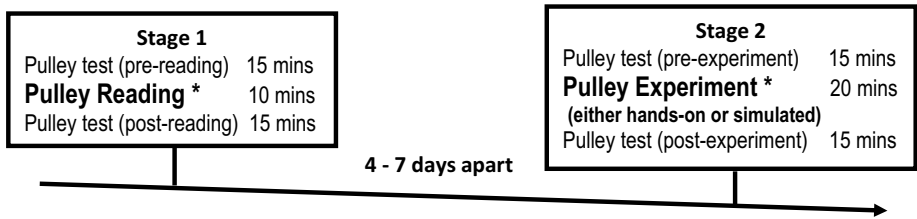


Fig. 1 Two-stage study procedures. The same procedure was applied to all participants, except for the mode of pulley experiment. The two experiment modes (hands-on and simulated) were matched with post-reading pulley test results

task ensured that our participants had acquired some basic knowledge about pulley mechanics before performing the experiment. Upon arrival, participants were briefed about the study before signing the informed consent form. They were then instructed to sit in front of a tablet device and complete a spatial ability test within a 10-min time limit, followed by a pulley test (pre-test) that did not exceed 15 min. Thereafter, they began to read a nine-page expository text about pulley mechanics on a computer screen for 10 min maximum. As the text was a rather short passage (see the next section for details), no participants reported that the reading time was insufficient. After reading the pulley text, they performed the post-reading pulley test on the tablet device for 15 min. Stage 1 took approximately an hour.

Based on the participants' scores in the post-reading pulley test, those with scores above the median and those with scores below the median were each equally assigned to either the hands-on experiment condition ($n=52$) or the computer simulation condition ($n=53$) in the subsequent pulley experiments. The difference in the mean of post-reading pulley test scores between the hands-on group ($M=20.04$, $SD=4.38$) and the simulated group ($M=20.16$, $SD=4.44$) was non-significant, as revealed by a Welch's independent t-test, $t(102.57)=0.14$, $p=.89$.

Participants returning to Stage 2 of the study completed, again, a computerized pulley test that did not exceed 15 min. Afterwards, in preparation for the pulley experiment, the participant put on a pair of eye-tracking glasses, with their position fixed by a band that was fastened around the participant's head. The participant was seated around 35–50 cm in front of the objects at the workstation. Before experimentation began, participants were given ample time to familiarize themselves with the way pulley manipulations should be performed, under the experimenter's standardized guidance. This is especially important for the hands-on experiment group, due to the variety of objects involved and the need to assemble the objects and set up pulley systems. However, operations on the pulley software for the simulation group were also thoroughly instructed. After the participant acknowledged to the experimenter that the operations were clear to them, they were given a maximum of 20 min to work on either a hands-on or computer-simulated experiment and to complete a laboratory report worksheet. The experimenter withdrew to the adjoining room, observing participants using the scene camera on the head-mounted eye tracker. Upon reaching the time limit, the experimenter entered the room again and removed the eye tracker from the participants, who then sat for the 15-min post-experiment pulley test. Finally, participants were debriefed, compensated, and released. Stage 2 lasted for approximately an hour.

Materials

Reading materials

The reading materials on pulley systems were adapted from relevant chapters of the Mandarin translation of a science multi-volume book written for high school students; it is entitled *Conceptual Physics* (Hewitt, 2018), carrying the following sub-sections in order of appearance: “Conservation of Energy,” “Three Orders of Levers,” “Friction,” “Features of Pulley System: Energy Saving and Time Saving,” “Work,” and “Mechanical Advantage,” which corresponded to the six subtopics of pulley mechanics in the analysis. The text was divided into nine pages – participants had to finish reading one page before moving on to the next. The text was 1817 characters long.

Assessment tests and laboratory report worksheet

In order to measure spatial ability—specifically the ability to perform spatial rotation mentally—the Purdue Visualization of Rotations Test (Bodner & Guay, 1997) was administered as it is shown to facilitate learning in physics (Kozhevnikov & Thornton, 2006) and to predict outcomes in science learning (Hodgkiss et al., 2018). The test required participants to match a rotated, irregular, and rod-shaped 3D object from among five alternatives. It contained 20 matching questions, and each of these contributed equally to the overall score. For our study, the internal consistency of this tool showed split-half reliability coefficients of 0.78–0.85.

The pulley test (Online Appendix 1) was modified based on the assessment tests deployed in the Concept Mapped Project-based Activity Scaffolding System (CoMPASS, 2014) curriculum and in Sullivan et al’s (2017) study. The test comprised 32 questions, including 17 multiple-choice questions and 15 short-answer questions. The questions and the marking scheme were subsequently edited by a researcher in physics. The test covered the six aforementioned subtopics of pulley mechanics. This test served to measure the learning outcome of both reading and experimentation. The short-answer questions were graded by two research assistants, with a high degree of inter-rater agreement due to the nature of questions. With Cronbach’s alpha values ranging between 0.701 and 0.771 across the four administrations of the test on the same sample, the test showed satisfactory internal consistency. The test was used to assess the learning outcome of reading and experimentation.

Regarding question types, the test contained 18 factual, 10 applied, and four experiment questions. The factual questions asked about information that were explicitly stated in the reading materials present in the previous section (e.g., the location of the fulcrum, point of force, etc.). The application questions required participants to apply conceptual knowledge when solving a problem (e.g., how far from the fulcrum of a seesaw one needs to sit in order to achieve balance). Experiment questions covered the specifics of pulley experiment (e.g., which kind of pulley generates a larger mechanical advantage). The variety of question types was aimed to assess different aspects of conceptual understanding. Each participant performed the same test four times: Before and after the reading task, and before and after the pulley experiment. The test was computerized and administered on a Windows Surface tablet device with a time limit of 15 min.

During experimentation, participants were asked to complete a lab report worksheet by recording observations from the experiment and by deriving generalizations about pulley-related concepts (Online Appendix 2). The worksheet contained 18 items, six of which were records of parameter values that covered a fixed pulley, movable pulley, and their combination. The remaining 12 items comprised mostly two-choice questions about work, mechanical advantage, and the discrepancy between an ideal and a real pulley. These questions were adapted from the Pulley Challenge section of the CoMPASS (2014) curriculum. The mean score of the two-choice questions was 9.55 out of 12 ($SD=2.10$), suggesting a satisfactory basic understanding of relevant concepts during the experimentation task.

Apparatuses

Eye tracker

As the activity required considerable body movements, a head-mounted eye tracker, Tobii Pro Glasses 2 (Tobii Technology, Stockholm, Sweden), was deployed in the pulley experiments, with a sampling rate of 100 Hz. A scene camera was attached to the device, and recorded the environment, while the eye cameras kept track of the participant's gaze directions.

Pulley experiment workstation

The experiment workstation was set up on a desk in our laboratory according to the specifications of the experiment mode assigned to the participant (Fig. 2). The desk was electrically adjustable in height, to match with the participant's eye level. In the computer-simulated experiment, a Windows surface tablet device with a 13.5" screen and a keyboard was placed on a stand for participants to perform the pulley manipulations (CoMPASS, 2014). In the hands-on experiment, a metal rack, with two vertical supporting poles, was set up on the desk, allowing the participants to hang pulleys, put ropes around the pulleys, and add weights to form a pulley system (Fig. 3). The left supporting pole was marked with

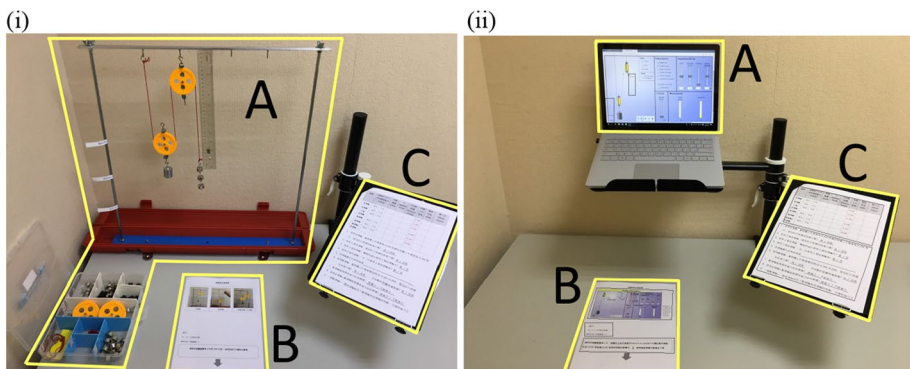


Fig. 2 Workstation for (i) hands-on and (ii) simulated experiments. The yellow frames mark the boundary of three regions under which the areas of interest (AOIs) for processing eye-movement data were grouped. Region A is where pulley manipulations took place; Region B contains an experiment guideline sheet; and Region C consists of a lab report worksheet for learners to complete

reference measurements for the easier gauging of how far the test weight was lifted from the desk surface. Furthermore, a tool box containing materials for the pulley system was placed in front of the metal rack; it included four pulley wheels (6.5 cm in diameter), one long rope (80 cm), one short rope (40 cm), a set of test weights (100 g: 1 piece; 20 g: 15 pieces; 10 g: 12 pieces; 5 g: 15 pieces; 2 g: 10 pieces; 1 g: 10 pieces), and a 30-cm ruler for measuring the distance between the pulley wheel and test weights. In both experiments, two A4-size documents were placed in front of the participants: a laboratory report worksheet to be completed (Fig. 3; Online Appendix 2) and an information sheet that served as a guide for completing the task (Fig. 3; Online Appendix 3a,b). The AOIs are depicted in Figs. 2 and 3, which include every region in the do experiments and worksheets.

