Computers & Education 196 (2023) 104726

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Teaching thermodynamics with augmented interaction and learning analytics

Wanli Xing [a,*], Xudong Huang [b], Chenglu Li [a], Charles Xie [b]

[a] School of Teaching & Learning, College of Education, University of Florida, Gainesville, FL, 32611, USA
[b] Institute for Future Intelligence, Natick, MA, USA

ARTICLE INFO

Keywords:
Thermodynamics learning
Infrared imaging
Image processing
Concreteness fading
Science education

ABSTRACT

Learning thermodynamics concepts remains challenging to students through grades K-12. In this study, we presented a design-based research program that aims to design and examine an innovative application, Infrared Explorer, that organically integrates physical experiments with a virtual environment. Relying on the theory of concreteness fading, we use infrared imaging and multimodal analytics to augment students' interaction with physical materials. Through two iterations of design and testing, we analyzed a variety of data sources and found that students actively used the application during their physical experimentation process. This technological application can allow students to highlight salient information and remove confusing details such that students' conceptual understanding of thermal concepts can develop further. Through this iterative development process, we also synthesized factors (e.g., multitasking and cognitive load) to consider when designing and implementing technologies to blend physical and virtual environments supporting learning. This research showed the promise of using concreteness fading to design science learning experiences through innovative technologies and identified important factors to consider during the design and implementation phases.

1. Introduction

Heat and temperature are intimately related to daily life. The concepts and principles of thermodynamics are the foundation for the natural sciences of physics, chemistry, and biology (Koh & Paik, 2002; Nottis, Vigeant, Prince, Golightly, & Gadoury, 2019). They are among the essential concepts that students must learn through the K-12 school science curriculum in most countries. However, the physics of heat has been identified as a notoriously difficult concept for students to grasp. Studies have found that students have problems linking temperature with a measurement of a physics attribute and often think that temperature and heat are the same (Lewis & Linn, 1994). Thomaz, Malaquias, Valente, and Antunes (1995) found that students have problems identifying the nature of heat, believing that heat is a kind of substance that can reside in objects, and which can further move from one object to another instead of a theoretical idea of energy transfer between objects of different temperatures. Moreover, students tend to misinterpret the temperature of an object based on their thermal sensation regardless of the fact that the initial temperature of the object is supposed to reach thermal equilibrium with ambient temperature (Russell, Lucas, & McRobbie, 2004). For example, students often respond that the temperature of a metal should be lower than wood because it feels cooler during the same physical interaction (Donnelly, Vitale, &

* Corresponding author.
E-mail addresses: wanli.xing@coe.ufl.edu (W. Xing), xudonghuang999@gmail.com (X. Huang), li.chenglu@ufl.edu (C. Li), charles@intofuture.org (C. Xie).

https://doi.org/10.1016/j.compedu.2023.104726
Received 19 April 2022; Received in revised form 5 January 2023; Accepted 8 January 2023
Available online 9 January 2023
0360-1315/

Linn, 2015). While learning the basic thermodynamics is crucial, it is challenging for K-12 students to understand these concepts as they are difficult to extrapolate from students' sensory experiences.

Previous research has primarily examined two contrasting approaches to facilitate the learning of thermodynamics concepts. The first approach is to design various physical lab activities. For example, Hitt and Townsend (2015) proposed to use household materials – food coloring and water – for students to visualize the connections between heat and particle motion. Rascoe (2010) presented a "how fluids move" experiment to allow students to engage in the process of scientific inquiry that leads them to understand a variety of thermodynamics concepts. The second approach examined a diverse set of virtual technologies that support thermodynamics learning. To illustrate, Tanahoung, Chitaree, Soankwan, Sharma, and Johnston (2009) conducted research to understand the effectiveness of interactive lecture demonstrations over traditional instruction on students' learning of heat and temperature. Several other studies applied simulation and modeling tools (e.g., microcomputer-based laboratory, MBL, Donnelly et al., 2015) and interactive heat transfer simulation (Xie, 2012). There are some other studies that explore various pedagogies to aid students in developing their understanding of heat and temperature, e.g., developing context-based materials (Bilgin, Nas, & Çoruhlu, 2017), or using engineering design activities (Schnittka & Bell, 2011), these works are usually investigated without any technological element. Therefore, they are not fully relevant to our research.

Two popular approaches, physical experiments and virtual tools and simulations, each have their unique affordances in supporting thermodynamics learning. For physical labs, Balamuralithara and Woods (2009) summarized the three key affordances for physicality, the acquisition of psychomotor skills, awareness of safety procedures, and learning how to use human senses for observation. Particularly, students can take advantage of tactile information in an experiment with thermos, according to theories of embodied cognition, to foster the development of conceptual understanding in thermodynamics (De Jong, Linn, & Zacharia, 2013). For virtual technologies, they provide rich information and varied representation (e.g., numerical, pictorial, graphical, conceptual, etc.) and offer capabilities to alternate reality (e.g., simplifying real-world models, visualizing objects and processes that are normally beyond perception, improving experiment efficiency, De Jong et al., 2013). Given these differing affordances, some researchers have explored the combing use of physical experiments and virtual tools. Most of these studies have primarily focused on utilizing both physical labs and virtual simulations in supporting students' learning (e.g., Jaakkola & Nurmi, 2008; Olympiou & Zacharia, 2012) rather than synthetically combining them to support learning.

In this study, we aim to design, develop, and investigate an innovative technology, Infrared Explorer[1], that systematically blends physical experimentation with a virtual tool to foster comprehension of thermodynamics concepts. Relying on the theory of concreteness fading (Fyfe, McNeil, Son, & Goldstone, 2014), the developed learning tool uses infrared imaging and data analytics to augment the learning of thermal concepts (Xie, 2011; Xie & Hazzard, 2011). Students begin by physically interacting with the experimental materials, then "seeing" the thermal phenomenon through infrared imaging, which is otherwise invisible. This experience then translates to an eventual abstract understanding of thermal concepts visualized by multi-representational data analytics. To examine the effectiveness of the learning tool, we took a design-based research (DBR) approach. DBR involves iterative cycles of design, evaluation, and improvement of learning interventions to identify meaningful educational practices (Anderson & Shattuck, 2012). Specifically, we conducted two iterations of design and testing to understand how students use the virtual tool, to determine how this tool augments students' scientific practices, and to further support the development of students' conceptual understanding of thermodynamics. This research method also enables us to distill design factors to guide the development of learning technologies for thermodynamics which mixes physical manipulatives and virtual technologies.

2. Theoretical foundation

2.1. Concreteness fading

The theory of concreteness fading is the framework that guides the development of Infrared Explorer such that it synthesizes the experience of a tactile physical experiment with dynamic virtual visualizations (Fyfe et al., 2014). In many learning settings, concepts can be presented using a diverse set of representations, some of which are more concrete than others (Fyfe & Nathan, 2019). In order to teach heat and temperature, for example, a metal spoon and a plastic spoon can be placed with ice cubes for a period of time to create a felt temperature difference. There are various benefits to using concrete materials. They can provide a practical context that can activate real-world knowledge (Schliemann & Carraher, 2002) and induct physical action to enhance memory and understanding (Glenberg, Gutierrez, Levin, Japuntich, & Kaschak, 2004), further enabling students to construct their own knowledge of abstract concepts (Berthold & Renkl, 2009). Using different virtual visualizations to teach thermodynamics can eliminate extraneous perceptual details and increase the portability and generalizability of what is learned in alternate contexts (Son, Smith, & Goldstone, 2008). Using abstract materials can also direct students' attention to structural and representational aspects, rather than towards superficial features (Kaminski & Sloutsky, 2009).

