Journal of the Learning Sciences

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hlns20

Intelligent science exhibits: Transforming hands-on exhibits into mixed-reality learning experiences

Nesra Yannier, Kevin Crowley, Youngwook Do, Scott E. Hudson & Kenneth R. Koedinger

To cite this article: Nesra Yannier, Kevin Crowley, Youngwook Do, Scott E. Hudson & Kenneth R. Koedinger (2022): Intelligent science exhibits: Transforming hands-on exhibits into mixed-reality learning experiences, Journal of the Learning Sciences, DOI: 10.1080/10508406.2022.2032071

To link to this article: https://doi.org/10.1080/10508406.2022.2032071

Published online: 28 Feb 2022.

JOURNAL OF THE LEARNING SCIENCES
https://doi.org/10.1080/10508406.2022.2032071

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Intelligent science exhibits: Transforming hands-on exhibits into mixed-reality learning experiences

Nesra Yannier[a], Kevin Crowley[b], Youngwook Do[c], Scott E. Hudson[a], and Kenneth R. Koedinger[a]

[a]Human Computer Interaction, Carnegie Mellon University; [b]Learning Sciences and Policy, University of Pittsburgh; [c]School of Interactive Computing, Georgia Institute of Technology

ABSTRACT

Background: Museum exhibits encourage exploration with physical materials typically with minimal signage or guidance. Ideally children get interactive support as they explore, but it is not always feasible to have knowledgeable staff regularly present. Technology-based interactive support can provide guidance to help learners achieve scientific understanding for how and why things work and engineering skills for designing and constructing useful artifacts and for solving important problems. We have developed an innovative AI-based technology, Intelligent Science Exhibits that provide interactive guidance to visitors of an inquiry-based science exhibit.

Methods: We used this technology to investigate alternative views of appropriate levels of guidance in exhibits. We contrasted visitor engagement and learning from interaction with an Intelligent Science Exhibit to a matched conventional exhibit.

Findings: We found evidence that the Intelligent Science Exhibit produces substantially better learning for both scientific and engineering outcomes, equivalent levels of self-reported enjoyment, and higher levels of engagement as measured by the length of time voluntarily spent at the exhibit.

Contribution: These findings show potential for transforming hands-on museum exhibits with intelligent science exhibits and more generally indicate how providing children with feedback on their predictions and scientific explanations enhances their learning and engagement.

ARTICLE HISTORY
Received 04 May 2019
Revised 04 December 2021
Accepted 11 December 2021

CONTACT Nesra Yannier nyannier@andrew.cmu.edu Human Computer Interaction, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213.

This article has been republished with minor changes. These changes do not impact the academic content of the article.

Introduction

Much learning happens outside of school. Informal learning settings can support out-of-school learning by offering opportunities for children and families to learn together in an engaging way. However, learning support in such settings is not always adequate. Without scaffolding and support, people often miss the point of the learning activity in these settings (Hammady & Ma, 2019; Puchner et al., 2001). Current exhibits rely on parents, signage and staff/facilitators in the museums and science centers to provide support and guidance. Since it's not always feasible to have knowledgeable staff and not all parents have the same knowledge, children do not get the same support (Eberbach, 2009). New forms of technology can help, but what kinds of support are best in different contexts? There is evidence that children explore the evidence space of interactive science exhibits more widely and are more likely to construct explanations in the context of structured facilitation (Crowley et al., 2001; Fender & Crowley, 2007; Kim & Crowley, 2010). At the same time, some forms or levels of guidance can be counterproductive, particularly when they reduce chances for children to engage in their own inquiry (Hmelo-Silver et al., 2007). Kolodner el al. emphasize the importance of having learners engage in design challenges and construct a personally meaningful physical artifact (Kolodner et al., 2003). Some educational effect studies (Alfieri et al., 2011; Kirschner et al., 2006; Klahr and Nigam, 2004) also show that open-ended exploration activities often lead to learners demonstrating prolonged investigation, but also to difficulties with observing and interpreting observations.

Evidence from a meta-analysis of guided inquiry contexts (Furtak et al., 2012) concludes that a combination of guidance and exploration in inquiry learning environments is beneficial. Nevertheless, there are many open questions about what combinations of methods work best (cf., Koedinger et al., 2013). Within science education, there have been calls for more "adequate answers … concerning the conditions under which various types of scaffolded learning environments are most effective" (Hmelo-Silver et al., 2007). More recent research indicates progress (Eysink et al., 2015; Gormally et al., 2009; De Jong & Lazonder, 2014; Pedaste et al., 2015), but many open questions remain about what kinds of scaffolding and guidance work best—especially in the STEM context where "there is a lack of systematic evaluations and reliable experimental designs" (Alimisis, 2013). Further, there continue to be indications of skepticism about the desirability of scaffolding, for example, a recent meta-analysis comments that "Adding scaffolding has often been disregarded in the context of PBL [Problem-Based Learning]." (Kim et al., 2018). This meta-analysis found that the most prominent strategy of computer-based scaffolding, question prompts, has positive benefits, "but its effect was not as strong as many scholars believed."

We demonstrate, through a controlled random-assignment experiment, the effectiveness of scaffolding that not only prompts with questions but also gives automated feedback on students' decisions during scientific inquiry with physical phenomenon.

More generally, this project explores innovation both in scaffolding techniques and in technology to automate feedback during in-the-world experimentation. We evaluate whether techniques for personalized guidance enhance science learning outcomes, while we introduce a novel AI technology, an "Intelligent Science Exhibit," that provides a means to scale these techniques to more children. Intelligent Science Exhibits are a new educational technology genre that supports children in learning science while doing science in the physical world. An Intelligent Science Exhibit uses intelligent camera sensing to track children's interactions in physical and virtual spaces and provides interactive personalized feedback through the help of an engaging character. With this sensing capability, the exhibit adapts to the different needs of children, provides personalized support, and helps achieve equity goals (e.g., all children get support regardless of who they are and whether or not they have a science-literate caregiver with them). This type of support is especially important for children who do not have other opportunities to deepen their inquiry and understand physical phenomena, and in museums where it's not always feasible to have knowledgeable staff around the exhibits.