Eye-movement measures and data processing

Three eye-movement measures were used to investigate the learning processes that were involved in manipulating hands-on and computer-simulated experiments and in completing a problem-solving task. The first measure was **total fixation durations**, which were calculated by aggregating the durations of all fixations on a specific AOI. This measurement represented the overall level of visual attention paid to this region, which can be linked to the level of cognitive effort (Jian, 2021, 2022b, 2022c; Mason et al., 2013; Miller, 2015; Wu & Liu, 2021). The second measure was **pupil size**, which was derived from the average pupil size of all fixations during a specific AOI. This measurement reflected the cognitive load that is required of a learner to complete a task as extensively shown in previous work (Alemdag & Cagiltay, 2018; Peterson et al., 2015; van Gog & Jarodzka 2013; Van Gog et al., 2009). The last measure was **transitions of fixations**, whereby the overall transition count refers to the number of shifts in fixations that occurred from one AOI to another, and by disregarding intermediary fixations on locations outside all AOIs (Jian, 2016, 2022a; Jian & Ko, 2017; (Chiou et al., 2021, 2022). This measurement depicted the learning sequences that occurred through the interplay of different workstation regions in the experimental task, and how they were affected by experiment mode and prior knowledge. The nine AOIs in this study included the following: observational area; fixed-pulley lift marks area; movable-pulley lift marks area; material area; diagram area; formula area; reminder area; record area; and question area. The definition of every AOI is provided in Fig. 2.

This study adopted two approaches to the analysis of cross-regional fixation transitions, which we believe could complement each other in examining different dimensions of fixation transitions. The first approach, namely lag sequential analysis (LSA), is a common approach to examining between-AOI fixation transitions (Jian, 2016, 2022a; Jian & Ko, 2017; (Chiou et al., 2021, 2022; Lai et al., 2013; Tsai et al., 2022). It compares the probability of transitions from a specified area toward a host of destinations by aggregating—from all the participants—the number of transitions for each cross-regional transition route and converting the aggregates to standardized scores. An incidence of transition routes that are significantly higher than the usual number which is associated with a random occurrence is viewed as an important finding. This approach is particularly useful in examining fixation transitions between individual regions and in identifying frequent and infrequent transitions. The second approach, entropy analysis, is relatively less common in eye-movement analyses, but is frequently used to analyze the uniformness of a discrete probability distribution in other fields (e.g., Schieber & Gilland 2008; Krejtz et al., 2014). A higher entropy value represents a higher level of uniformness in the probability distribution. In the case of fixation durations, a single value of entropy could be derived to represent the

Fig. 3 Nine areas of interest where pulley manipulations took place (i.e., Region A in this figure and Fig. 2) in the (i) hands-on and (ii) simulated experiments; areas of interest on the experiment guideline (i.e., Region B) in the (iii) hands-on and (iv) simulated experiments, and (v) areas of interest within the lab report worksheet (i.e., Region C) in both experiment modes. Refer to the table above for the description of the areas of interest labelled in the pictures

distribution of conditional probabilities on destination AOIs given a source AOI for each participant. This allows for straightforward comparisons between individuals and groups. The downside is that the entropy metric only represents an overall distribution of probabilities; it pales in comparison with LSA in terms of reflecting the pattern of fixation transitions on the regional level. This study adopted both approaches to examine fixation transitions between workstation regions at the macro- and regional levels. For the purpose of performing lag sequential analysis (LSA), the aggregate number of fixation transitions between each AOI for each participant were tabulated to form transition matrices. Also, for deriving conditional probabilities for each transition path, the total number of fixations on each AOI was compiled for each participant, from which the probabilities of events preceding transitions were generated. Transition counts for each of the nine destination AOIs in relation to a source AOI were subsequently converted to Z-scores.

In addition to LSA, we conducted entropy analysis on the distributions of stationary fixations and fixation transitions for comparisons between experiment mode conditions and among prior knowledge groups. The entropy analysis started by calculating the conditional entropy for the transition matrix of individual participants. In order to do so, based on the total number of eye fixations on each of the nine AOIs, we first computed the probabilities of eye fixations on each AOI preceding cross-AOI transitions (i.e., stationary probabilities) for every individual. These probabilities showed the overall distribution of stationary fixations, which could be represented by a Shannon's entropy measure henceforth referred to as stationary entropy (Krejtz et al., 2015). Stationary entropy (\hat{H}_s) was computed for each participant:

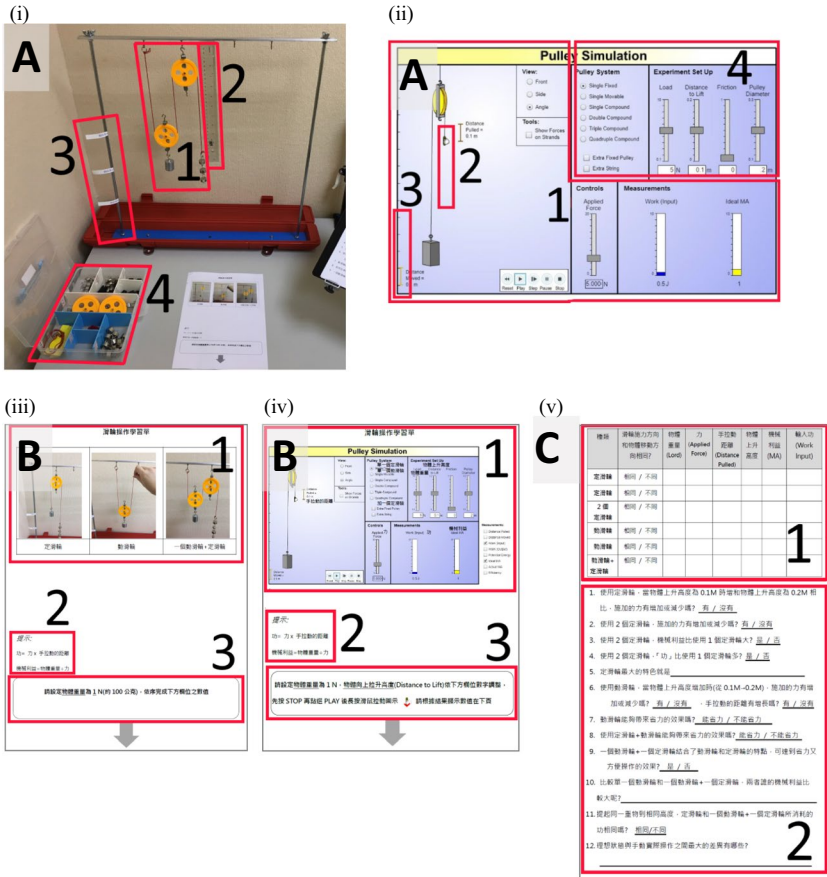
$$\hat{H}_s = - \sum_{i \in S} \pi_i \log_2 \pi_i, \quad (1)$$

where π_i is the stationary probability of an AOI (i.e., i). A high value in entropy suggests a more dispersed distribution of eye fixations among the nine designated regions. After obtaining stationary probabilities, we calculated the conditional probabilities of eye fixations transiting to a destination AOI given the stationary probability of being in a source AOI. These conditional probabilities were then used to compute the transitional entropy (\hat{H}_t) for the whole transition matrix (S) of a participant,

$$\hat{H}_t = - \sum_{i \in S} \pi_i \sum_{j \in S} p_{ij} \log_2 p_{ij}, \quad (2)$$

where π_i is the stationary probability, p_{ij} is the probability of transitioning to a destination AOI (i.e., j), given a source AOI (i.e., i), and $i, j \in S$. The transitional entropy reflected the level of distribution dispersion in relation to the destinations of fixation transitions.

Due to frequent shifts in the scene, the data on eye fixations that were overlaid onto the video—which were recorded by the glasses during the experiment—were subsequently viewed and manually coded by research assistants. Specifically, the AOI in the scene (see Figs. 2 and 3) where every fixation fell on was identified manually. This was performed with the use of Tobii Pro Lab, a software that was provided by the manufacturer of the eye-tracking glasses. During the coding process, two windows were displayed on the user



Region	No.	Area of interest	Function
A	1	Observational Area	For performing the pulley manipulations and observing their outcome.
A	2	Fixed-pulley Lift Marks Area	For taking the measurement on how far the weight attached to the fixed pulley is pulled down from the above.
A	3	Movable-pulley Lift Marks Area	For taking the measurement on how far the weight attached to the movable pulley is pulled up from the ground.
A	4	Material Area	For obtaining the materials needed to build a pulley system.
B	1	Diagram Area	A diagram showing how to build a pulley system on the metal rack using pulley wheels, ropes, and test weights (hands-on experiment), or a diagram showing the virtual experiment interface with Mandarin translations to the items in English (simulated experiment).
B	2	Equation Area	Mathematical equations of work and mechanical advantages are presented for easy referencing.
B	3	Reminder Area	It contained a message on how the weights should be added to the pulley system, physically or virtually.
C	1	Record Area	A table for recording relevant values observed from the manipulations and related parameters, such as work and mechanical advantage, derived from the observed values.
C	2	Question Area	A list of two-choice or short-answer questions about the experiment and related concepts.

interface. The first window played back the video that was recorded by the glasses, which showed the location of eye fixations that were dynamically marked in the video. The coder controlled the video playback, manually clicked on the static picture in the second window to record the location of the eye fixations, and repeated the same procedure for all the fixations in the video. Repeated mapping was not required when a series of fixations were maintained at the same location.

As the static picture and video playback of the experiment were launched simultaneously, temporal information (i.e., the start and end time of a fixation) was automatically obtained through manual mapping. The coordinates of the nine AOIs were also entered into the software to derive the fixation data for every one of them. Since the focus of this study was on eye movements that were defined by AOIs, the data quality was not affected by slight spatial errors in manual mapping, except for cross-boundary errors. The coding was performed by four research assistants, and every one of them began by working on the video playbacks of the hands-on and simulation groups separately. They then discussed and resolved differences in judgment to establish common principles. After the discussion, every assistant proceeded to work on the video playbacks that were allocated to them. In order to ensure the quality of coding, spot checks were made on them by the senior research assistant.