The theory of concreteness fading offers a framework to combine the advantages students gain from making connections in these different representational settings – by beginning with physical interactions with concrete instantiations of a target concept, and gradually and explicitly fading toward more abstract representations (Fyfe et al., 2014). Concreteness fading theory proposes that learning usually occurs by going through three stages of representation as in Fig. 1: (1) an enactive stage, which is action based (e.g.,

[1] Infrared Explorer is open to public use and can be accessed at https://intofuture.org/ie.html.

acting on a physical concrete model); (2) an iconic stage, which is image-based (e.g., using a pictorial or pictorial model); (3) a symbolic stage, which is notation-based (e.g., using an abstract model). This concreteness fading framework exploits the continuum from the concrete to the abstract and allows students to benefit from the grounded and tactile experience while still encouraging them to abstract and generalize beyond it. There is increasing theoretical and empirical support of concreteness fading to demonstrate how various features of this theory are helpful in learning contexts.

2.2. Multitasking and cognitive load

The design and development of Infrared Explorer are also informed by the affordances of multitasking on students' cognitive load. The implementation of multiple tasks within a short time is often referred to as concurrent multitasking (Salvucci, Taatgen, & Borst, 2009). However, multitasking is more than simultaneously or rapidly conducting different activities. Sequential multitasking is an additional type of multitasking in which the time gap between related tasks can be longer (e.g., an hour) and each task is equally important (Salvucci et al., 2009). Cognitive load describes the simultaneous mental activity used for information processing (Paas, Renkl, & Sweller, 2004). Excessive cognitive load, when information needed to be processed by learners exceeds their capacity, can impair students' learning outcomes (Seufert, Wagner, & Westphal, 2017). The excess of cognitive load can be caused intrinsically (e.g., interpreting learning materials using prior knowledge) and extraneously (e.g., the presentation or required manipulations with learning materials) (Sweller, 2005).

There have been studies examining the cognitive load affordances of concurrent and sequential multitasking, with concurrent multitasking having the tendency to increase cognitive load and sequential multitasking not significantly affecting the cognitive load. For example, Örün and Akbulut (2019) conducted a randomized controlled experiment to examine the effects of concurrent and sequential multitasking on participants' perceived mental effort and retention of knowledge. Their results showed that compared to participants in the control group (no multitasking), concurrent multitaskers tended to have lower knowledge retention and higher reported mental effort, while sequential multitaskers had no significant difference. In the proposed Infrared Explorer which involves both physical (e.g., conducting experiments) and virtual interactions (e.g., visualization and analysis using the provided tool), there is inevitably multitasking required that may negatively affect students' learning. It is important to consider multitasking and cognitive load in our design of Infrared Explorer. We discuss more details of multitasking and cognitive load for the two design iterations in the next section.

3. Infrared Explorer development

3.1. Concreteness fading in Infrared Explorer

Infrared Explorer was designed to embody the three stages of concreteness fading. The development of the enactive stage in Infrared Explorer is natural as the learning tool aims to augment students' exploration with the physical world through mobile devices and an infrared-ray (IR) camera enhanced by analytic tools. In this enactive stage, students can directly interact with the physical materials to gain the learning experience with physical objects. As shown in Fig. 2a, students can form a concrete perceptual model with the physical materials (e.g., rulers) used in experiments in Infrared Explorer. To enable graphic representations in the iconic stage, we have utilized thermal imaging powered by FLIR ONE Pro thermal camera and its software development kit (SDK). This thermal camera provides images rendered with a color heatmap to serve as direct indicators of temperature differences (see Fig. 2b).

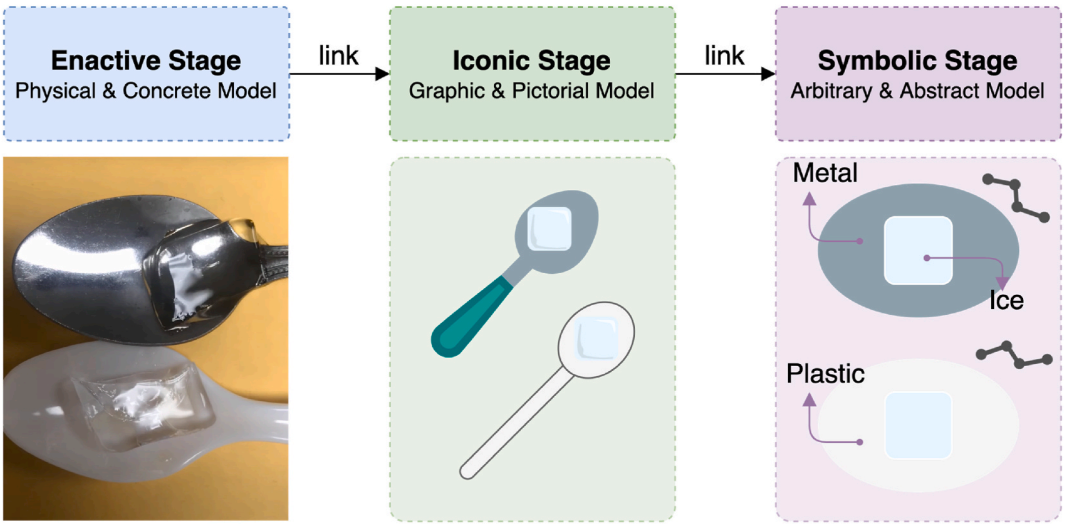


Fig. 1. Illustration of concreteness fading.