There has been progress in bringing AI advances into learning environments (e.g., Aleven et al., 2016; Holmes et al., 2019; Sottilare et al., 2018; Timms, 2016) and much progress in new AI sensing technologies (e.g., Benko et al., 2012; Gupta et al., 2012; Laput & Harrison, 2019; Zhang et al., 2018) that have potential for supporting learning. Recent developments in museum exhibits provide design guidance toward technology support for engaging and sustaining visitors' explorations (Hammady & Ma, 2019; Ma et al., 2019; Yoon et al., 2012). While building on this past work, our main technical contribution is in creating tangible interfaces that provide interactive feedback and guidance that adapts to changes in the physical environment by tracking student actions in the physical environment.

To do so, we developed new AI vision technology and Intelligent Science Exhibits that allows the system to *observe* students' actions—to accurately monitor and evaluate predictions, experiments, and explanations and provide an intelligent guided learning experience (Yannier et al., 2016). The technology also utilizes effective learning mechanisms of intelligent tutoring systems such as contrasting cases (Chase et al., 2010), predict-observe-explain (White & Gunstone, 1992), inquiry-based learning (Eysink et al., 2015; Gormally et al., 2009; Pedaste et al., 2015), guided-discovery (De Jong & Lazonder, 2014) and menu-based self-explanation (Aleven & Koedinger, 2002; Chi et al., 1989). More specifically, the physical set-up of the Intelligent

Science Exhibit consists of an earthquake table, physical towers, a display screen, a depth camera and a specialized AI technology to track what kids are doing as they do experiments in the physical world (e.g., making predictions about which of the towers on the table will fall first when the table shakes or building structures based on the challenges given by the system) and provide feedback accordingly (See, Figure 1).

Using this technology, we explore fundamental learning science questions about the active ingredients of active learning (Yannier et al., 2021). Our general research question is: Does adding more guidance and scaffolding structure to an exhibit enhance visitor learning? We consider two competing hypotheses. One hypothesis is that learning will be more effective when children explore freely and construct physical materials. The ICAP framework (Chi & Wylie, 2014), for example, suggests greater learning from constructive learning experiences than from active ones (which may include feedback and guidance). A competing hypothesis is that learning is more effective when guidance and scaffolding are provided in the context of structured inquiry activities. This hypothesis is consistent with meta-analysis of scaffolding in science learning environments (Kim et al., 2018) and theories recommending active learning (Klahr et al., 2009) or deliberate practice (Ericsson, 2008) with purposely designed tasks and explanatory feedback.

A second research question is: Does providing guidance enhance visitor engagement? One hypothesis is that guidance (e.g., prompting a visitor to explain a prediction or observations) will interrupt their flow and decrease their engagement. Bamberger and Tal (2007) found that more "directed learning in limited choice activities," while yielding more scientific discussion reduced social interactions and expressions of excitement in comparison to free choice exploration. In the authors' own experience on a prior project, game designers suggested that adding self-explanation prompts would reduce player enjoyment and thus such prompts were not included

Figure 1. Children and families interacting with the Intelligent Science Exhibit at Carnegie Science Center.

(Christel et al., 2012). An alternative hypothesis is that helping visitors make sense of their world is engaging and thus, guidance will enhance engagement. Anderson et al. (2000) found that enhancing student understanding of content, for example, by repeating episodes rather than showing them just once, increased engagement. UFranse et al. (2021) also found benefits of guidance in a museum setting whereby “interventions of museum educators positively affected the families’ learning process by reducing the number of scientifically incorrect remarks.”

Our prior experimental research has evaluated the importance of physical experience. While some (mostly non-experimental) research has suggested promise in physical interaction supporting learning (Manches & O’malley, 2012; Manches & Price, 2011; Marshall, 2007; Olympiou & Zacharia, 2012), there is also experimental evidence that found no learning benefit in the direct comparison of physical versus virtual in science learning (Klahr et al., 2007). Some have argued that empirically based studies are needed to understand better how tangible interfaces actually work and whether and why they might have benefits (Marshall et al., 2007). Thus, it was important to test whether children learn much more from our mixed-reality interaction and guidance than from that same interaction and guidance around flat-screen video of the earthquake table. We demonstrated, in randomized controlled studies, that children learn much more from mixed-reality interaction and guidance around the physical experience of observing actual blocks falling on an earthquake table than from that same interaction and guidance around flat-screen video of blocks falling on an earthquake table. In fact, a controlled experiment showed *five times* greater pre to post learning gains from mixed-reality compared to a screen-only tablet or computer version (Yannier et al., 2016, 2015).

The current experiment and our mixed-method approach evaluate the importance of the mixed-reality experience and, in particular, whether and what types of guidance can produce large learning gains while maintaining or even improving visitor engagement. In this paper, we explore benefits of STEM learning particularly with respect to improving students’ understanding of science concepts of early physics, as assessed through changes in the quality of their explanations, students’ competence in applying these concepts to make accurate predictions, and students’ ability to transfer these science concepts in engineering activities, as assessed by the quality of constructed artifacts (stable towers). While we do not pursue an investigation of impact on long-term student agency in science, we do evaluate impact on their immediate enjoyment and their choice in extended engagement in these activities.