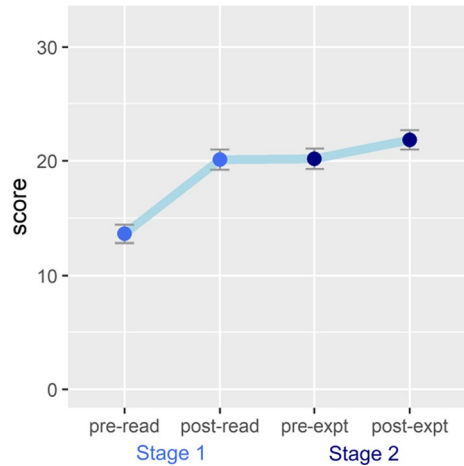
Prior knowledge about pulley mechanics with regards to experimentation was defined in terms of the post-reading test scores, such that those who ranked the upper third, middle third, and bottom third were labeled as high ($n=34$), middle ($n=35$), and low ($n=36$) in prior knowledge respectively. For each group, the number of participants were approximately the same in both experiment modes.

Results

Statistical analyses were conducted on R (version 4.0.4; R Core Team, 2013). The fitting of linear mixed-effects models was performed using the lme4 package (version 1.1–26; Bates et al., 2015). Confidence intervals, calculated by the basic bootstrap method and based on 1000 iterations, were computed using the boot package (version 1.3–28; Canty & Ripley 2021). Post-hoc pairwise comparisons were conducted with the multcomp (version 1.4–16; Hothorn et al., 2008) and emmeans packages (version 1.5.5-1; Lenth 2021).

The conceptual understanding of pulley mechanics (learning outcome) was improved by reading and experimentation, a one-way repeated-measures ANOVA was conducted on the scores for pre-reading ($M=13.63$, $SD=4.13$, bootstrapped 95% CI = [12.77, 14.41]), post-reading ($M=20.10$, $SD=4.56$, bootstrapped 95% CI = [19.19, 20.97]), pre-experiment ($M=20.18$, $SD=4.59$, bootstrapped 95% CI = [19.26, 21.08]), and post-experiment ($M=21.81$, $SD=4.37$, bootstrapped 95% CI = [21.01, 22.68]) pulley tests (Fig. 4). There were significant differences between these tests, $F(2.62, 272.52)=247.33$, $p<.001$, $\eta^2_p=0.70$; we used Greenhouse-Geisser correction ($\epsilon=0.87$) to adjust for sphericity violation. Moreover, all but one post-hoc Tukey comparisons were significant, $ps<0.001$. The exception was the pairwise comparison between post-reading and pre-experiment tests, $p=.995$; still, this was an expected exception, as no intervention occurred between the two tests. Given that the pulley test was repeatedly administered, this non-significant difference could allay concerns about potential retest effects. Focusing on the experimentation stage,

Fig. 4 Mean scores of the pulley test at the pre-reading, post-reading, pre-experiment, and post-experiment time points ($N=105$). The max score was 32. Error bars represent 95% confidence intervals of the mean



upward trends in pulley test scores were observed across prior knowledge groups (high, middle, and low).

Effects of prior knowledge and experiment mode on the learning outcome of experimentation

To answer Research Question 1, whether the learning outcome of experimentation is affected by prior knowledge and experiment mode, a series of linear mixed-effects models were fit on the data, using the restricted maximum likelihood (REML) method; while controlling for spatial ability, we specified Experiment Mode (hands-on vs. computer-simulated) and Prior Knowledge (high, middle, and low) as between-subjects factors, and Test (pre- vs. post-experiment) as a within-subjects factor for the fixed effects of the model, and subjects as the random effect (Fig. 5). A total of seven models were fitted, one for the overall test score and the rest for each of the six pulley subtopics, along with their corresponding random-effects model. Likelihood-ratio tests that compared the goodness of fit of the full model and the random-effects model revealed superiority of five out of seven full models, which includes the model on the overall test scores, and the models on lever, work, mechanical advantage, and the basic principles of pulley system, $ps < 0.05$, with the p-values adjusted using the false discovery rate approach (Benjamini & Hochberg, 1995). However, the full models on the subtopics of friction and conservation of energy did not provide a better data fit than the random-effect-only model, $ps > 0.05$. In other words, for the subtopics of friction and conservation of energy, the specification of the factors did not provide satisfactory account of the variance in the test scores. Thus, no further analyses were carried out on these two subtopics. On violations of assumptions, Shapiro test on the residuals of the model on total test scores returned no significant deviation from normality ($W=0.99$, $p=.18$); however, on the residuals of other subtopic models, normality tests returned significant results [Lever: $W=0.95$, $p < .001$; Work: $W=0.99$, $p=.03$; Mechanical advantage: $W=0.97$, $p < .001$; Pulley system: $W=0.98$, $p=.002$]. All the p-values reported on fixed effects and their interactions below had been adjusted for multiple comparisons using the false discovery rate approach.

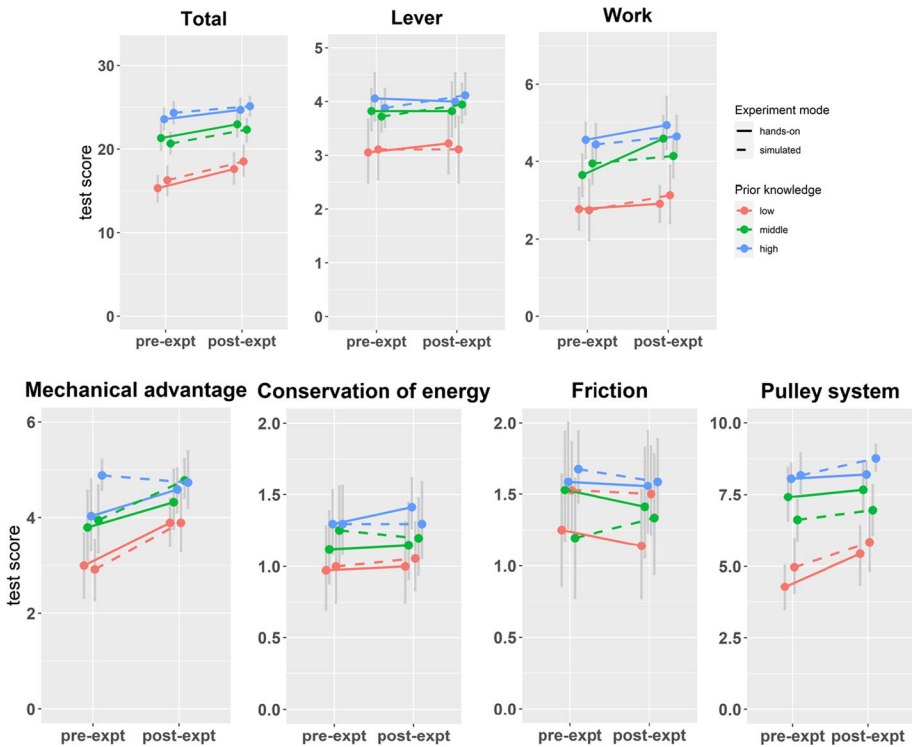


Fig. 5 Mean total scores and mean scores for each of the six pulley subtopics for the pre-experiment and post-experiment tests by prior knowledge and experiment mode. Prior knowledge was divided into high, middle and low, defined by the post-reading scores. Experiment mode was divided into hands-on vs. simulated modes. Error bars represent bootstrapped 95% confident intervals of the mean

On the total test scores, the linear mixed-effects model returned significant fixed effects for Prior Knowledge [$F(2,99)=58.22$, $p<.001$; $\eta_p^2=0.54$] and Test [$F(1,99)=40.04$, $p<.001$, $\eta_p^2=0.29$]; no interaction between them was found. Post-hoc Tukey tests showed that post-experiment test scores were higher than pre-experiment test scores, $p<.001$. Improvement on the conceptual understanding of pulley mechanics was observed. On the fixed effects of Prior Knowledge, high prior knowledge group attained the highest mean scores, followed by the middle and then low prior knowledge groups ($ps<0.001$).

Separate linear mixed-effects models were then conducted on test scores for four of the six subtopics of pulley mechanics that showed satisfactory goodness of fit; all models had Prior Knowledge and Test as predictors. Prior Knowledge returned significant for all these subtopics [lever: $F(2,99)=8.23$, $p=.003$, $\eta_p^2=0.14$; work: $F(2,99)=23.16$, $p<.001$, $\eta_p^2=0.32$; basic principles of pulley system: $F(2,99)=37.75$, $p<.001$, $\eta_p^2=0.43$; mechanical advantage: $F(2,99)=9.46$, $p<.001$, $\eta_p^2=0.16$]; Test, however, was significant in all the models except for lever [work: $F(1,99)=8.92$, $p=.01$, $\eta_p^2=0.08$; basic principles of pulley system: $F(1,99)=13.67$, $p=.001$, $\eta_p^2=0.12$; mechanical advantage: $F(1,99)=24.44$, $p<.001$, $\eta_p^2=0.20$]. In none of the models above was the interaction between Prior Knowledge and Test significant, $ps>0.05$.

Overall, where post-experiment test scores were found to have been improved, prior knowledge also tended to significantly boost test performance as well.