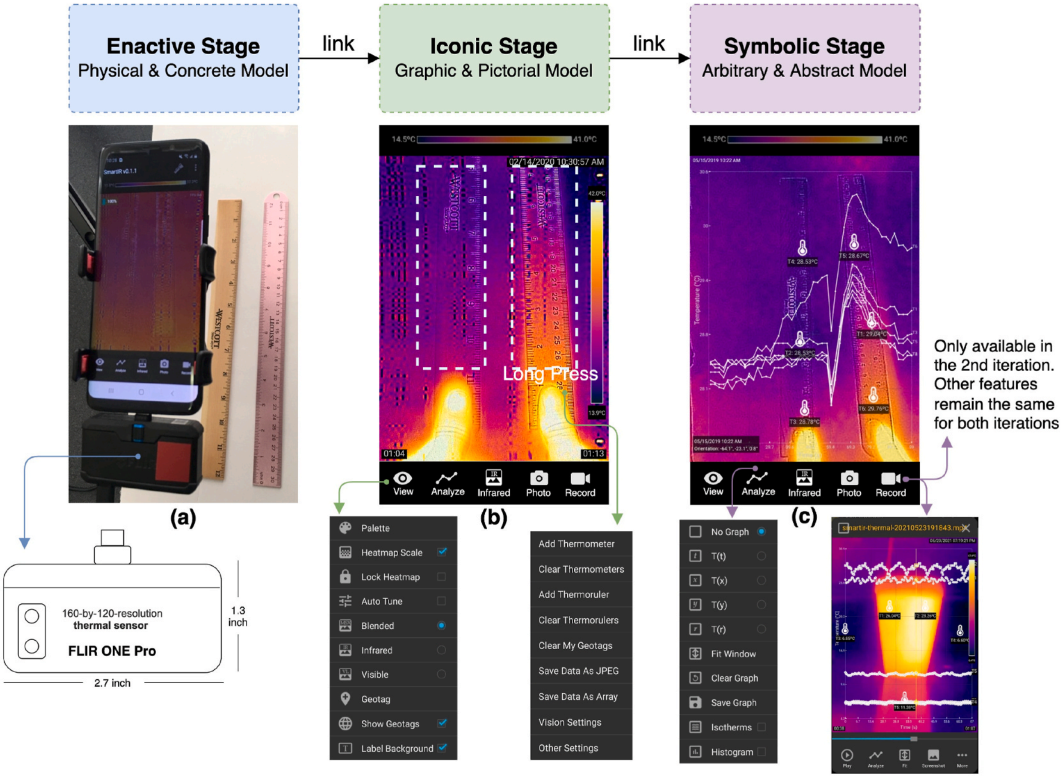


Fig. 2. Development of infrared explorer incorporating concreteness fading.

Meanwhile, buttons on the user interface are designed with icons to provide mental shortcuts for comprehending the purposes of each button. Finally, to assist with learning in the symbolic stage, we have developed a series of temporal and spatial graphs (e.g., temperature (Y) by time (X), temperature by horizontal distance (X)) to allow students to collect relevant structural patterns and representational features through data analytics (more detail in the next section). Fig. 2c demonstrates the use of temporal graphs to assist with students' information generation.

3.2. Data analytics in Infrared Explorer

Infrared Explorer was developed to allow students to conduct multi-representational data analytics to support their scientific inquiry. Studies have suggested that students' science learning can greatly benefit from data analysis with multiple representations to help them observe, measure, and decode scientific phenomena (Madden, Jones, & Rahm, 2011; Quintana et al., 2018; Xing, Lee, & Shibani, 2020). In Infrared Explorer, three representations of data can be utilized for analysis, all of which derive from sensor-collected thermal data. The first representational form visually analyzes using color. For example, students can use the heatmap superimposed on thermal images to easily identify temperature differences. The second representation utilizes graphing such that students can understand and compare trends of temperature change in various thermal image locations. Finally, students can analyze and interpret thermal data in a textual form in which they can quantify the temperature in a specific pixel location by placing and reading thermometers. Other than allowing students to learn through data analytics, Infrared Explorer also empowers researchers to conduct learning analytics to evaluate, analyze, and report students' learning processes. For example, other than thermal images and thermal data, Infrared Explorer also records students' every interaction with the learning tool (e.g., thermometer creation, movement, and removal) as system logs. Researchers can understand students' scientific inquiry triangulated from the perspectives of content (lab reports), dynamics (thermal images), and behaviors (system logs).

3.3. Two iterations of Infrared Explorer

Infrared Explorer experienced two iterations of design and development. The major distinction between the two iterations resides in how students can use Infrared Explorer to analyze thermal data for learning: real-time (mainly concurrent multitasking) or post hoc (mainly sequential multitasking), while other features (e.g., thermometers and graphing) remain the same. In the first iteration, students can only analyze data (e.g., the temperature at the pixel level) in real time as they are being collected by the thermal camera. When the thermal camera is disconnected or the app is closed, students will not be able to re-analyze the observed thermal phenomenon. The nature of real-time mode requires students to be well-prepared for an experiment; temperature changes can occur and disappear within seconds. To enable students to focus on initial observations and have second thoughts in experiments, we provided a post hoc analysis mode in the second iteration (Fig. 2, “Record” button), in which students can analyze data in their recorded thermal videos. In the post hoc mode, recorded videos retain the captured temperature data at the pixel level across time. Moreover, students can edit the recorded videos to eliminate noise in an experiment (e.g., trim the video to skip experiment preparation footage). Both real-time and post hoc modes support the three stages in concreteness fading well through access to various analytical tools (e.g., thermometers and temporal and spatial graphs). However, graphing in the post hoc mode differs from the real-time mode in that students can preview the complete temperature trend located by student-created thermometers. This is natural as students cannot have future data in real-time mode, but they have already concluded data collection in the post hoc mode.

4. Methodology

4.1. *The current study*

This study intends to investigate a technological innovation which mixes physical materials and virtual tools to transform the learning of thermodynamics using infrared imaging and data analytics. The resulting tool, Infrared Explorer, is primarily informed by the theory of concreteness fading and supports students’ various stages of thermodynamics learning. Following the design-based research approach, this study will develop, test, and refine this tool in an iterative process. We are guided specifically by the following research questions.

(1) To what extent do students use the virtual tool in the integrated learning environment and how does this use vary across the two iterations?
(2) To what extent do students conduct augmented observation using the virtual tool in the integrated learning environment and how does the augmentation vary across the two iterations?
(3) To what extent does students’ conceptual understanding of thermodynamics differ across the two iterations?

4.2. *Research context and participants*

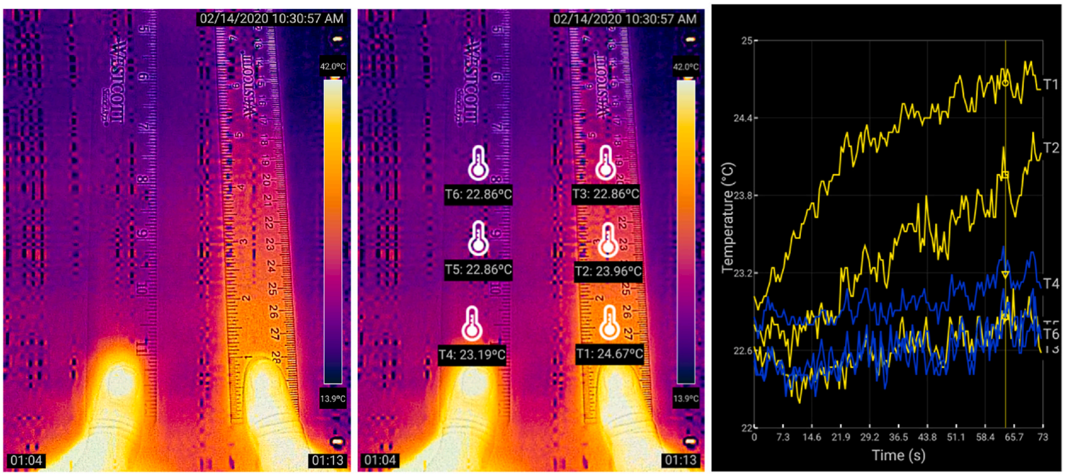
Students in the 9th grade (ages 14 to 15) from three suburban high schools in the Northeastern US participated in the study. In the first iteration of the study, 111 students from five physical science classes in one school participated and were taught by one male teacher. In the second iteration of the study, 132 students from seven science of energy classes, taught by three male teachers, participated, and another 72 students in four earth science classes were involved in the study taught by one female teacher from two schools, respectively. See Table 1 for student demographic details. Students without parental consent forms were excluded from this study. Thermal concepts taught included radiation, convection, conduction, and latent heat. The classical approach to conceptual learning, prediction-observation-explanation (POE, Bakirci & Ensari, 2018), was adopted. That is, before implementation of the lab, students used their prior knowledge to predict the experiment outcome (P) and gave initial explanations. After the scientific observation (O), students reconstructed their explanations (E) based on the collected real-world evidence.