As many traditional museum exhibits, maker spaces and constructivist theories suggest (Jeffery-Clay, 1998), might students learn as well as or better, without explicit guidance when in the context of more open-ended hands-on construction? Or perhaps better results can be found by selecting from and automating some of the many forms of instructional guidance and learning support that have been developed, explored, and tested (e.g., Clark & Mayer, 2016; Hattie, 2012; Koedinger et al., 2013; Schwartz et al., 2016)? We find it useful to distinguish between more *explicit* forms of guidance that use verbal instructions, prompts, or feedback from more *implicit* forms of guidance that use the structure and sequence of tasks to aid learning. We discuss next how these different forms of guidance are implemented in our Intelligent Science Exhibit and contrast them with the forms of guidance in a similar existing museum exhibit, which we argue is typical of many museum exhibits. These exhibits tend to have explicit guidance in the form of a sign and implicit guidance in the materials available for playful experimentation but tend not to have interactive facilitation from an always-present staff person or from automation. We thus call this an "Unfacilitated Exhibit" and use it as our control condition in the experiment described below.

In the section titled "**Use-Driven Iterative Design process of An Intelligent Science Exhibit**," we describe the design process we used in developing and improving our Intelligent Science Exhibit. We then return to the key research questions of whether an Intelligent Science Exhibit can yield substantial benefits, for both visitor learning and engagement, in comparison to an "Unfacilitated Exhibit" control condition. This control condition is meant to test beliefs that guidance is not necessary (c.f., Hovious, 2015) or may be counter to worthy educator goals such as "awakening the spark of curiosity in students and giving them the chance to explore and discover for themselves" (Cody, 2010). Providing guidance may reduce student opportunities to experience "desirable difficulties" (Bjork et al., 2013) or risk reduced student enjoyment or persistence in the activity if it comes off as "just another procedure to follow" (Schwartz et al., 2016, p. 97). Following the implications of this prior work, we strive for a design whereby guidance does not eliminate difficulties or provide recipe-style procedures, but rather prompts for genuine inquiry whereby students predict and explain novel situations they have not previously experienced *followed* by reactive feedback on outcomes of that inquiry.

We describe a random assignment field-based experiment that addresses these practical and scientific questions by comparing learning and engagement outcomes for visitors experiencing either the Intelligent Science Exhibit or the Unfacilitated Exhibit. Results indicate visitors learned substantially more from and were more engaged by the Intelligent Science Exhibit. We explain these results with support from illustrative case studies of observations of student interactions with the two different kinds of exhibits. We also report on exhibit engagement (measured by time spent voluntarily at the exhibit) in the natural museum setting.

The intelligent exhibit: A guided mixed-reality interactive experience

Our Intelligent Science Exhibit uses a specialized AI vision technology to track physical objects and children's actions as they experiment and make discoveries in the real world, facilitating guided inquiry thinking processes through a predict, explain, observe, explain cycle. A gorilla character appears on a large display to guide visitors as they make predictions, observations and explain results in the physical environment.

To scaffold learning, out design employs particular principles from learning science (e.g., Chi et al., 1989; Clark & Mayer, 2016; Gibson & Gibson, 1955; White & Gunstone, 1992) and game design (e.g., Falloon, 2010) including contrasting cases, self-explanation, predict-observe-explain, and real-time interactive feedback. The principle of *contrasting cases* originally came from work in perceptual learning (Gibson & Gibson, 1955) and suggests use of aligned cases with minimal differences to help students notice information that they might otherwise overlook. Gick and Paterson have shown that contrasts that differ on a single dimension can improve schema induction and transfer (Gick & Paterson, 1992). Based on this literature and more recent literature (e.g., Rittle-Johnson & Star, 2009; Schwartz et al., 2011), we decided to include contrasting cases in our design. Students are shown two towers that differ on only one principle (symmetry, wide-base, height or center of mass principles) and are asked to make a prediction about which one will fall first when the table shakes. Figure 2 shows a height-only contrast.

A second scaffolding feature, *self-explanation*, is supported by a number of studies. For example, students studying examples or textbook text learn with greater understanding when they explain the materials to themselves (Chi et al., 1989). Others have demonstrated that menu-based self-explanation enhances learning and transfer (Aleven & Koedinger, 2002). Similarly, in our system, we utilize a self-explanation menu to help children explain the reasoning for why a tower falls. The menu consists of explanation choices that express the underlying physical principles in child-friendly language such as: "because it is taller," "because it has more weight on top than bottom," "because it is not symmetrical," "because it has a thinner base."

We also utilize a *predict-observe-explain (POE) cycle* in our mixed-reality system. The POE method is a teaching-learning strategy proposed by White and Gunstone (White & Gunstone, 1992), which includes three stages: Prediction, Observation and Explanation. The POE process can be an effective teaching strategy to facilitate students' understanding of concepts, with the help of active discussions among students as they