The influence of prior knowledge and experiment mode on fixation durations and pupil size

To answer Research Question 2, whether fixation durations and pupil size for individual workstation AOIs were affected by prior knowledge and experiment mode, we

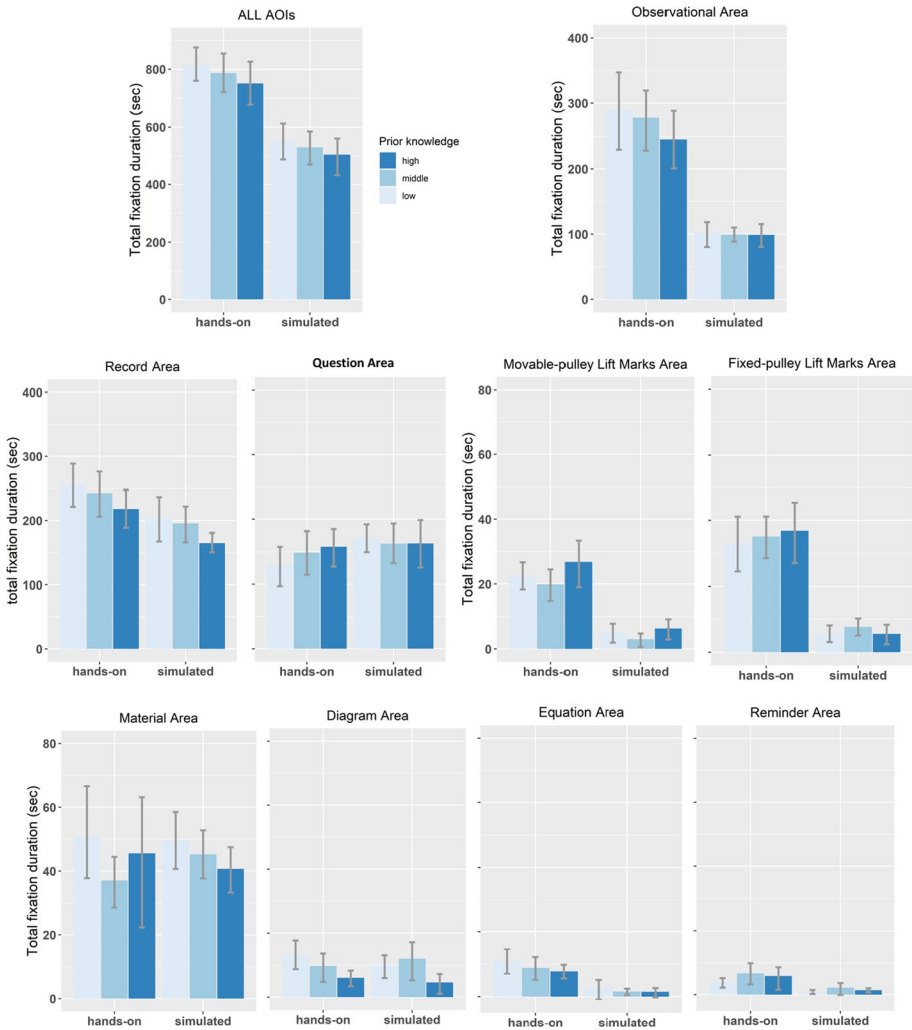


Fig. 6 Mean total fixation duration (in sec) on workstation AOIs (combined and the nine base-level AOIs) by Experiment Mode (hands-on: left vs. simulated: right) and Prior Knowledge (high: darkest; middle; low: lightest). Y-axes without values share the same scale as the graph to the left on the same row. Error bars represent bootstrapped 95% confidence intervals of the mean

(i) Hands-on experiments

(i) Hands-on experiments

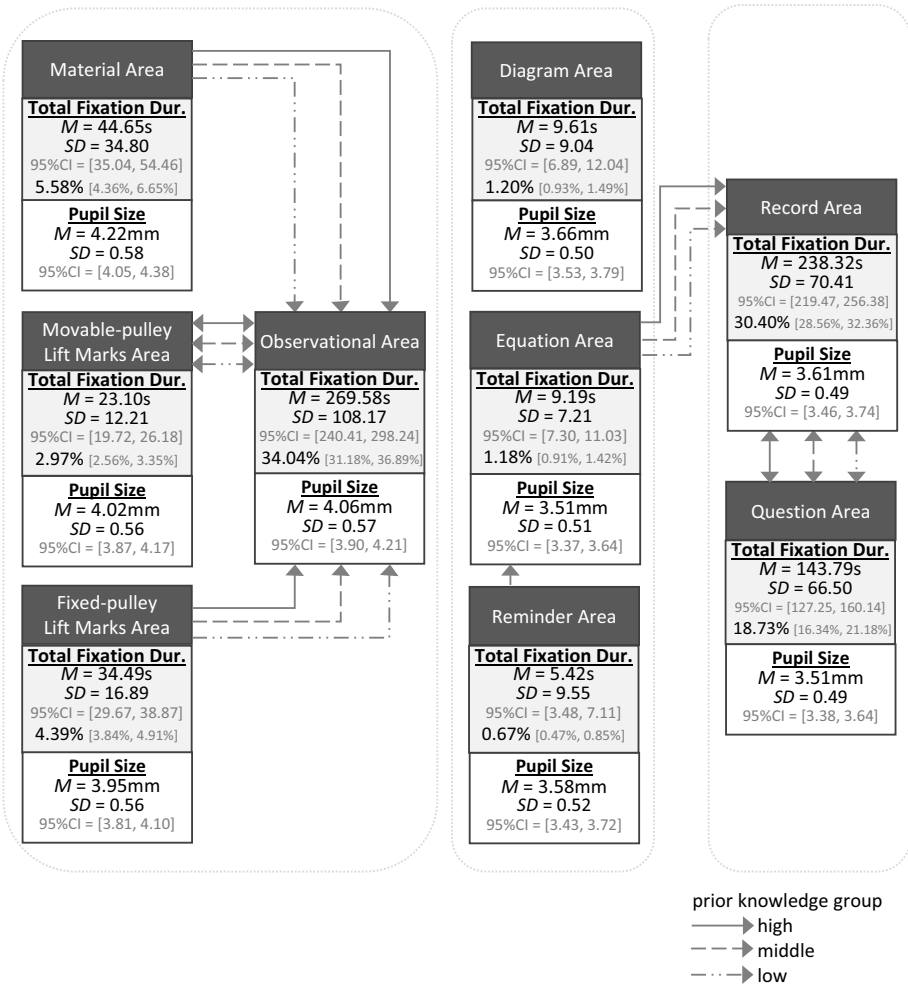


Fig. 7 AOIs used in this study. The arrows indicate the directions of significant aggregate transitions. The line on the arrow represents prior knowledge group (high/ middle/ low). Three paths among various prior knowledge groups were not shown due to low number of aggregate transitions (< 15). The total fixation duration and average pupil size for each workstation area and significant aggregate transitions ($p < 0.05$) between these areas in a (i) hands-on and (ii) simulated experiment. Under each area label, the mean, standard deviation and bootstrapped 95% confidence interval of total fixation duration and average pupil size are shown. For total fixation duration, the mean proportion, in percentage, of fixation durations out of the grand total fixation durations covering the entire workstation, is also presented

conducted a series of 2 (Experiment Mode: hands-on vs. simulated) X 3 (Prior Knowledge: high, middle, and low) factorial ANOVAs on the total duration of fixations and average pupil size that fell on each individual AOI (Figs. 6 and 7).

(ii) Simulated experiments

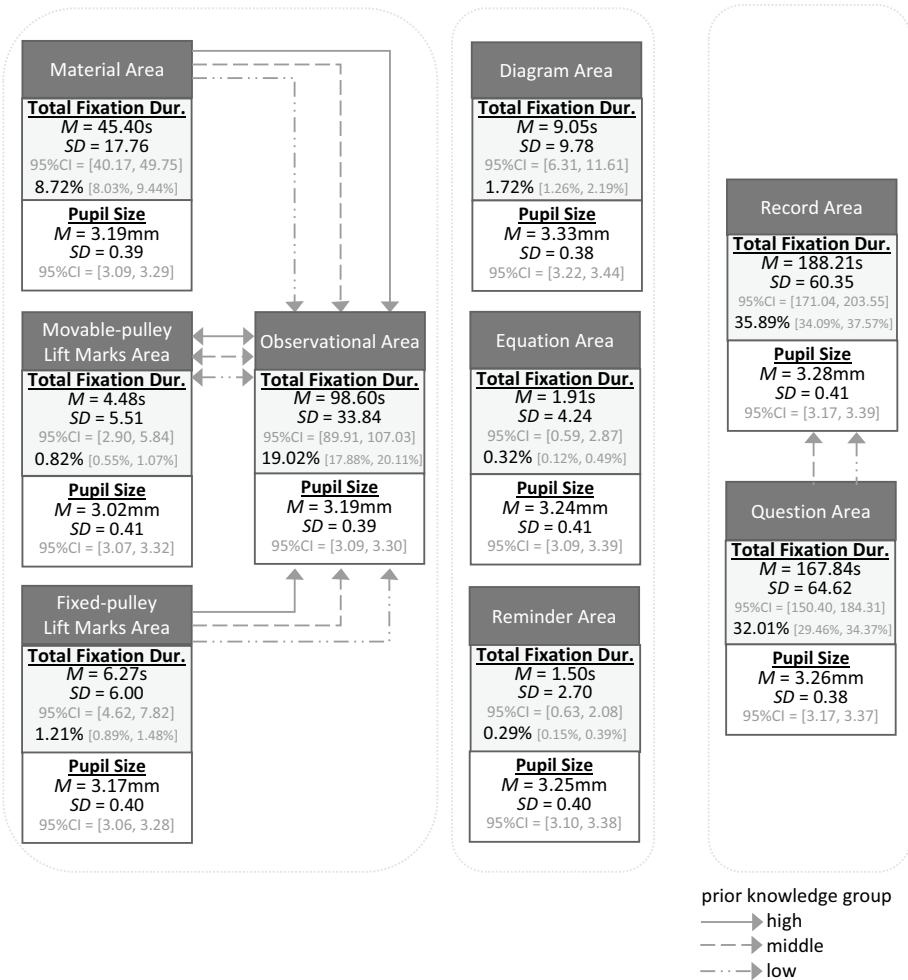


Fig. 7 (continued)