All activities took five days, with one class per day. The first day was the warmup and Infrared Explorer tutorial. During the second to fifth days, the labs for four thermal concepts were conducted. Students completed a lab report each day and filled in the corresponding section of the lab report. The student data from the conduction concept was selected for this analysis because it was implemented during the latter part of the 4-day curriculum, which minimized the errors that came from learning a new tool. Also, this activity contained the most completed lab reports. The lab implementation of conduction (Fig. 3b) asked students to put two thumbs on metal and wood rulers for 1 min and observe the thermal differences between the rulers and thumbs. In the first iteration, the experiment implementation and data analysis happened simultaneously for students, while the second iteration separated the data analysis from implementing the lab. See Table 2 for the details.

To conduct experiments, students were paired together. Students were seated together in pairs in the classroom, so we followed this natural arrangement. Experiments were carried out by the student pairs. At the beginning of class, each pair received a printout of the experiment’s instructions, a phone and IR camera set, and the necessary supplies. The student pairs worked together and carried out the experiments following the steps listed on the handout. The teacher kept order in the room, circulated to answer students’ questions, and corrected any mistakes in following the written instructions. The researchers served as the technical experts in the classroom and helped resolve issues with the phone, camera, and software. At the end of the class, the same pair of students completed one lab report collaboratively. Therefore, the N reported in this paper’s result section indicates n pairs, not N individual students.

Table 1
School demographics of participants.

	Students	African American	Asian	Hispanic	White	Male	Female
First iteration	111	3.8%	10.9%	6%	73.8%	47%	53%
Second iteration School A	132	26%	6.9%	12.2%	48.5%	52%	48%
Second iteration School B	72	2.8%	36.6%	5.9%	50.7%	52%	48%



(a) (b) (c)

Fig. 3. The three methods to capture thermal data using Infrared Explorer app.

Table 2
Lab behavioral sequences in the first and second iterations.

Phase	First Iteration	Second Iteration
Before experiment	•Add thermometers to rulers. •Turn on the temperature-time T(t) graph	
During experiment	•Thumbs on rulers for 1 min •Take four images depicting thermal changes	•Record video •Thumbs on rulers for 1 min •Stop recording video
After experiment	Export the T(t) graph as an image	Open video for analysis: •Add thermometers on rulers •Turn on the T(t) graph •Export the T(t) graph as an image •Take four images of thermal changes

Note. The two iterations provided students with the same phone, IR camera, and lab materials. The second iteration used an updated version of Infrared Explorer, which provided a new feature of recording videos. Therefore, in the second iteration, students recorded the video while conducting the experiment, then analyzed the video to save snapshots as evidence. In this way, students could focus on observing the changes in real time with no need to worry about when to take a picture as evidence.

4.3. Data collection and analysis

System log *data analysis for RQ1*. Several major indicators from log data relevant to the experiment process were identified to understand students' usage of the tool and to determine whether the two iterations of implementations were significantly different from each other. These variables are: (1) duration of the experiment, which shows the total time each student dyad spent from the beginning to the end of the experiment and is identified by reviewing the logged screen video; (2) repetition times, which indicates each group's number of attempts to perform the experiment and to collect all the graphs as required in the lab report; (3) number of pictures taken during the experiment; (4) number of thermometers placed in the app to measure designated areas; (5) whether each pair turned on the T(t) graph during the experiment to analyze the temperature trend over time. Welch's two-sample t-tests were conducted to analyze whether students from two iterations differ from each other on duration, repetition frequency, number of pictures, and thermometer counts. A chi-square test of independence was used to determine whether students of both iterations were different in the toggling state of the temperature-time graph.

Lab report for RQ2. The way in which students carried out observations and interpreted data were measured using students' descriptions of five thermal images captured. The images include (1) the initial state of two rulers; (2) the two rulers with the thumbs pressed to them for 1 min; (3) the two rulers immediately after the thumbs are removed; (4) the thumbs immediately after they are removed from the rulers; and (5) the T(t) graph for the whole process. Students' descriptions of phenomena revealed how they utilized multiple pieces of evidence to support their observations. Infrared Explorer provided three types of evidence: the color heat map visualization (V) in which brighter color indicates the warmer area (Fig. 3a), temperature reading (R) from a virtual thermometer (Fig. 3b), and the T(t) graph (G) for trends over time (Fig. 3c). Besides the evidence from the app, students might also use their perception (P) and abstract concept (A) to interpret data. Two independent researchers rated the student writings from the first iteration of the design. They identified the evidence used in each image's interpretation and allowed the existence of multiple evidence

Table 3
Lab-report questions to measure students' conceptual understanding.

Iteration	Prediction	Explanation
First Iteration	Q1: Guess when you touch the two rulers, which one will feel cooler? Explain why. Q2: What will happen to the temperature patterns of the two rulers after you touch them for 1 min? Explain why. Q3: What temperature pattern will happen to your two fingers right after you move them away from the rulers? Explain why.	Q1: When you touched the two rulers, which one felt cooler? Explain why. Q2: As the image you took for Table 4(b) in the lab report shows, two rulers' temperatures were different at the places 2 inches away from the thumbs. Explain why. Q3: As the image of Table 4(c) in the lab report shows, the places two thumbs touched had different temperatures. Explain why. Q4: As the image of Table 4(d) in the lab report shows, one thumb was cooler than the other after touching two rulers. Explain why.
Second Iteration	Q1: After you press on the rulers for 1 min, what will happen to the area 2″ away from your thumbs on the two rulers? The area 2″ away from your thumb ___. a) on the metal ruler will be warmer b) on the wood ruler will be warmer c) on both rulers will be equally as warm Explain why. Q2: After you press on the rulers for 1 min, what will happen to the areas where your thumbs were pressing on the rulers? The area under your thumb ___. a) on the metal ruler will be warmer than that on the wood ruler b) on the wood ruler will be warmer than that on the metal ruler c) on both rulers will be at the same temperature Explain why. Q3: After pressing on the rulers for 1 min, which of the following relationships might best describe the temperatures (T) of the thumbs? a) $T_{thumb\text{-}on\text{-}metal} > T_{thumb\text{-}on\text{-}wood}$ b) $T_{thumb\text{-}on\text{-}metal} < T_{thumb\text{-}on\text{-}wood}$ c) $T_{thumb\text{-}on\text{-}metal} \approx T_{thumb\text{-}on\text{-}wood}$ Explain why.	Q1: Does thermal energy diffuse at different rates in different materials? Q2: Why does the thumb on the metal ruler feel colder?

sources for a single picture. The inter-rater reliability was 0.82. Automatic multi-label text classification using machine learning techniques was adopted in the second iteration to analyze students' lab reports (P, V, A, G, R) (see more details in Sung et al., 2021). In the construction of machine learning models, the researchers compared the state-of-the-art deep learning model, bidirectional encoder representations from transformers (BERT, Devlin, Chang, Lee, & Toutanova, 2018), with support vector machine (SVM), a classical machine model that has been reported with robust results in natural language processing applications (Dessì, Fenu, Marras, & Recupero, 2019; Kadhim, 2019). Different linguistic feature engineering techniques in natural language processing (NLP), such as part-of-speech tags (e.g., extracting grammatical components in sentences) and named entity recognition (e.g., extracting entities such as person, unit, time expression, location, etc.) were examined to achieve an optimized result. The model evaluation suggested that linguistic features slightly increased the predictive performance of SVM while degrading that of BERT. BERT trained with raw text data could achieve the best predictive accuracy, with an average area-under-the-curve (AUC) score of 0.94, showing an outstanding performance to accurately identify multiple possible sources of evidence in students' responses. Descriptive statistics and case analysis were used to determine how the tool improved students' observation of the physical experiment and was compared between the two design iterations.