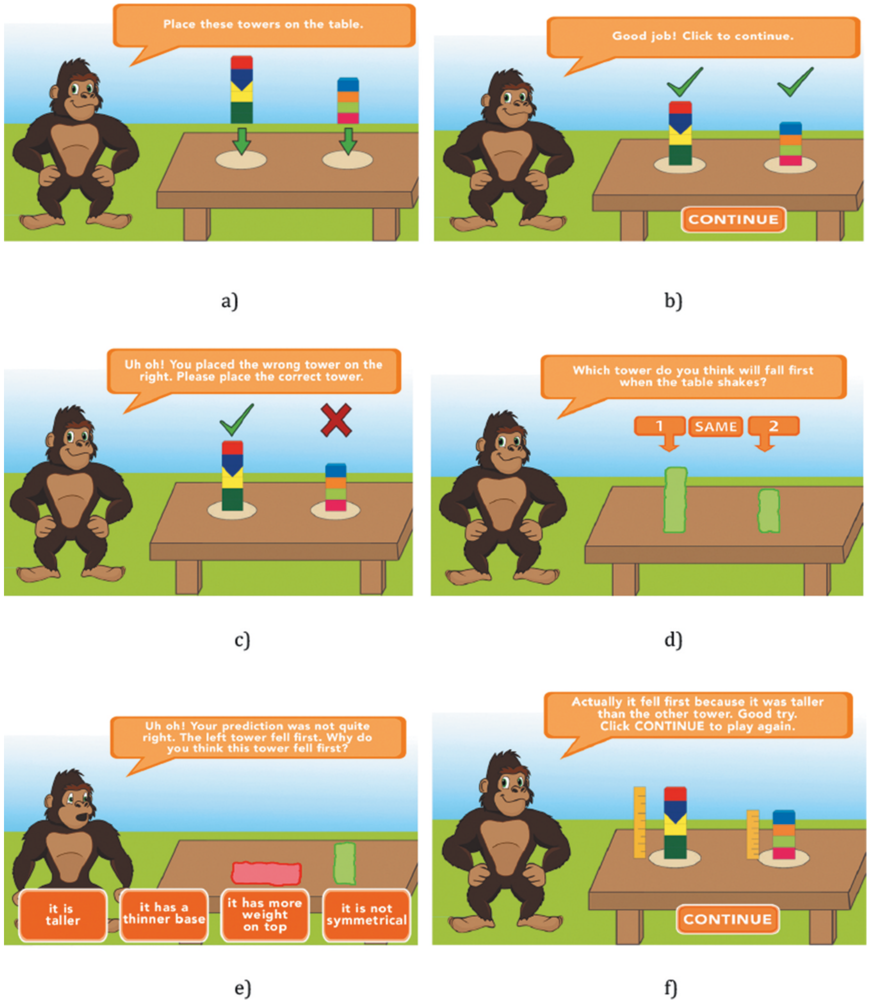

Figure 2. Screenshots from the "Guided-discovery" mode of the Intelligent Science Exhibit. Users are guided to place the given towers on the table, and are given feedback whether they placed the right tower or not. The game also gives feedback about their explanations with visualizations to help them understand the underlying physics principles.

explain any discrepancies between predictions and observations (Kearney, 2004). In our system, we utilize POE to help children understand the physics principles of balance and stability and why structures fall.

Our system provides *real-time interactive feedback*, which is a critical component. Immediate feedback and immediate error correction has been shown to be more effective than feedback on-demand and student control of error correction in intelligent tutoring systems (Corbett & Anderson, 2001). There is also research showing that without scaffolding and support, people often miss the point of the learning activity (Puchner et al., 2001). Also, the phenomenon of "confirmation bias" (Nickerson, 1998) suggests that children are likely to see their predictions as confirmed even when they are not, so explicit feedback otherwise can reduce this tendency and enhance learning.

In addition to learning science principles, we also employed game design principles including use of an animated character, a game-like scenario, and interactive physical construction and experimentation. Our system also includes *a character and scenario* that helps give immediate feedback to users. Avatars and virtual environments have been shown to have potential to act as powerful communication mediums for students to display knowledge and understanding (Falloon, 2010). Also, there is evidence that animated graphics enhance motivational appeals of instructional activities and embodied conversational characters, and virtual peers facilitate conversation in a virtual environment (Cassell, 2000).

Scenario

The features of our Intelligent Science Exhibit are illustrated in Figure 2. In the "Guided-discovery" mode, the users are first asked to place the towers shown on the screen on the physical earthquake table (See, Figure 2a). These towers are designed to be contrasting cases with only one difference between them so that the kids can focus on isolated principles. The Kinect camera and our specialized computer vision detects if towers placed on the table match the ones on the screen and gives feedback accordingly. If the tower they place on the table matches the tower on the screen, a green check appears above the tower on the screen and the gorilla character says "Good job! Click to continue" (See, Figure 2b). Otherwise, if the tower they place does not match the tower on the screen that they were asked to place, the computer vision system detects that it was not the correct tower, and a red cross appears on the tower on the screen, and they're asked to place the correct tower (See, Figure 2c).

Our computer vision algorithm uses depth information from the depth-sensing camera in algorithms for blob detection and physics-based moment of inertia calculations (Yannier et al., 2013) to recognize and differentiate objects of experimentation, the towers in this case, and to recognize critical experimental outcomes, a falling tower in this case. Using the steps described above as an example, when students try to place a requested tower on the table, the vision algorithm computes a moment of inertia value that serves as a kind of signature of the identity of the tower. If this value matches the stored value for the requested tower, the system provides positive feedback as described above (see, Figure 2b). If this value does not match, the system provides corrective feedback (see, Figure 2c).

Once they click continue, the gorilla character prompts them to make a prediction saying, "Which tower do you think will fall first when I shake the table?" (See, Figure 2d). They can choose either 1, 2 or same by clicking on the buttons or one of the towers on the screen (a virtual projection of the towers as a blob is drawn on the screen).

After they make a prediction, the gorilla character says: "You chose the left tower. Why do you think so? Discuss and then click SHAKE to see what happens." Here they can discuss their prediction with their partner/friends/family, why they think the tower they chose will fall first. They then click the "Shake" button on the touch screen and the physical earthquake table starts shaking (the touch screen triggers the relay that is connected to the motor of the earthquake table).