Total fixation durations

On the total fixation durations for **all the AOIs combined**, Experiment Mode was significant [$F(1,99) = 88.22, p < .001, \eta^2_p = 0.47$], but not Prior Knowledge [$F(2,99) = 1.40, p = .38, \eta^2_p = 0.03$], or the interaction between Experiment Mode and Prior Knowledge [$F(1,99) = 0.03, p = .97, \eta^2_p < 0.001$]. Post-hoc Tukey test suggested that simulated experiments ($M = 784.06, SD = 133.85$) gave rise to shorter total fixations than hands-on experiments ($M = 528.95, SD = 142.91$), $p < .001$. Regarding those **AOIs where pulley manipulations** took place, Experiment Mode was a significant factor influencing total fixation durations on the Observational Area [$F(1,99) = 116.78, p < .001, \eta^2_p = 0.54$], the Fixed-pulley Lift Marks Area [$F(1,99) = 125.24, p < .001, \eta^2_p = 0.56$], and the Movable-pulley Lift Marks Area [$F(1,99) = 102.10, p < .001, \eta^2_p = 0.51$]. Neither Prior Knowledge nor its

interaction with Experiment Mode was significant, $ps > 0.05$. Post-hoc Tukey test suggested that shorter total fixations were associated with simulated experiments, $ps < 0.001$. For the Material Area, none of the factors was significant, $ps > 0.05$. **On the guideline sheet**, divergent results were found on subordinate AOIs. For both the Equation Area and the Reminder Area, only Experiment Mode was found to significantly affect fixation durations [Equation Area: $F(1,99) = 39.55$, $p < .001$, $\eta^2_p = 0.29$; Reminder Area: $F(1,98) = 15.70$, $p < .001$, $\eta^2_p = 0.14$]. Post-hoc Tukey tests showed that simulated experiments were associated with shorter fixations ($ps < 0.001$). In contrast, for the Diagram Area, Prior Knowledge was instead the only significant factor [$F(2,99) = 4.94$, $p = .009$, $\eta^2_p = 0.09$]. Post-hoc Tukey tests suggested that the high prior knowledge group showed shorter total fixation durations in the Diagram Area than the middle group ($p = .03$) and the low groups ($p = .01$). **On the lab report worksheet**, for the Record Area, Experiment Mode [$F(1,99) = 15.74$, $p < .001$, $\eta^2_p = 0.14$] was the only significant factor. Post-hoc Tukey tests suggested that simulated experiment was associated with shorter fixations ($p < .001$). As for the Question Area, none of the factors reached statistical significance ($ps > 0.05$).

Pupil size

The results showed that experiment mode was a significant factor that affected pupil size [$F(1,99) = 44.28$, $p < .001$, $\eta^2_p = 0.31$], but not prior knowledge or the interaction between both factors ($p > .05$). Post-hoc Tukey test results suggested that the pupil was more dilated during a hands-on experiment than in a simulated experiment, $p < .001$, Cohen's $d = 0.33$. In addition, pairwise comparisons between both experiment modes were conducted on individual workstation area. After adjusting for multiple comparisons, the average pupil size was found to be consistently larger across all workstation areas in the hands-on group ($p < .05$). A summary of the statistical findings on average pupil size for individual workstation area in the hands-on and simulated experiments are shown in Fig. 7. The salient findings on total test scores, fixation durations, and pupil size are summarized in Table 1.

Lag sequential analysis on fixation transitions and entropy analysis on the distribution

To address Research Question 3, lag sequential analysis on fixation transitions and entropy analysis on the distribution were conducted.

Lag sequential analysis on fixation transitions

On average, a hands-on experiment gave rise to 309.96 cross-AOI transitions ($SD = 89.18$, bootstrapped 95% CI = [286.36, 334.55]), while a simulated experiment produced 143.48 transitions ($SD = 39.66$, bootstrapped 95% CI = [132.23, 153.65]). The difference was partly driven by the stark contrast between the two experiment modes in terms of overall duration. Since our focus was on comparing the distribution of fixation transitions among all the possible transition paths, aggregate transition count was converted to Z scores (see Appendices 4 and 5). Paths with significant fixation transitions recorded (i.e., Z scores larger than 1.96, with an alpha level of 0.05) were shown in Fig. 7; these significant transition paths were taken in exceptionally high number. Three of these paths were common to both experiment modes and to all levels of prior knowledge (high, middle, and low); these paths originated from the

Table 1 A summary table on the findings on how prior knowledge affected the test scores on individual pulley subtopics and how prior knowledge and experiment mode affected the total fixation durations and pupil size for individual workstation areas

DV	Pulley subtopic/ Area of interest	Significant factor	Pairwise comparison
Test scores	Lever, work, pulley system, mechanical advantage	Prior knowledge	High > <u>middle</u> > <u>low</u>
	All subtopics combined	Prior knowledge	High > middle > low
Total fixation durations	<i>Guideline sheet</i>		
	Diagram Area	Prior Knowledge	High < <u>middle</u> < <u>low</u>
	Equation Area, Reminder Area	Experiment Mode	Simulated < hands-on
	<i>Pulley manipulation regions</i>		
	Fixed-pulley Lift Marks Area, Movable-pulley Lift Marks Area, Observational Area	Experiment Mode	Simulated < hands-on
	<i>Lab report worksheet</i>		
	Record Area	Experiment Mode	Simulated < hands-on
Pupil size	All areas combined	Experiment Mode	Simulated < hands-on
	All areas combined	Experiment Mode	Simulated < hands-on
	All of the nine workstation areas		

Trends that are underlined did not reach statistical significance

Mean post-test scores were higher than mean pre-test scores for the subtopics of work, basic principles of pulley system, mechanical advantage, and all the subtopics combined (i.e., total test scores); the subtopics of friction and conservation of energy are not presented due to poor model fit; with regards to total fixation durations, Material Area, one of the regions where pulley manipulation took place, and Question Area, within the lab report worksheet, are not presented above as neither factor was significant

Material Area, Fixed-pulley Lift Marks Area, or Movable-pulley Lift Marks Area, and were all bound for the Observational Area. Three other paths were only significant in hands-on experiments and all the prior knowledge groups in that experiment mode: They were the paths from the Observational Area to the Movable-pulley Lift Marks Area, from the Equation Area to the Record Area, and from the Record Area to the Question Area. The path from the Question Area to the Record Area was significant in both experiment modes except for the high knowledge group in simulated mode. As for paths that were significant only in either one of the experiment modes, the path from the Observational Area to the Material Area was significant only in simulated experiments, covering all prior knowledge groups; the path from Reminder Area to the Equation Area was significant only in hands-on experiments and, specifically, for the high prior knowledge group only. Other than these paths, three statistically significant paths had been disregarded in hands-on experiments, as the results were skewed by a few outliers among the overall small count of transitions. For a bigger picture of the significant paths, considering the three larger regions (i.e., pulley manipulation, guideline, and lab report) within which the AOIs were located, only the transition path from the Equation Area to the Record Area (significant in hands-on experiments only) was cross-regional; other paths did not involve movements beyond their respective region.

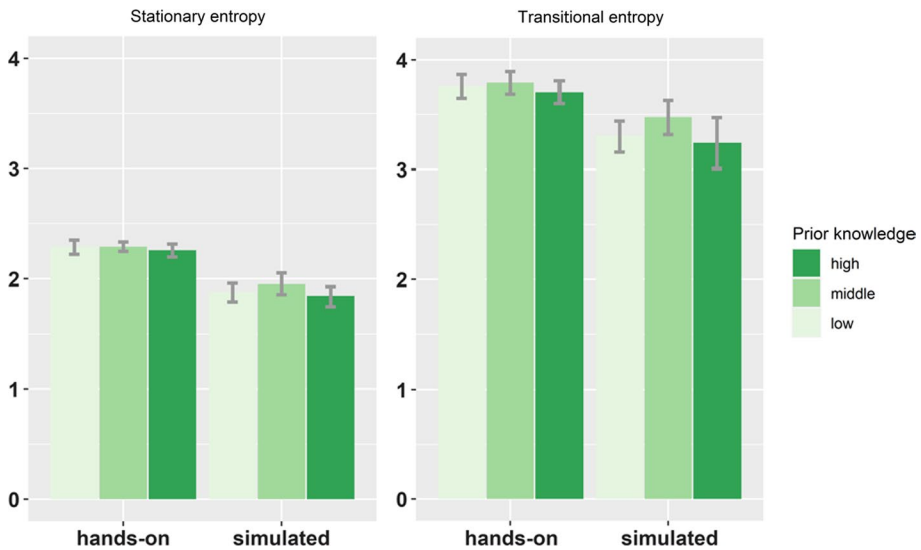


Fig. 8 Mean stationary entropy (left) and mean transitional entropy (right) by experiment mode and prior knowledge. Error bars represent bootstrapped 95% confidence interval of the mean

Entropy analysis on the distribution of eye fixations and fixation transitions

On the entropy analysis, the results of Fig. 8 showed that only the main effect of Experiment Mode was significant, $F(1, 99) = 144.31$, $p < .001$, $\eta^2_p = 0.59$, suggesting that the mean stationary entropy was higher for hands-on experiments than simulated experiments; the main effect of Prior Knowledge [$F(2, 99) = 1.51$, $p = .23$] and the interaction between the two factors [$F(2, 99) = 0.59$, $p = .56$] were not significant. On the simple main effects, pairwise comparisons showed that all differences between the two experiment modes were significant within each of the three prior knowledge groups [High: $t(99) = 7.31$, $p < .001$, Cohen's $d = 2.21$; Middle: $t(99) = 6.09$, $p < .001$, Cohen's $d = 2.01$; Low: $t(99) = 7.50$, $p < .001$, Cohen's $d = 2.50$], with p-values adjusted by Tukey's method. Regarding the assessment of ANOVA assumption violations, Shapiro test on the residuals returned no significant deviation from normality [$W = 0.98$, $p = .051$]; Levene's test on homoscedasticity also returned non-significant [$F(5, 99) = 2.05$, $p = .08$].

On transitional entropy (Fig. 8), similar to the results on stationary entropy, only the main effect of Experiment Mode was found significant [$F(1, 99) = 43.41$, $p < .001$, $\eta^2_p = 0.30$], suggesting that the transitional entropy was higher (i.e., more dispersed destinations) for hands-on experiments than for simulated experiments, while the main effect of Prior Knowledge [$F(2, 99) = 2.33$, $p = .10$] and the interaction between Experiment Mode and Prior Knowledge [$F(2, 99) = 0.58$, $p = .56$] were both non-significant. On the simple main effects, all the differences between hands-on and simulated experiments within each of the three prior knowledge groups were significant [High: $t(99) = 4.28$, $p < .001$, Cohen's $d = 1.47$; Middle: $t(99) = 2.97$, $p = .04$, Cohen's $d = 1.00$; Low: $t(99) = 4.29$, $p < .001$, Cohen's $d = 1.43$], with p-values adjusted using Tukey's approach. On the assessment of ANOVA assumption violations, Shapiro test on the residuals revealed no significant deviation from normality [$W = 0.99$, $p = .70$]; Levene's test on homoscedasticity, however, returned significant [$F(5, 99) = 2.40$, $p = .04$].