Conceptual understanding for RQ3. Conceptual understanding was measured by students' scientific explanations during the POE cycle. Specific questions used are listed in Table 3. Based on the experience of the first iteration, we updated the POE prompts by explicitly listing possible outcomes for selection in prediction and asking integrated questions in the explanation. Student answers were scored by two independent raters on a scale of 0 (no answer) to 3 (fully correct). Inter-rater reliabilities were 0.80 and 0.84 for the

Table 4
Statistical test results of the first iteration and second iteration-lab major activities.

t-test	First Iteration		Second Iteration	t	df	Sig.
Duration (second)	269.27		256.89	−0.373	40.212	0.711
Repetition Times	1.68		2.36	2.543	37.638	0.015
Picture Taken	6.10		5.75	−0.581	52.917	0.564
Thermometer No.	9.73		8.68	−0.745	40.961	0.460
Chi-square test				X^2	df	Sig.
Use T(t) Graph		Used	Not Used	3.818	1	0.051
	First Iteration	24	4			
	Second Iteration	25	16			

* Note: The analysis contained different numbers of students from two iterations. This is because (1) The two iterations included a different total number of students; (2) A different number of students that did not provide consent were excluded from the analysis; (3) We lost some student log data due to the technical issues of the software.

first iteration and second iteration, respectively. Scoring rubrics are provided in Appendix A. Paired-sample t-tests were used to analyze the data to understand whether students' conceptual understanding of the thermal concept improved after the experiment.

5. Results

5.1. Infrared Explorer usage for RQ1

Students' usage of the tool is mainly reflected by several major indicators in the log data relevant to the experiment process. These variables are: the duration of each experiment, repetition times showing each group's number of attempts to perform the experiment and to collect all the graphs as required in the lab report, number of pictures taken during the experiment, and number of thermometers placed in the app to measure designated areas (see Table 4). The results showed that students spent about 5 min conducting the experiment, took 5–6 infrared images, and placed 8–10 thermometers using the virtual tool. Students also used the T(t) graph to support their understanding. Welch's two-sample t-tests were conducted to analyze whether student activities from the two iterations differ from each other. Only repetition times were found statistically significant between the two implementations ($t\,(37.638) = 2.543$, $p = 0.015$), which showed that students of the later implementation tried significantly more times to complete the experiment. An additional chi-square test of independence on whether students of both implementations turned on the temperature-time graph during their experiment showed that there was a clear tendency for statistically significant association between the two implementations, X^2 $(1) = 3.818$, $p = 0.051$.

Our explanation for the observed difference in retry times on the part of the students from the two iterations is that students were more willing to reiterate the experiment steps to collect all the evidence required in the lab report, which corresponded to certain features added in the latter implementation to facilitate data collection by allowing students to be able to review their recorded observations and then adjust measurements rather than to redo the entire process. Such features eased up students' cognitive load (Mayer & Moreno, 2003) and might also explain why visibly more students turned on the T(t) graph to observe the temperature trend over time during the first iteration.

5.2. Augmented observation for RQ2

Students were able to select evidence from multiple data sources to interpret the observed phenomena. The data sources include the three representations provided by the app (heat map visualization, V; temperature reading, R; and T(t) graph, G) and their tactile perceptions (P) and abstract concepts (A). A comparison was conducted on the evidence selection between the two iterations (Table 5). The first part of Table 5 shows that, for student interpretations using one evidence source, students in the first iteration reported notably more use of T(t) graph and slightly more temperature reading, while the students of the second iteration discussed apparently more heat map visualization and slightly more temperature reading. Students from the second iteration were more likely to interpret data using more than one resource from the app. When combining evidence from the app with perception, the two iterations showed similar percentages.

Several examples were provided to show how exactly one or more pieces of evidence were used in data interpretation. When interpreting the image taken for the state "*the two rulers with the thumbs pressing on them for 1 min,*" a student from the second iteration used a single source, color map visualization, as the evidence. The student wrote (for Fig. 4a) that, "*the metal rulers heat dispersed to cover the whole ruler, but the wooden rulers heat stayed concentrated in the place where the thumb was.*" Mentioning heat covering the whole metal ruler but only concentrated on a spot of the wood ruler is a direct indicator of using visualization as evidence. The other evidence, such as the temperature reading, was not mentioned in the description. In contrast, a student from the first iteration interprets the same state (image as Fig. 4b) purely with temperature reading: "*Towards the top of the computer the temperature was around 28 C, towards the middle closer to where the thumbs were placed the temperature was starting to go up but only by a decimal difference. Down at the bottom where the thumbs were placed the temperature had gone up to around 29.*" The student did not mention any color differences between the two rulers, even though they are obvious in the image.

A typical image interpretation using the T(t) graph as the single point of evidence is like this - "*When the thumbs first began touching the rulers the temperature began rising rapidly but the temperature increased more in the metal ruler than in the wooden ruler.*" The corresponding image is Fig. 4c. In the description, the student explicitly pointed out that the speed of temperature change over time is larger

Table 5
The percentage of different pieces of evidence used in students' observation description.

	Single source						Multiple sources	
	V	R	G	P	A	O	V/R/G	P + V/R/G
First Iteration	26.6%	16.6%	23.6%	7.4%	0.9%	0.4%	10.5%	14.0%
Second Iteration	48.8%	11.3%	3.8%	6.3%	–	–	16.3%	13.8%

Note 1. V = color map visualization; R = temperature reading; G = T(t) graph; P = perception; A = abstract concept; O = other uncategorized evidence; V/R/G = any combination of V, R, and G; P + V/R/G = the combination of perception and any V, R, G.
Note 2: The percentages in this table do not represent the proportion of students who use a specific type of data source. It calculates the percentage of image descriptions that use a particular type of single or combined source. A single unit in this analysis is the description of a single thermal image saved in lab reports. Students should ideally describe 5 thermal images.