After the table starts shaking, when one of the towers falls, the Kinect camera and our specialized computer vision algorithm detects the fall and stops the earthquake table from shaking. The vision algorithm detects whether the left or right tower fell and if it matches with the prediction of the user. If the user's prediction was correct and the right tower fell first, then the gorilla character says "Good job! Your hypothesis was correct! Why do you think this tower fell first?." The gorilla on the projector screen is happy and starts jumping and dancing to give them positive feedback and reward. On the other hand, if the user's prediction was wrong and the tower that fell does not match the tower that was predicted by the user, then the gorilla character says "Uh oh! Your prediction was not quite right! Why do you think this tower fell first?." This time, the gorilla on the projector screen looks surprised. Users are asked to explain why they think the tower that fell, fell first. They are given an explanation menu with four different choices that they can choose from: "It is taller," "It has a thinner base," "It has more weight," "It is not symmetrical." These explanations correspond with the four different principles of stability and balance that are primary science content learning objectives for the exhibit. This explanation menu comes up even if their prediction is correct.

When they click on one of the choices in the explanation menu, the gorilla character tells them if their explanation was correct or wrong, with a visualization laid over the images of the towers on the screen to explain why the tower actually fell (See, Figure 2f). For example, if the reason was because it had more weight on top than their bottom, and their explanation was not correct, he says: "Actually it fell first because it had more weight on top than bottom. Good try. Click CONTINUE to play again." The visualization of the

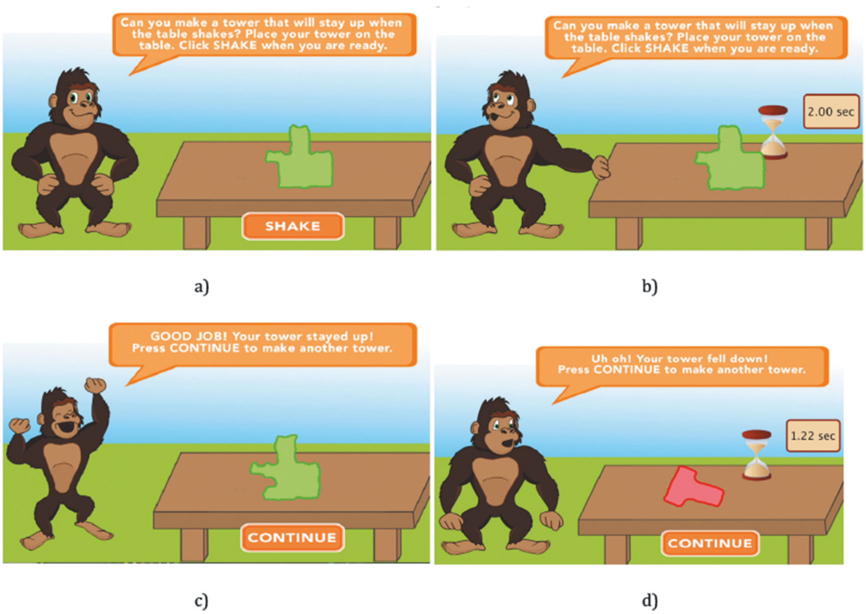

Figure 3. Screenshots of the Exploration mode of the Intelligent Science Exhibit. In this mode, the gorilla character asks the users to make a tower that will stay up when the table shakes. The AI vision algorithm tracks any tower they build, and gives feedback if the tower stayed up or not and for how long.

towers shows circles on the parts of the towers that have more weight. Or if it fell because it was taller, the ruler visualizations this time highlight the height of each tower.

This scenario is repeated for different contrasting cases of towers (Yannier et al., 2020).

In addition to the “Guided-discovery” mode, users can try the “Exploration” (or “Test My Tower”) mode. In this mode, they can build any tower they want using the wooden, Lego or magnetic blocks in the bins. The gorilla character says: “Can you make a tower that will stay up when the table shakes? Place your tower on the table and click SHAKE when you are ready.” (See, Figure 3a) When they have built their tower and click the SHAKE button on the touch screen, it triggers the motor in the physical earthquake table, and the table starts shaking. If their tower falls down, the Kinect camera and our specialized computer vision algorithm detects the fall, and the gorilla character gives feedback “Uh oh! Your tower fell down! Press CONTINUE to make another tower” (Figure 3d). The system also displays how many seconds it took for the tower to fall down. If the tower does not fall down in 5 seconds, the earthquake table stops shaking. Then the gorilla character starts jumping/dancing (similar to that in the guided-discovery mode) and says: "Good job! Your tower stayed up! Press CONTINUE to make another tower" (Figure 3c). They are then challenged to make another tower (e.g., taller) to test on the earthquake table.

Based on previous research (Yannier et al., 2020, 2021), the Intelligent Science Exhibit condition in this experiment combines the Guided-discovery and Exploration modes as it led to better learning outcomes than use of either mode alone.

Use-driven iterative design process of an intelligent science exhibit

Early iterations at the children's museum of Pittsburgh

We engaged in iterations of use-driven design inspired by approaches of design-based research (Barab & Squire, 2004; Design-Based Research Collective, 2003; Wang & Hannafin, 2005), but with a practical focus on enhancing the usability and user engagement of the resulting Intelligent Science Exhibit design and not as a primary research output in itself. These iterations were done mainly on the Guided-discovery mode of the system. Our goals were to gain insights about museum visitors, both children and parents, and toward improving the design of the Intelligent Science Exhibit, before coming up with the latest design of the exhibit (Figure 1). Key usability questions included:

1) Do children find the Intelligent Science Exhibit interesting and intuitive to use and can they use it independently without researcher assistance?
2) Does the Intelligent Science Exhibit engage parents? Do they help their children?
3) Is the Intelligent Science Exhibit engaging enough to compete with other museum exhibits?
4) Does the guidance we added enhance learning discourage or increase engagement?

On-site observation of system use in two iterations

System design and usability observations: Iteration 1

Our first usability investigation took place in a pull-out room of the MakeShop of Children's Museum of Pittsburgh (See, Figure 4). Around 40–50 people interacted in small groups with the exhibit over two days. Typical groups included 1–2 children with 1 parent or responsible adult.