Discussion

This study investigated the learning performances and the eye movements during a computer-simulated versus hands-on science experiment, from which the learning process could be inferred, and how the eye movements were affected by prior knowledge and experiment mode. In the following sections, we discussed the three main findings that addressed our research questions.

Experiment mode and prior knowledge on conceptual understanding

For answering the first research question, is the learning outcome of experimentation affected by prior knowledge and experiment mode (hands-on versus simulated), we found that there had no significant difference between hands-on and simulated experiments in terms of learning outcome, consistent with some of the prior studies on the comparison between the two (e.g., Dalgarno et al., 2009; Han, 2013; Triona & Klahr 2003; Zacharia & Olympiou, 2011). No difference was found in the pulley subtopics either. Based on our findings, both experiment modes were effective in improving conceptual understanding about pulley mechanics, as a significant improvement in result was seen in the post-experiment test. Specifically, performing either a hands-on or simulated experiment, the process of which includes the reflections involved in the completion of the lab report, led to a better understanding of subtopics of pulley mechanics including work, mechanical advantage, and the basic principles of pulley mechanics. The results of this study also corroborated the finding by earlier studies on an advantage of simulated experiments, which is their helpfulness in assisting learners to integrate science knowledge, graphs, processes, and phenomena through the provision of dynamic visualizations and analogies for abstract and unobservable phenomena (Chien et al., 2015; Olympiou & Zacharia, 2012; Zacharia & Olympiou 2011). Moreover, our study found that learners with more prior knowledge had better results in the post-experiment test than those who had less prior knowledge, which supported the findings by previous research that such knowledge can influence the achievements of students in science learning, even after controlling for their motivations and learning strategies (Song et al., 2016; van Riesen et al., 2018; Wen et al., 2020).

Experiment mode and prior knowledge on eye movements

To address our second research question on whether eye-movement indicators (e.g., fixation durations and pupil size) were affected by prior knowledge and experiment mode in manipulating experiments and worksheets, our study found that only experiment mode had an effect on eye movements. The comparative results between both experiment modes showed consistently shorter total fixation durations and smaller average pupil size in most workstation regions (except the Material Area) in the simulation group, including the space where pulley manipulations were performed. Interestingly, compared to the eye-movement results of Chien et al. (2015)—which showed that the simulation group had more trials and paid more attention to the experiment while the physical-laboratories group spent more time on worksheets—our results showed the opposite pattern: participants of simulated experiments in the present study did not appear to exploit the ease of experimental manipulations on the computer simulations and to try out more varied manipulations. Moreover, the results of this study showed that the simulation group had smaller pupil size in most areas in the worksheet than the hands-on group. Since fixation durations (Jian, 2019, 2021;

Mason et al., 2013; Miller, 2015; Wu & Liu, 2021) and pupil size (Alemdag & Cagiltay, 2018; van Gog & Jarodzka, 2013; Van Gog et al., 2009) are used to measure visual attention and cognitive load, we inferred that one possible explanation was that the learners in the hands-on group experienced a higher cognitive load than those in the simulation group based on the cognitive load theory (Plass et al., 2010), which states that additional demands for tactile information and kinesthetic sequences may increase cognitive load. Another possible explanation was that the specific features of the experimental setups explain differences between the two modalities, not the fundamental nature of the hands-on and virtual environments themselves (Bumbacher et al., 2018). For example, we make the same argument—the equation sheet plays a greater role in the hands-on condition because the simulation computes the work done for participants. In contrast, the hands-on participants must calculate it independently. Therefore, a part of the greater fixation time for the hands-on groups may be attributed to the increased time needed for physically changing the equipment, rather than the increased cognitive load in processing the outcome of the experimental trial. Further empirical studies were needed to confirm the possible explanations.

Fixation transitions and experiment mode

To address our last research question on whether cross-regional fixation transitions were affected by prior knowledge and experiment mode, we conducted sequential and entropy analyses to examine the cross-AOI fixation transitions from regional and macro perspectives respectively, offering multi-faceted insights to the two experiment modes. Starting from a macro perspective on the overall distribution of fixation transitions among AOIs, the transitional entropies revealed that participants in hands-on experiments tended to make more diversified fixation transitions across AOIs, while those in simulated experiments showed a higher level of concentration in the spatial pattern of fixation transitions. This reflected the tendency of concentrating on a smaller portion of the workstation in simulated experiments. The more focused pattern of fixation transitions could partially explain the higher efficiency of virtual experiments and the minimal occurrence of trial and error. Beyond experiment mode, the level of prior knowledge was unlikely to influence the distribution of fixation transitions.

On the regional level, despite a rather high degree of overlapping between hands-on and simulated experiments in terms of prominent fixation transition routes, a few differences between the two experiment modes are noteworthy. First, the transition route from the Record Area to the Question Area, both of which are located within the lab report, is exceptionally salient in hands-on experiments but not in simulated experiments. Interestingly, the transition route in reverse (i.e., from the Question Area to the Record Area) was also significant among all but one prior knowledge groups for hands-on and simulated experiments, with the exception being the high prior knowledge group for simulated experiments. Salient bi-directional routes in this case indicate the need to integrate information from the figures recorded during the experiment and the concepts raised in the lab report questions in order to derive answers to those questions. The salient bi-directional routes between the Record Area and the Question Area in the hands-on experiment suggested that, when participants of the hands-on mode were working on the lab report, they tended to focus on the lab report and attempt the experimental questions on the report based mostly on the information from the experiment record. In contrast, participants of simulated experiments tended to make transitions from the Record Area towards not only the Question Area but also the Observational Area, instead of concentrating on making transitions to the Question

Area. In other words, more varied references were made to complete the lab report questions for those performing simulated experiments. This is particularly the case in the high prior knowledge group for simulated experiments, where transitions out of the Question Area were much more evenly distributed than the middle and low prior knowledge groups. One possible explanation is that the participants with high prior knowledge about pulley mechanics in the simulation group could answer the lab report questions without frequently referring to the experiment record because they had acquired a comprehensive understanding before and during the experiment. It is noteworthy that such an advantage in prior knowledge is absent in hands-on experiments because of the relative cognitive overload resulting from manipulating the abundant physical objects for all participants regardless of their prior knowledge levels. Some other differences in significant transition routes between the two experiment modes could be attributed to the fundamental difference in the software design. For example, the hands-on experiment required participants to calculate the values of work on their own, and they must first check the equations in the Equation Area and then transferred to the Record area. In other words, the software embedded in the simulation could automatically display and calculate the required values in the lab report, and this might free a considerable amount of working memory capacity and processing time for the simulation group to engage in integrating the results of simulation experiments with relevant knowledge to achieve a deeper understanding of the pulley mechanics.

Research contributions and limitations

A contribution of this study was its methodology. Although a few studies on eye movements used sequential analysis to investigate learning transition patterns in reading (Jian, 2016; Jian & Ko, 2017; Tsai et al., 2022) and in manipulating experiments (Jian, 2022a; Chiou et al., 2022), the use of entropy analysis to examine the overall distribution of transitions from a macro perspective is, as far as we know, the first of its kind in this field of research. As expected, sequential and entropy analyses provide different perspectives to the pattern of fixation durations. In the current study, LSA offered insights into the differences in transitions on the AOI-level between the two experiment modes and, in several occasions, among levels of prior knowledge. Notwithstanding that it is possible to use LSA to examine the similarities and differences of cross-regional fixation transitions among individuals, its strength lies in revealing the spatial pattern of transitions on a group level, as it is typically deployed in eye-tracking studies. Meanwhile, entropy analysis offers a more straightforward overview in the overall distribution of fixation transitions, and, as we have demonstrated, allows conventional statistical modeling that examines group differences while taking individual differences into account. It is noteworthy that entropy analysis is not necessarily restricted to the overall distribution of fixation transitions; rather, entropy could be derived from the distribution of transitions towards destination AOIs for each individual source AOI. However, this was not attempted for the present study, as the LSA has proven its merit in identifying important transition routes, which could not be delivered by entropy analysis. Depending on the research questions, a dual-method approach in examining transition patterns is desirable. We propose that such an approach could be applied to the analysis of learning process in other virtual and hands-on learning activities.

Another contribution of this study are the implications of its findings in science education. To the best of our knowledge, the use of an eye tracker to examine the learning process that is involved in the conduct of scientific experiments is rarely used (e.g., Jian 2022a; Chien et al., 2015). This study aimed to examine whether—and how—the learning

outcome and learning process of experimentation in science education were affected by the mode of experiment and level of prior knowledge. Our results on the learning outcome of experimentation concurred with previous findings that prior knowledge is a robust factor in shaping learning outcome, and that hands-on and computer-simulated experiments could bring about similar learning outcome. Beyond learning outcome, the most consistent difference between the two experiment modes was found in total fixation durations and pupil size in most workstation regions, where shorter fixation durations were observed in simulated experiments. Given that there is no difference in terms of learning outcome between the two experiment modes, computer simulations offered higher learning efficiency across the board, which validated the findings of previous research on the advantage of time efficiency that is provided by virtual experiments (Gibbons et al., 2004). Computer technology allows for virtual forms of scientific inquiry with the benefits of increased availability, higher flexibility, lower cost, precise manipulations, and support for self-paced learning (Chen, 2010; del Mar Quiroga & Choate, 2019; Hofstein & Lunetta 2004), virtual experiments have become even more widely adopted in educational practice, especially during the COVID-19 pandemic (Klein et al., 2021; Radhamani et al., 2021). Our findings show that mode of experiment took the place of prior knowledge as the factor that shaped the learning process. Specifically, computer-simulated experiments showed a more even distribution of fixation transitions among both source AOIs and destination AOIs. Note that the difference is not restricted to the regions where actual manipulations occurred, where operations were understandably different, but could also be spotted in such directly comparable areas as in the lab report. The important findings in this study were that the total fixation durations, pupil size and fixation transition analyses revealed that experiment mode affected the completion of the Question Area in different ways: Where total fixation durations and pupil size showed no difference between the two experiment modes, fixation transition analysis was able to detect subtle difference in the transition routes between the two, which informs their distinctions in the learning process, especially in terms of how the various visual and verbal information, from the perspective of the Cognitive Theory of Multimedia Learning (Mayer, 2009), is integrated to build new knowledge about pulley mechanics. Fixation transition analysis appears to be a promising method of examining learning process.