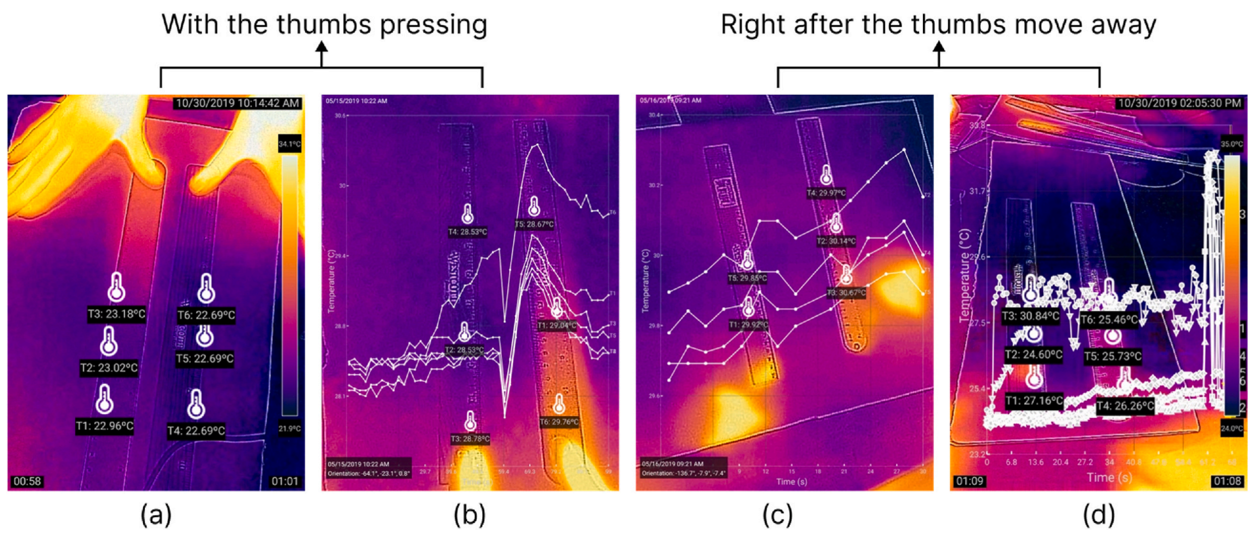


Fig. 4. (a) and (b) are two example images taken by students for the state “the two rulers with the thumbs pressing on them for 1 min” (c) and (d) are images for the state “two rulers immediately after the thumbs move away.”

on the metal ruler (i.e., “increase more”), which is clearly shown by the slope of the curves. Students also used more than one piece of evidence to interpret the data. A student from the second iteration took an image for the two rulers immediately after the thumbs moved away (Fig. 4d), and interpreted it as: “*The temperature changed on the rulers. It only changed in the spot the thumb was on the wood one, but it was more gradual on the metal one*.” Mentioning there was a temperature change implies that he or she had interpreted the temperature numbers. And pointing out the area of heat spread (e.g., “only … in the spot”) shows that the student also used information from visualization. In sum, students used various features in the virtual tool during their observation of the physical experiment and interpretation.

Color map visualization (V) is the most popular type of data used by students to interpret the results (first iteration, 26.6%; second iteration, 48.8%). This could be because of its intuitiveness and sensitivity to changes. First, the visualization is dynamic, displaying emerging temperature changes with high color contrast. More importantly, students’ actions on rulers are directly responsible for the changes. Students were most interested in this intuitive visualization of the causal relationship. When comparing two iterations, students from the second iteration used more visual evidence. This could be because they were allowed to watch the video several times after the operation, giving them more opportunities to observe the dynamic visualization. In contrast, in the first iteration, students only had one opportunity to observe this dynamic when performing experimental steps on rulers. Some students’ attention to the visualization is limited by competing tasks (operating on rulers vs. observing the outcome).

Given plenty of data sources, students seldom use abstract concepts (A) and other uncategorized evidence (O) as their evidence. One example answer that is coded as an abstract concept is “*It is showing that the metal ruler had a greater conductivity rate than the wood one*.” This answer directly quotes the concept of conduction, not mentioning any clues from their observation. One example answer that is coded as uncategorized evidence is “*the thermometers*.” This answer lacks enough detail to be categorized into a V, R, G, P, or A type.

5.3. Conceptual understanding for RQ3

Paired-sample t-tests were conducted to analyze whether students’ conceptual understanding was improved after the experiment. Only the students who completed both the prediction and explanation questions were included in this analysis. In the first iteration, there was an improvement in their conceptual learning, but the differences were only significant at the 0.1 level (prediction $M = 2.27$, $SD = 0.55$; explanation $M = 2.44$, $SD = 0.57$; $t\ (43) = 1.74$, $p = 0.089$). In the second iteration, the improvements were larger and

Table 6
T-test results comparing before- and after-lab conceptual understanding.

Iteration	Phase	n	Mean	SD	*t*-test	df	*p*
First Iteration	Prediction	44	2.27	.55	1.74	43	.089
	Explanation	44	2.44	.57			
Second Iteration	Prediction	19	1.60	.41	4.64	18	<.001
	Explanation	19	2.37	.50			

* Note: The total number of students analyzed for this research question is not identical to the number of students being analyzed for RQ1 (see Table 4). This analysis contains all the students who answered the prediction and explanation questions. However, some of these students’ log data were lost (especially from the first iteration) due to technical issues.

statistically significant in the 0.05 level (prediction $M = 1.60$, $SD = 0.41$; explanation $M = 2.37$, $SD = 0.50$; $t\,(18) = 4.64$, $p < 0.001$). Please see Table 6 for details.

Some students in the first iteration did not obtain accurate information from the experiment, probably due to a hurried observation that occurred simultaneously with lab operations. Therefore, their post-observation explanation of the thermal concept was problematic. For instance, a student predicted that the metal ruler would feel colder than the wood one, and briefly explained that it was "*because it's a better conductor.*" A visualization of how heat flows in the rulers in the observation phase should help this student expand this abstract explanation. However, this student's observation seems to have an issue because the re-explanation of the same answer after the observation became that "*because it is conducting the cooler temperature,*" which was the opposite of what was visually presented to the student - metal conducts heat.

The students in the second iteration improved their explanation after the lab observation by reasoning using evidence provided by the app. For example, in the prediction phase, a student made a correct prediction that, after the thumbs touched two rulers for 1 min, the metal one would be warmer than the wood one at the place 2 inches away from the thumb (prediction Q1). But he or she failed to give an explanation. After visually observing how quickly heat transferred on metal with unaided eyes, he/she accurately explained that "*Yes, we know this because heat in metal disperses faster than heat in wood*" (explanation Q1). On the same prediction question, another student even made an incorrect prediction, claiming that the area on the wood ruler would be warmer. This student's explanation for the prediction was confusing - "*The wood ruler is colder than the metal ruler meaning the wood ruler will continue to be hotter than the metal ruler.*" But his/her understanding of the concept improved after the experiment, by accurately explaining (for explanation Q1) that "*Yes it does, it depends on if it's a conductor or not. If it is, then it would diffuse faster than an insulator because it's easy to let heat in.*"

Students in the second interaction showed a lower prediction score (1.60) than the one of the first iteration (2.27). A possible reason is that we revised the questions in the second iteration to prevent vagueness in student answers. Students had to select one answer from three options and then explain the reason. This reduced the chance that students gave a vaguely right answer and got a higher score than they should be.

6. Discussion

To answer the research questions regarding the usage of the Infrared Explorer, we found that students actively used the virtual tool from the log data analysis. The activities included the collection of infrared images, placement of thermometers, and repetition of experiments. Overall, students spent a significant amount of time doing these activities. By comparing the first and second iterations of the design, we found that students' repetition times were significantly increased, and students tended to use more additional features (e.g., turn on the T(t) graph) and conduct experimental actions more effectively (e.g., fewer pictures taken, fewer thermometers placed). One hypothesis is that because the students can analyze the recorded video after finishing the observation and experiment, the experimental process is much simplified, and they are more likely to repeat the experiment more times to collect all the evidence for the lab report. Also, because students are much more focused during the experimental phase and become more engaged in the tactile experience (De Jong et al., 2013), they can use more features of the virtual tools and conduct the experiment more effectively.