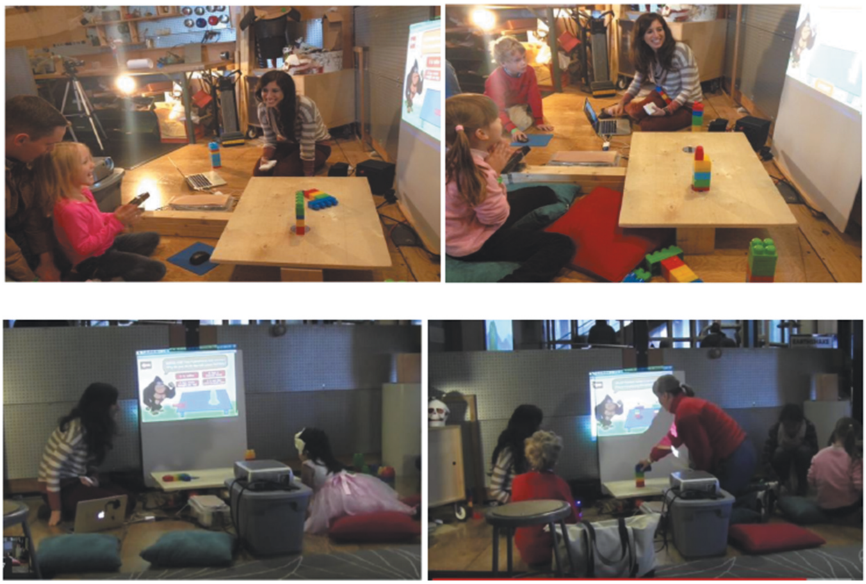

Figure 4. First iteration of play testing at the Children's Museum of Pittsburgh. Note the early, more primitive form of the prototype exhibit, with the projector, mouse control, simple shaking table etc (in contrast to Figure 1). The exhibit fostered engagement and interaction between the children and with parents. Both children and parents seemed to be quite engaged in the exhibit.

Consistent with prior observations and surveys in the lab (Yannier et al., 2020), children appeared quite interested and engaged in the Intelligent Science Exhibit. One child played with the game early in the morning before anyone else came, then she brought her whole birthday party group to play later on. In addition to positive observations, we found opportunities to improve the system. Younger children had a hard time using the mouse. Therefore, we decided to use a touch-screen tablet as an input device instead of the mouse. More substantially, our computer vision algorithm did not perform as well in the museum as it had in the lab. We discovered this performance degradation was a consequence of the perceived colors of the blocks being affected by the lighting conditions in the museum. To address this issue, a new vision algorithm was developed that summarizes a tower shape into a unique set of "moment of inertia" values (Yannier et al., 2020) that does not depend on color.

We also discovered opportunities to improve physical elements of the Intelligent Science Exhibit. The prebuilt towers that we had created by sticking Lego blocks together were not durable. Children thought that they could be taken apart when they saw the Lego blocks (glued together to create prebuilt towers) and tried to separate them. Instead, we created prebuilt

Figure 5. Children imitated the gorilla character when their prediction was correct and the gorilla started dancing on the screen. The parents were also very much engaged with the game. At some points they started playing on their own.

towers made of wood blocks that would be secured tightly to each other and painted to create a visual cue. We continued use of individual Lego blocks for children to build their own towers in the exploration mode.

Parents were generally quite involved in working with the kids. Some of the parents were also curious about how the system worked. For example, one of the dads said: "Very cool!" and asked how we had built the system and if he could set it up at home. In general, the parent-child interaction seemed quite natural and indeed parents prompted their children to follow the on-screen guidance.

System design and usability observations: Iteration 2

Based on input from the first iteration, we redesigned the system and produced a new version with a more robust mechanism and more polished physical design (See, Figure 5). Specifically, we iterated on our computer vision algorithm, so it uses depth information instead of color information that is more reliable in different lighting conditions at the museum. We have also made the physical design more durable for children based on the findings from Iteration 1 described above. This new version was playtested in a second design iteration that took place in the MakeShop of Children's Museum of Pittsburgh. The MakeShop provided a setting that was more open to the public than the private room used in iteration 1. Our Intelligent Science Exhibit was available over two days in the weekend, from 10 am until 5 pm. Around 40–50 people interacted in small groups with the exhibit over two days. Again, typical groups included 1–2 children with 1 parent or responsible adult. There were two researchers observing and taking notes. In this iteration, we were not only interested in child engagement and parent involvement, but also in Usability Questions 3 and 4: Is the Intelligent Science Exhibit engaging enough to compete with other museum exhibits? and does the guidance we added to enhance learning discourage or increase engagement?

As in Design Iteration 1, both kids and parents were quite engaged by the exhibit. Children were particularly engaged by the new animated gorilla character. When the gorilla responded by dancing when a child's prediction was correct, many kids celebrated by mimicking the gorilla's dance moves (see, Figure 5a).

We observed that parents and children seemed to discuss and engage in productive dialogue, discussing the reasons why towers fall, more in the guided-discovery mode. When the game prompted them to explain why they think one of the towers fell first, the parents started asking children why they think the tower fell and helped them understand the underlying principles. For example, one of the moms put the towers side by side, asking her son what the difference is between the two, if one has more weight on top than the other. Similarly, many other parents got involved and tried to guide the kids to understand why and to celebrate success (e.g., giving a high five; Figure 5b). The interactive guidance from the system appeared to prompt parents to interact with their children by encouraging inquiry toward discovering scientific explanations rather than up-front telling. For example, they would repeat the questions that the exhibit asked to their children "Which tower do you think will fall first?," "Do you know why?." Also, the self-explanation options in the exhibit seemed to guide them to understand the underlying principles e.g., "because it's not symmetrical," "because it has a thinner base," "because it has more weight on top than bottom" etc. The parents repeated these explanations to the children, especially for younger children.