Finally, this study allows for drawing some pedagogical implications for science education.

First, simulated experiments could be a more time-efficient and, based on our findings of eye fixations, probably a less cognitively-taxing option of guided inquiry, although they are as effective as their hands-on counterpart in delivering learning outcome. Also, although there is no denying that prior knowledge gives learners a competitive edge to the learning outcome of experimentation, regardless whether it was hands-on or computer-simulated, the learning process as demonstrated by our eye-movement analyses points to a leveling-up effect for learners of various prior knowledge background. Moreover, regarding the manipulation of virtual simulations, our analyses have shown the tremendous potential for using fixation transitions between functional regions to uncover the learning process involved in manipulating simulations. Teachers can monitor students' learning progress by examining the patterns of their fixation transitions. For example, if a student fixates on the Material area without transferring to the Observational area, then it can be assumed that the student may encounter difficulty in connecting the changing variables with the resulting phenomena in the pulley system. In addition, if a student pays too much attention to the Report area and frequently transfers between the Equation area and Report area, the teacher can assume that the student may not know the simulation software has automatically calculated

the required values for the lab report, and then guide the student to focus on the numbers shown on the simulation. By doing this, the student could free a certain amount of working memory capacity to engage in higher-order cognitive processes, such as analyzing the trends and patterns of obtained data instead of arithmetically calculating. In other words, as suggested by Chiou et al. (2022), to successfully learn with scientific simulations, learners not only need to perform fixation transitions in logical sequences but also integrate the obtained information as a meaningful whole. Accordingly, analyses of fixation transitions may provide feedback to both educators and students for examining the learning processes involved in manipulating simulations and for designing instructional activities to guide students on how to manipulate the simulations, observe target phenomena, take measures, and analyze the data in a logical sequence. Furthermore, science teachers with no access to eye-tracking systems at hand can also benefit from the findings of this study. By knowing students' potential fixation transition patterns before actual instruction, science teachers can design instructional activities that can either guide students to perform expected fixation transitions or prevent them from undertaking undesired fixation transitions.

Some limitations were observed in this study. First, the pulley test was repeatedly administered, so retest effect cannot be ruled out; however, this would not affect the between-subjects comparison of experiment modes. Second, since eye fixations reflect visual attention—and to a lesser extent, cognitive processing—one cannot assert, with absolute confidence, that they entail active processing of visual objects. Thus, an element of uncertainty is involved when an inference of cognitive load is made based on fixation durations and pupil size. Third, the development of conceptual understanding from experimentation and guided inquiry could depend on the subject matter (Pouw et al., 2014) and its connection with the actual inquiry activities; hence, our study findings should be generalized with caution. Finally, we tried to investigate if the participants' learning processes (reflected by the eye-movement indicators) were relevant to their test performance, but we found neither a stable correlation of the learning processes and test performance for the hands-on group nor for the simulation group. We thought it might be because the sample size (about 50 participants for the hands-on and the simulation groups) was not large enough for conducting correlation analysis or predictive analysis.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11251-023-09634-8>.

Acknowledgements This research was financially supported from the grant MOST 111-2636-H-003-009-under Young Scholar Fellowship Program by Ministry of Science and Technology in Taiwan, NSTC 111-2410-H-003 -013-MY3 by National Science and Technology Council, and the “Institute for Research Excellence in Learning Sciences” and “Higher Education Deep Cultivation Project” of National Taiwan Normal University (NTNU), sponsored by the Ministry of Education, Taiwan. We also thank for Pro. Puntambekar agreeing us to use their test items of CoMPASS, and Miss Chan-Mam Yu for collecting the data in this study.

References

- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, 125, 413–428. <https://doi.org/10.1016/j.compedu.2018.06.023>.
- Appel, T., Sevchenko, N., Wortha, F., Tsarava, K., Moeller, K., Ninaus, M., & Gerjets, P. (2019). Predicting cognitive load in an emergency simulation based on behavioral and physiological measures. *In 2019 International Conference on Multimodal Interaction* (pp. 154–163).

- Baltzis, K., & Koukias, K. D. (2009). Using laboratory experiments and circuit simulation IT tools in an undergraduate course in analog electronics. *Journal of Science Education and Technology*, 18, 546–555. <https://doi.org/10.1007/S10956-009-9169-Z>.
- Bates, D., Mäechler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Bodner, G. M., & Guay, R. B. (1997). The Purdue visualization of rotations test. *The Chemical Educator*, 2(4), 1–17. <https://doi.org/10.1007/s00897970138a>
- Bumbacher, E., Salehi, S., Wieman, C., & Blikstein, P. (2018). Tools for science inquiry learning: Tool affordances, experimentation strategies, and conceptual understanding. *Journal of Science Education and Technology*, 27, 215–235.
- Canty, A., & Ripley, B. (2021). *Boot: Bootstrap R (S-Plus) Functions*. R package version 1.3–28.
- Chang, K. E., Chen, Y. L., Lin, H. Y., & Sung, Y. T. (2008). Effect of learning support in simulation-based physics learning. *Computers & Education*, 51(4), 1486–1498. <https://doi.org/10.1016/j.compedu.2008.01.007>.
- Chen, S. (2010). The view of scientific inquiry conveyed by simulation-based virtual laboratories. *Computers & Education*, 55(3), 1123–1130. <https://doi.org/10.1016/j.compedu.2010.05.009>.
- Chien, K. P., Tsai, C. Y., Chen, H. L., Chang, W. H., & Chen, S. (2015). Learning differences and eye fixation patterns in virtual and physical science laboratories. *Computers and Education*, 82, 191–201. <https://doi.org/10.1016/j.compedu.2014.11.023>.
- Chini, J. J., Madsen, A., Gire, E., Rebello, N. S., & Puntambekar, S. (2012). Exploration of factors that affect the comparative effectiveness of physical and virtual manipulatives in an undergraduate laboratory. *Physical Review Special Topics – Physics Education Research*, 8, 010113. <https://doi.org/10.1103/PhysRevSTPER.8.010113>.
- Chiou, G. L., Hsu, C. Y., & Tsai, M. J. (2021). An investigation of visual and manual behaviors involved in interactions between users and physics simulation interfaces. In I. Devetak & S. Glazar (Eds.), *Applying bio-measurements methodologies in science education research* (pp. 277–294). Springer. https://doi.org/10.1007/978-3-030-71535-9_14
- Chiou, G. L., Hsu, C. Y., & Tsai, M. J. (2022). Exploring how students interact with guidance in a physics simulation: Evidence from eye-movement and log data analyses. *Interactive Learning Environments*, 30(3), 484–497. <https://doi.org/10.1080/10494820.2019.1664596>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Academic Press.
- CoMPASS. (2014). Temporary home for CoMPASS project. Concept Mapped Project-based Activity Scaffolding System. <https://github.com/cashaffer/CoMPASS>
- Dalgarno, B., Bishop, A. G., Adlong, W., & Bedgood, D. R. (2009). Effectiveness of a virtual laboratory as a preparatory resource for distance education chemistry students. *Computers & Education*, 53(3), 853–865. <https://doi.org/10.1016/j.compedu.2009.05.005>.
- del Mar Quiroga, M., & Choate, J. K. (2019). A virtual experiment improved students' understanding of physiological experimental processes ahead of a live inquiry-based practical class. *Advances in Physiology Education*, 43(4), 195–503. <https://doi.org/10.1152/advan.00050.2019>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioural, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/bf03193146>.
- Gibbons, N. J., Evans, C., Payne, A., Shah, K., & Griffin, D. K. (2004). Computer simulations improve university instructional laboratories. *Cell Biology Education*, 3(4), 263–269. <https://doi.org/10.1187/cbe.04-06-0040>.
- Han, I. (2013). Embodiment: A new perspective for evaluating physicality in learning. *Journal of Educational Computing Research*, 49(1), 41–59. <https://doi.org/10.2190/EC.49.1.b>
- Hewitt, P. G. (2018). *Conceptual Physics: The High School Physics Program* (Vol. 1) (Y.-H. Chang, K.-S. Cai, M.-R. Shih, & K.-G. Chen, Trans.). Taipei: Commonwealth Publishing.
- Hodgkiss, A., Gilligan, K. A., Tolmie, A. K., Thomas, M. S. C., & Farran, E. K. (2018). Spatial cognition and science achievement: The contribution of intrinsic and extrinsic spatial skills from 7 to 11 years. *British Journal of Educational Psychology*, 88(4), 675–697. <https://doi.org/10.1111/bjep.12211>.
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, 88(1), 28–54. <https://doi.org/10.1002/sce.10106>.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3), 346–363. <https://doi.org/10.1002/bimj.200810425>.