For the augmented interaction, we found that almost all students used certain evidence from the virtual tool (e.g., heat map visualization, temperature reading, T(g) graph) for the physical experiment. This is important because focusing on salient information and removing confusing details (Kaminski & Sloutsky, 2009; Pei, Xing, Zhu, Antonyan, & Xie, 2022; Xing et al., 2021) is essential for students to develop evidence-based scientific practices. By comparing the two iterations of design, students used more evidence in their lab report and less perception and abstract ideas as evidence to support their observation and reasoning. They are also likely to use more heat map visualization and much less T(g) graph as evidenced in the second iteration of the design. The difference between the two iterations is probably because it is easier for students to understand images (iconic stage) than much more abstract graphs and figures (abstract stage) as in concreteness fading (Fyfe & Nathan, 2019).

When it comes to scientific understanding of thermodynamics, this learning technology improves students' learning of a set of thermal concepts including heat and temperature. By analyzing the comparison between prediction and explanation, we found that students' conceptual learning improved in both iterations of design. Particularly, in the second iteration, the difference in prediction and explanation is statistically significant. From the qualitative case analysis, we found that students relied on the infrared image and data analytics techniques to quantify the heat amount that is carried in a specific direction. Such an explanation is based on the tactile experience of feeling the metal ruler and wood rule combined with the iconic imaging and image analysis. Eventually, this explanation is transformed into conceptual understanding facilitated by the abstract representation of the heat and temperature graph using Infrared Explorer. The particularly significant increase of the conceptual understanding for the second iteration, post hoc analysis of videos rather than live analysis, may be a result from students' having more time to reflect (Denton, 2011) for the thermal concepts in the concernedness fading process.

Through design-based research, this study contributes to domain understanding of concreteness fading in supporting thermodynamics learning by demonstrating a number of factors as design considerations. The first factor to consider is cognitive load (Mayer & Moreno, 2003). In designing learning technologies based on concreteness fading, students have the opportunity to work with both physical and virtual objects. These objects will be further used to enhance conceptual understanding. The exposure of these various objects and concepts may easily lead to students' cognitive overload and reduce students' thinking and reflection time. The second factor to consider is the transition time between each concreteness fading stage. Little research illustrates how much time ought to be spent in each stage, physical, iconic, and abstract in the concreteness fading process. In the first iteration of the experiment, students can be easily moved from physical experimental to iconic and abstract representation of the experiment. In the second iteration, students are more focused on the physical experiment first and then engage in the iconic and abstract representations. The findings in this study provide greater support to engage in the physical interaction for a certain period of time before transition to iconic and abstract manipulations.

There are some possible limitations in this study. First, it is not fully fair to compare the first iteration and second iteration as students, teachers, and class subjects are different in the two rounds of experiments. The demonstrated difference may result from other factors, e.g., nature of the class, implementation quality, differences across teachers etc. Therefore, the findings should be generalized with caution across contexts. Second, for the observation coding, while the first iteration is fully coded by researchers, the coding for second iteration is conducted by machine learning algorithms. Though the performance for the machine learning model is very robust, it may still produce errors and contaminate the comparisons for the two rounds of studies. Meanwhile, although automatic text coding can provide valuable insights into students' textual artifacts at scale, it might not fully reveal students' actual learning. Third, the derived design factors for concreteness fading are inducted from the single thermodynamic experiment. These principles and factors may have limitations when apply to other situations as well.

7. Conclusion

Thermodynamics concepts are challenging for K-12 students to learn. Relying on the theory of concreteness fading, this paper presented a technological tool, Infrared Explorer, to augment student's interaction with physical materials with virtual manipulatives to promote the learning of thermodynamics. Using infrared imaging and data analytics, this design-based research study showed that the proposed technology to engage students' and support their learning of thermal concepts. This research demonstrates the promise of the concreteness fading framework to inform learning technology design and further identify important factors (e.g., cognitive load and multitasking) to consider when integrating physical and virtual learning environments. For future research directions, it is important to conduct control experiments to rigorously examine the effect of this technology and identify how to manage the time and transition between different concreteness fading stages across contexts and topics.

Credit author contribution statement

Wanli Xing: Writing – Original draft, Conceptualization, Methodology, Supervision. Xudong Huang: Data collection, Formal analysis, Writing. Chenglu Li: Formal analysis, Writing. Charles Xie: Software development.

Statements on ethics, and consent

Ethical statement

The data set was collected under an IRB approval through the Institute for Future Intelligence.

Consent statement

Consent forms were collected from all the participants of this study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the National Science Foundation (NSF) of the United States under grants #2054079 and #2131097. Any opinions, findings, and conclusions or recommendations expressed in this paper, however, are those of the authors and do not necessarily reflect the views of the NSF.

Appendix A. The rubric for scientific understanding

Score	Criteria	Examples
0	No answer	No response.
1	Prediction Q1-3; Explanation Q1: Selection/claim is wrong or off-topic.	Both rulers after touching them for 1 min they feel the same temperature (First iteration prediction Q2). (a) $T_{\text{thumb-on-metal}} > T_{\text{thumb-on-wood}}$. Metal is a conductor of heat (second iteration prediction Q3).
	Explanation Q2-4 (claim is given): Explanation is the rephrasing of question stem, off-topic, or containing big error(s).	The temperature was different between the 2 inches of the thumbs because the heat from the thumb cannot radiate the ruler all the way through using only a thumb (First iteration explanation Q2). Because the metal ruler is colder than the wooden one (second iteration explanation Q2).
2	Prediction Q1-3; Explanation Q1: Selection/claim is correct but the explanation is wrong, off-topic, or missing	The metal ruler will be cooler because heat leaves the ruler faster into your hand (First iteration prediction Q1). (b) on the wood ruler will be warmer than that on the metal ruler. The metal ruler is harder to warm up (second iteration prediction Q2).
	Explanation Q2-4 (claim is given): Explanation is partial correct or having minor error(s).	One thumb was cooler than the other after touching two rulers, because it gave away more heat to the metal ruler. The metal ruler started out with a cooler temperature but is a conductor, which is why it took more heat from the fingers. While the wooden ruler was the opposite which is why the temperature of the finger that touched the wood was warmer, and the finger that touched the metal was cooler (First iteration explanation Q4). Because it is much better at conducting heat, and it has not been heated to any point higher than room temperature (second iteration explanation Q2).
3	Prediction Q1-3; Explanation Q1: Selection/claim is correct and explanation is correct.	The finger that touched the metal ruler will be cooler than the finger that touched the wood one because heat is leaving your finger to go to the ruler (First iteration prediction Q1). Yes, we know this because heat in metal disperses faster than heat in wood (second iteration explanation Q1).
	Explanation Q2-4 (claim is given): Explanation is correct.	The spot on the metal ruler was cooler because the heat travelled up the ruler, because metal is a good conductor. But the spot on the wooden ruler was warmer because the heat stayed relatively in the same spot, because wood is a bad conductor (First iteration explanation Q3). Because the metal absorbed more heat than the wood (second iteration explanation Q2).

References

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*(1), 16–25.