We also observed that children and families seemed to strategize more while building their towers when they did it after the guided discovery activity as opposed to doing it as a first activity when they came to the exhibit. If they chose to build a tower before interacting with the guided-discovery activity, they tended to do more random tweaking not aligned with the underlying principles of stability. Children and parents did seem to enjoy building their own towers in addition to doing experiments with prebuilt towers which suggested that it may be good to combine Guided discovery (making predictions and explanations with prebuilt towers) and Explore Construct modes (where they build their own towers based on what they learned in the Guided discovery mode). These observations informed the research question and hypotheses for the experiment below, namely do children learn better from exploration and construction in a traditional exhibit or from an Intelligent Science Exhibit with interactive guidance? How can we combine both interactive guidance and exploration for maximized learning and enjoyment outcomes?

Similar to Design Iteration 1, parents seemed to be very much engaged as well as the children. We heard from museum staff this is an important quality for museum exhibits—that it should appeal to people of different ages. In some cases, the parents even started playing on their own, including one father who played the game for over 15 minutes (Figure 5c).

Learning experiment: Intelligent guidance better enhances science and engineering outcomes

As described above, our experiment was motivated by open practical and scientific questions. The key practical question is whether concerns raised by some museum professionals that adding "intelligent" guidance is not necessary are warranted. One key scientific question is whether intelligent guidance may suppress "desirable difficulties" (Bjork et al., 2013) that produce effective constructive learning (Kafai & Resnick, 1996) experiences. Another is whether intelligent guidance may reduce student agency and subsequent engagement (Schwartz et al., 2016). Put simply, the experiment tests whether the Intelligent Science Exhibit produces better learning and engagement than typical museum exhibits that come with little or no automated facilitation.

Methods

Unfacilitated exhibit control condition

We designed the "Unfacilitated Exhibit" control condition to meet two important criteria. To enhance ecological validity, it should be as much like existing exhibits as possible. To enhance internal validity, it should be otherwise similar to the Intelligent Science Exhibit, but without the intelligent guidance. Thus, we developed an exhibit to represent current practice at museums. This mimics an earthquake exhibit at the science center where we

a) b)

Figure 6. a) Earthquake exhibit regularly used at the Science Center. b) Our matched Unfacilitated Exhibit Condition consisting of the same earthquake table set up used in the Intelligent Exhibit Condition, with the AI turned off to match the standard earthquake exhibit and the same materials used to minimize the variables.

conducted our experiment. It also makes use of the same base materials as the Intelligent Science Exhibit. The Unfacilitated Exhibit uses the same physical materials/blocks/table used in the mixed-reality Intelligent Science Exhibit. This approach also controls for exhibit location as there is just one exhibit, sometimes with the interactive guidance turned on (Intelligent Science Exhibit) and sometimes turned off (Unfacilitated Exhibit).

Figure 6a shows the earthquake exhibit at Carnegie Science Center where the study took place. Visitors control the shaking using the "Start" and "Stop" buttons. Visitors build towers using the given materials and may shake them on the earthquake table to observe if their towers stay up or not. Similarly, our Unfacilitated Exhibit Condition (see, Figure 6b) consists of the same earthquake table set up we used in the Intelligent Exhibit Condition. On the touch screen, there are "Start" and "Stop" buttons. Visitors can use all the same physical materials used in the Intelligent Exhibit condition to build towers and test if their structures stay up or not by shaking the table. In other words, the Unfacilitated Exhibit is the Intelligent Exhibit with the intelligent features turned off to match typical exhibits that come with no facilitation. Similar to other conventional exhibits at museums, we also had a sign for our Unfacilitated Exhibit condition, explaining concepts such as "symmetry, center of mass, height, weight."

We compared this Unfacilitated Exhibit with our Intelligent Science Exhibit described above including a combination of "Guided-discovery" and "Exploration" modes as found to be effective in a previous experiment (Yannier et al., 2020). In this Intelligent Science Exhibit condition, children are exposed to "Guided-discovery" activities followed by "Exploration activities" where the scaffolding fades, as described in the Scenario section above. The Intelligent Science Exhibit did not have any signage and the guidance was provided merely through the AI scaffolding.

Participants

Participants were elementary-aged students (first through fifth graders) attending a science summer camp at the science center. This user group is consistent with the original design goals and early learner testing with K-5th grade children. 32 children participated in pairs (16 pairs) at the exhibit floor away from the rest of the summer camp activities. The pairs were randomly selected and assigned to different conditions. The ages of the kids were balanced for both conditions. The median grade level of participants for both the Intelligent and Unfacilitated Exhibit participants was second grade. A previous study showed no significant difference in learning and enjoyment between kids working in pairs versus solo (Yannier et al., 2015). We conducted the study in pairs as previous feedback from parents and teachers indicated a preference to have kids interact collaboratively in pairs.