- Jian, Y. C. (2016). Fourth graders' cognitive processes and learning strategies for reading illustrated biology texts: Eye movement measurements. *Reading Research Quarterly*, 51(1), 93–109. <https://doi.org/10.1002/rtrq.125>.
- Jian, Y. C. (2019). Reading instructions facilitate signaling effect on science text for young readers: An eye-movement study. *International Journal of Science and Mathematics Education*, 17, 503–522. <https://doi.org/10.1007/s10763-018-9878-y>.
- Jian, Y. C. (2021). The immediate and delayed effects of text-diagram reading instruction on reading comprehension and learning processes: Evidence from eye movements. *Reading and Writing*, 34, 727–752. <https://doi.org/10.1007/s11145-020-10089-3>.
- Jian, Y. C. (2022a). Influence of science text reading difficulty and hands-on manipulation on science learning: An eye-tracking study. *Journal of Research in Science Teaching*, 59, 358–382. <https://doi.org/10.1002/tea.21731>.
- Jian, Y. C. (2022b). Reading in print versus digital media uses different cognitive strategies: Evidence from eye movements during science-text reading. *Reading and Writing*, 35, 1549–1568. <https://doi.org/10.1007/s11145-021-10246-2>.
- Jian, Y. C. (2022c). Using an eye tracker to examine the effect of prior knowledge on reading processes while reading a printed scientific text with multiple representations. *International Journal of Science Education*, 44(8), 1209–1229. <https://doi.org/10.1080/09500693.2022.2072013>.
- Jian, Y. C., & Ko, H. W. (2017). Influences of text difficulty and reading ability on learning illustrated science texts for children: An eye movement study. *Computers and Education*, 113, 263–279. <https://doi.org/10.1016/j.compedu.2017.06.002>.
- Klahr, D., Triona, L., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching*, 44(1), 183–203. <https://doi.org/10.1002/tea.20152>.
- Klein, P., Ivanjek, L., Dahlkemper, M. N., Jeličić, K., & Geyer, M. A. (2021). Studying physics during the COVID-19 pandemic: Student assessments of learning achievement, perceived effectiveness of online recitations, and online laboratories. *Physical Review Physics Education Research*, 17, 1010117. <https://doi.org/10.1103/PhysRevPhysEducRes.17.010117>.
- Kozhevnikov, M., & Thornton, R. (2006). Real-time data display, spatial visualization ability, and learning force and motion concepts. *Journal of Science Education & Technology*, 15(1), 111–132. <https://doi.org/10.1007/s10956-006-0361-0>.
- Krejtz, K., Szmidt, T., Duchowski, A. T., & Krejtz, I. (2014). Entropy-based statistical analysis of eye movement transitions. *ETRA '14: Proceedings of the Symposium on Eye Tracking Research and Applications*. (pp. 159–166). <https://doi.org/10.1145/2578153.2578176>
- Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., Perilli, F. G., Pires, A. C., Vilaro, A., & Villalobos, N. (2015). Gaze transition entropy. *ACM Transactions on Applied Perception*, 13(1), 1–20. <https://doi.org/10.1145/2834121>.
- Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., Lee, M. H., Chiou, G. L., Liang, J. C., & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, 10, 90–115. <https://doi.org/10.1016/j.edurev.2013.10.001>.
- Lazonder, A. W., Wilhelm, P., & Hagemans, M. G. (2008). The influence of domain knowledge on strategy use during simulation-based inquiry learning. *Learning and Instruction*, 18(6), 580–592. <https://doi.org/10.1016/j.learninstruc.2007.12.001>.
- Lenth, R. V. (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.5-1. <https://CRAN.R-project.org/package=emmeans>
- Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science*, 29(4), 587–625. https://doi.org/10.1207/s15516709cog0000_15.
- Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrated text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, 60, 95–109.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511811678>
- Mestad, I., & Kolstø, S. D. (2014). Using the concept of zone of proximal development to explore the challenges of and opportunities in designing discourse activities based on practical work. *Science Education*, 98(6), 1054–1076. <https://doi.org/10.1002/sce.21139>.
- Miller, B. W. (2015). Using reading times and eye-movements to measure cognitive engagement. *Educational Psychologist*, 50(1), 31–42. <https://doi.org/10.1080/00461520.2015.1004068>
- National Research Council. (2000). *Inquiry and the National science education standards: A guide for teaching and learning*. National Academy Press.


- Olympiou, G., & Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: An effort to improve students' conceptual understanding through science laboratory experimentation. *Science Education*, 96(1), 21–47. <https://doi.org/10.1002/sce.20463>.
- Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., Mnoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phase of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47–61. <https://doi.org/10.1016/j.edurev.2015.02.003>.
- Peterson, J., Pardos, Z., Rau, M., Swigart, A., Gerber, C., & McKinsey, J. (2015, June). Understanding student success in chemistry using gaze tracking and pupillometry. In *International Conference on Artificial Intelligence in Education* (pp. 358–366) Springer, Cham.
- Plass, J. L., Moreno, R., & Brünken, R. (2010). Cognitive load theory. *Cambridge University Press*. <https://doi.org/10.1017/CBO9780511844744>
- Pouw, W. T. J. L., van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, 36, 51–72. <https://doi.org/10.1007/s10648-014-9255-5>.
- Radhamani, R., Kumar, D., Nizar, N., Achuthan, K., Nair, B., & Diwakar, S. (2021). What virtual laboratories usage tells us about laboratory skill education pre- and post-COVID-19: Focus on usage, behavior, intention and adoption. *Education and Information Technologies*, 26, 7477–7495. <https://doi.org/10.1007/s10639-021-10583-3>.
- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from, <http://www.R-project.org/>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <https://doi.org/10.1037//0033-2909.124.3.372>
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10, 241–255. <https://doi.org/10.1207/s1532799xssr1003-3>.
- Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers and Education*, 58, 136–153. <https://doi.org/10.1016/j.compedu.2011.07.017>.
- Schieber, F., & Gilland, J. (2008). Visual entropy metric reveals differences in drivers' eye gaze complexity across variations in age and subsidiary task load. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(23), 1883–1887. <https://doi.org/10.1177/154193120805202311>
- Shapiro, L. (2011). *Embodied cognition*. Routledge.
- Song, H. S., Kalet, A., & Plass, J. L. (2016). Interplay of prior knowledge, self-regulation and motivation in complex multimedia learning environments. *Journal of Computer Assisted Learning*, 32(1), 31–50. <https://doi.org/10.1111/jcal.12117>.
- Sullivan, S., Gnesdilow, D., Puntambekar, S., & Kim, J. S. (2017). Middle school students' learning of mechanics concepts through engagement in different sequences of physical and virtual experiments. *International Journal of Science Education*, 39(12), 1573–1600. <https://doi.org/10.1080/09500693.2017.1341668>.
- Triona, L. M., & Klahr, D. (2003). Point and click or grab and heft: Comparing the influence of physical and virtual instructional materials on elementary school students' ability to design experiments. *Cognition and Instruction*, 21(2), 149–173. https://doi.org/10.1207/S1532690XC12102_02.
- Tsai, M. J., Hou, H. T., Lai, M. L., Liu, W. Y., & Yang, F. Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, 58(1), 375–385. <https://doi.org/10.1016/j.compedu.2011.07.012>.
- Tsai, M. J., Wu, A. H., Bråten, I., & Wang, C. Y. (2022). What do critical reading strategies look like? Eye-tracking and lag sequential analysis reveal attention to data and reasoning when reading conflicting information. *Computers & Education*, 187, 104544.
- van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In R. Azevedo & V. Alevin (Eds.), *International handbook of metacognition and learning technologies* (pp. 143–156). Springer.
- Van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior*, 25(2), 325–331. <https://doi.org/10.1016/j.chb.2008.12.021>.
- van Riesen, S. A. N., Gijlers, H., Anjewierden, A., & de Jong, T. (2018). The influence of prior knowledge on experiment design guidance in a science inquiry context. *International Journal of Science Education*, 40(11), 1327–1344. <https://doi.org/10.1080/09500693.2018.1477263>.
- Wang, T. N., Jian, Y. C., Wu, C. J., & Li, P. (2022). Science reading and self-regulated learning: Evidence from eye movements of middle-school readers. *The Journal of Educational Research*, 115(1), 11–24. <https://doi.org/10.1080/00220671.2021.2012637>.

- Weisberg, S. M., & Newcombe, N. S. (2017). Embodied cognition and STEM learning: Overview of a topical collection. *Cognitive Research: Principles and Implications*, 2, 38. <https://doi.org/10.1186/s41235-017-0071-6>.
- Wen, C. T., Liu, C. C., Chang, H. Y., Chang, C. J., Chang, M. H., Chiang, F., Yang, S. H., C.-W., & Hwang, F. K. (2020). Students' guided inquiry with simulation and its relation to school science achievement and scientific literacy. *Computers & Education*, 149, 103830. <https://doi.org/10.1016/j.compedu.2020.103830>.
- Winberg, T. M., & Berg, C. A. R. (2007). Students' cognitive focus during a chemistry laboratory exercise: Effects of a computer-simulated prelab. *Journal of Research in Science Teaching*, 44(8), 1108–1133. <https://doi.org/10.1002/tea.20217>.
- Wu, C. J., Liu, C. Y., Yang, C. H., & Jian, Y. C. (2021). Eye movements reveal children's deliberative thinking and predict their performance in arithmetic word problems. *European Journal of Psychology of Education*, 36, 91–108.
- Zacharia, Z. C. (2005). The impact of interactive computer simulations on the nature and quality of post-graduate science teachers' explanations in physics. *International Journal of Science Education*, 27, 1741–1767. <https://doi.org/10.1080/09500690500239664>.
- Zacharia, Z. C., & de Jong, T. (2014). The effects on students' conceptual understanding of electric circuits of introducing virtual manipulatives within a physical manipulatives-oriented curriculum. *Cognition and Instruction*, 32(2), 101–158. <https://doi.org/10.1080/07370008.2014.887083>.
- Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction*, 21(3), 317–331. <https://doi.org/10.1016/j.learninstruc.2010.03.001>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Yu-Cin Jian¹  · Leo Yuk Ting Cheung¹ · Yi-Jye Wu¹ · Fang-Ying Yang² · Guo-Li Chiou³

✉ Yu-Cin Jian
jianyucin@ntnu.edu.tw

Leo Yuk Ting Cheung
leo.yt.cheung@gmail.com

Yi-Jye Wu
pata.heath@gmail.com

Fang-Ying Yang
fangyang@ntnu.edu.tw

Guo-Li Chiou
glchiou@ntnu.edu.tw

¹ Department of Educational Psychology and Counseling, National Taiwan Normal University in Taiwan, No.129, HePing East Road, Section 1, Taipei, Taiwan

² Institute of Science Education, National Taiwan Normal University in Taiwan, 88 Sec. 4 Ting-Zhou Road, 116 Taipei, Taiwan

³ Program of Learning Sciences, National Taiwan Normal University in Taiwan, No.129, HePing East Road, Section 1, Taipei, Taiwan