Bakirci, H., & Ensari, O. (2018). The effect of common knowledge construction model on academic achievement and conceptual understandings of high school students on heat and temperature. *Education in Science, 43*(196), 171–188.

Balamuralithara, B., & Woods, P. C. (2009). Virtual laboratories in engineering education: The simulation lab and remote lab. *Computer Applications in Engineering Education, 17*(1), 108–118.

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology, 101*(1), 70.

Bilgin, A. K., Nas, S. E., & Çoruhlu, T.Ş. (2017). The effect of fire context on the conceptual understanding of students: "The heat-temperature". *European Journal of Education Studies, 3*(5), 339–359. https://doi.org/10.5281/zenodo.546161

De Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science, 340*(6130), 305–308.

Denton, D. (2011). Reflection and learning: Characteristics, obstacles, and implications. *Educational Philosophy and Theory, 43*(8), 838–852.

Dessì, D., Fenu, G., Marras, M., & Recupero, D. R. (2019). Bridging learning analytics and cognitive computing for big data classification in micro-learning video collections. *Computers in Human Behavior, 92*, 468–477.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805*.

Donnelly, D. F., Vitale, J. M., & Linn, M. C. (2015). Automated guidance for thermodynamics essays: Critiquing versus revisiting. *Journal of Science Education and Technology, 24*(6), 861–874.

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review, 26*(1), 9–25.

Fyfe, E. R., & Nathan, M. J. (2019). Making "concreteness fading" more concrete as a theory of instruction for promoting transfer. *Educational Review, 71*(4), 403–422.

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology, 96*(3), 424.

Hitt, A. M., & Townsend, J. S. (2015). The heat is on! using particle models to change students' conceptions of heat and temperature. *Science Activities, 52*(2), 45–52.

Jaakkola, T., & Nurmi, S. (2008). Fostering elementary school students' understanding of simple electricity by combining simulation and laboratory activities. *Journal of Computer Assisted Learning, 24*(4), 271–283.

Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review, 52*(1), 273–292.

Kaminski, J. A., & Sloutsky, V. M. (2009). The effect of concreteness on children's ability to detect common proportion. In N. Taatgen, & H. van Rijn (Eds.), *Proceedings of the conference of the cognitive science society* (pp. 335–340). Mahwah: Erlbaum.

Koh, H., & Paik, S. (2002). Analysis of conceptions of heat and temperature of the pre-service elementary school teachers. *Elementary Science Education, 21*, 81–100.

Lewis, E. L., & Linn, M. C. (1994). Heat energy and temperature concepts of adolescents, adults, and experts: Implications for curricular improvements. *Journal of Research in Science Teaching, 31*(6), 657–677.

Madden, S. P., Jones, L. L., & Rahm, J. (2011). The role of multiple representations in the understanding of ideal gas problems. *Chemistry Education: Research and Practice, 12*(3), 283–293.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.

Nottis, K. E., Vigeant, M. A., Prince, M. J., Golightly, A. F., & Gadoury, C. M. (2019). Computer simulations versus physical experiments: A gender comparison of implementation methods of inquiry-based heat transfer activities. *Chemical Engineering Education, 53*(4), 223–228.
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.
Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*(1/2), 1–8.
Pei, B., Xing, W., Zhu, G., Antonyan, K., & Xie, C. (2022). Integrating infrared technologies in science learning: An evidence-based reasoning perspective. *Education and Information Technologies, 1-21*. https://doi.org/10.1007/s10639-022-11538-y
Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., … Soloway, E. (2018). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences, 13*(3), 337–386. https://doi.org/10.1207/s15327809jls1303_4
Rascoe, B. (2010). What is heat? Inquiry regarding the science of heat. *Science Activities, 47*(4), 109–114.
Russell, D. W., Lucas, K. B., & McRobbie, C. J. (2004). Role of the microcomputer-based laboratory display in supporting the construction of new understandings in thermal physics. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 41*(2), 165–185.
Salvucci, D. D., Taatgen, N. A., & Borst, J. P. (2009). Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1819–1828).
Schliemann, A. D., & Carraher, D. W. (2002). The evolution of mathematical reasoning: Everyday versus idealized understandings. *Developmental Review, 22*(2), 242–266.
Schnittka, C., & Bell, R. (2011). Engineering design and conceptual change in science: Addressing thermal energy and heat transfer in eighth grade. *International Journal of Science Education, 33*(13), 1861–1887.
Seufert, T., Wagner, F., & Westphal, J. (2017). The effects of different levels of disfluency on learning outcomes and cognitive load. *Instructional Science, 45*(2), 221–238.
Son, J. Y., Smith, L. B., & Goldstone, R. L. (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. *Cognition, 108*(3), 626–638.
Sung, S. H., Li, C., Chen, G., Huang, X., Xie, C., Massicotte, J., et al. (2021). How does augmented observation facilitate multimodal representational thinking? Applying deep learning to decode complex student construct. *Journal of Science Education and Technology, 30*(2), 210–226.
Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. *The Cambridge handbook of multimedia learning, 3*(2), 19–30.
Tanahoung, C., Chitaree, R., Soankwan, C., Sharma, M. D., & Johnston, I. D. (2009). The effect of interactive lecture demonstrations on students' understanding of heat and temperature: A study from Thailand. *Research in Science & Technological Education, 27*(1), 61–74.
Thomaz, M. F., Malaquias, I. M., Valente, M. C., & Antunes, M. J. (1995). An attempt to overcome alternative conceptions related to heat and temperature. *Physics Education, 30*(1), 19.
Xie, C. (2011). Visualizing chemistry with infrared imaging. *Journal of Chemical Education, 88*(7), 881–885. https://doi.org/10.1021/ed1009656
Xie, C. (2012). Interactive heat transfer simulations for everyone. *The Physics Teacher, 50*(4), 237–240. https://doi.org/10.1119/1.3694080
Xie, C., & Hazzard, E. (2011). Infrared imaging for inquiry-based learning. *The Physics Teacher, 49*(6), 368–372. https://doi.org/10.1119/1.3628268
Xing, W., Lee, H. S., & Shibani, A. (2020). Identifying patterns in students' scientific argumentation: content analysis through text mining using Latent Dirichlet Allocation. *Educational Technology Research and Development, 68*(5), 2185–2214. https://doi.org/10.1007/s11423-020-09761-w
Xing, W., Li, C., Chen, G., Huang, X., Chao, J., Massicotte, J., & Xie, C. (2021). Automatic assessment of students' engineering design performance using a Bayesian network model. *Journal of Educational Computing Research, 59*(2), 230–256. https://doi.org/10.1177/0735633120960422

Wanli Xing is an Assistant Professor of Educational Technology at the University of Florida. His research interests are artificial intelligence, learning analytics, STEM education and online learning.

Xudong is a Learning Scientist at Institute for Future Intelligence. She received a Ph.D. degree in experimental psychology from the University of Memphis in 2016. Her research interests include technology-enhanced learning in science and engineering and learning analytics for educational research.

Chenglu Li is a doctoral student in the Educational Technology program at the University of Florida. His research interests include learning analytics, Bayesian statistics, fair AI for educational use, educational software development and educational Games.

Charles Xie is the founder and chief scientist at the Institute for Future Intelligence. His research includes computational science, learning science, machine learning, artificial intelligence, computer-aided design, mixed reality, Internet of Things, molecular simulation, solar energy engineering, and infrared imaging.