Measures

We used assessments of both scientific and engineering outcomes to measure learning. For scientific outcomes, we evaluated whether children correctly used the four principles of stability and balance when explaining their predictions of given tower contrasts. For engineering outcomes, we evaluated the quality of the towers that children built before and after learning interactions based on which principles were exhibited or violated. We also assessed students' ability to predict which tower would fall first in given pairs of towers.

o measure pre- to posttest changes on the tower building task, we scored each student's towers according to three principles: height, symmetry, and center of mass (we did not use the fourth principle, wide base, as all students were instructed to use the same base block). For each principle, students were given one point if their towers improved from pre- to posttest, −1 for the reverse, and 0 for no change. Comparing pre- and post- towers for the height principle, a shorter post-tower scores 1, a taller post-tower scores −1, and towers of the same height score 0. For example, if a student's tower was shorter on the post-tower compared to the pre-tower, it would get a 1 for Height Score (see second image in Figure 7). However, if the post tower is taller than the pre-tower, it would get a −1 (see first image in Figure 7). If the post tower is the same height as the pre tower, it would get a 0 (see fifth image in Figure 7). Likewise, post-towers with more symmetry and a lower center of mass score one for each of those principles. Adding the scores for each principle yielded the student's total score (Figure 7). The students'

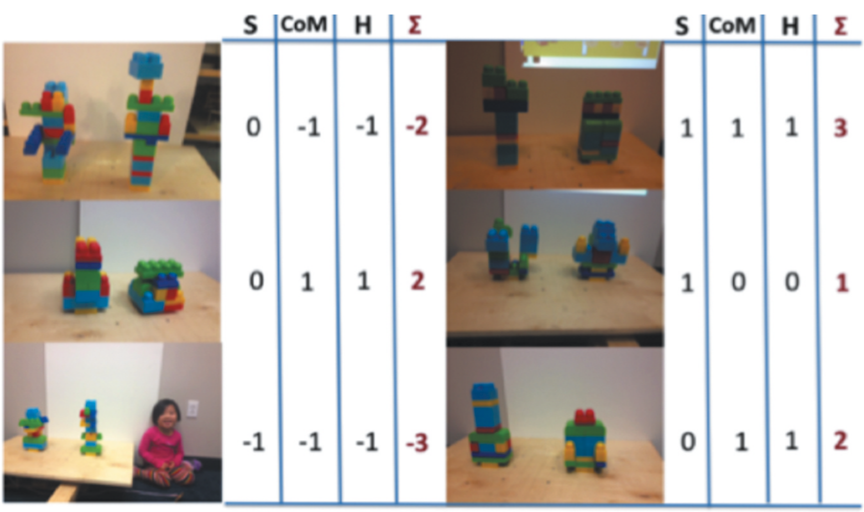

Figure 7. Coding scheme for Tower pre/post tests change.

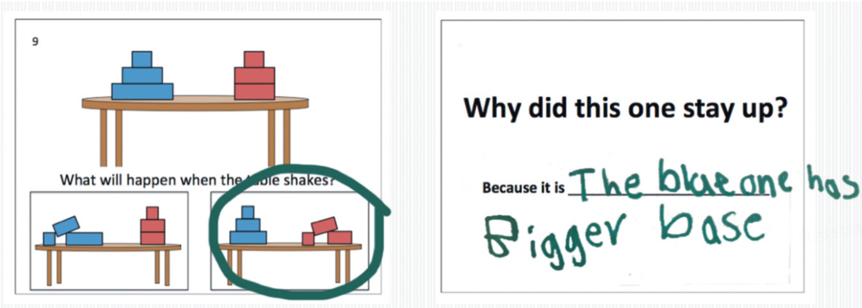

Figure 8. Prediction (left) and explanation (right) items used in the paper pre/posttests.

explanations were also coded based on their mentioning of the principles related to height, symmetry, center of mass and base. For example, if they responded to a question about height principle saying "because it is taller," or a question about symmetry principle saying "because it is symmetrical" or "because it is the same on both sides" their answer was coded to be correct.

Paper posttests were developed based on the concepts covered in the National Research Council's Framework & Asset Science Curriculum (National Research Council, 2012), and targeted the four principles of balance: symmetry, wide base, height, and center of mass. Questions presented a picture of two towers on a table, and asked students to indicate which tower will fall when the table shakes. Prediction items asked students to select which tower would fall first, and explanation items asked for their reasoning (Figure 8). These questions were phrased as "What will happen when the table shakes?" and "Why did this one stay up?" Counting prediction and explanation items as individual questions, the tests had 9 questions in total. If the kids had any trouble reading or writing, the experimenter helped them read or write.

A previous study showed that the tasks worked for different age groups in elementary school and the pre/posttest measures produce a good range of outcomes (e.g., not close to ceiling) to distinguish learning outcomes for students at these ages (Yannier et al., 2016).

Procedure

Before interacting with the exhibit, children were individually given a tower building task, which took approximately 3–5 minutes. They were asked to use a given set of blocks (using all the blocks in the set) to build a tower that would stay up when the earthquake table shook. Students were told to use a specific block as the base of the tower as a challenge for making a small base. They did not test their towers before interacting with the exhibit. They were told that they would get a chance to test their towers after they play the game.

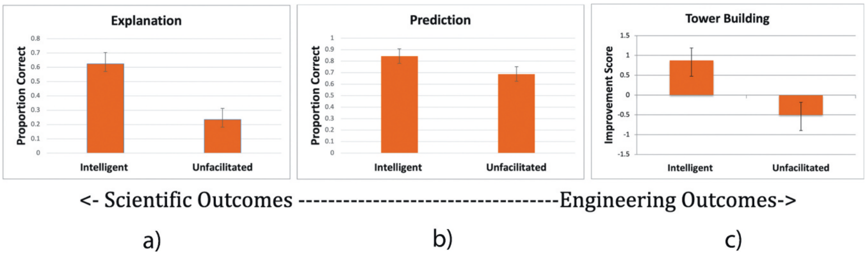

Figure 9. Learning results of a random assignment experiment (N=32). In comparing outcomes from the Intelligent Science Exhibit versus those from the Unfacilitated Exhibit, children were a) significantly better able to explain scientific principles, b) marginally better able to predict scientific outcomes, c) significantly better at engineering stable towers even though children did more tower building with the Unfacilitated Exhibit.

The pair of children then completed their assigned condition, either the Intelligent Science Exhibit or the Unfacilitated Exhibit. The timing was matched for both conditions based on the average time it takes for children to go through the tasks in the Intelligent Science Exhibit (~15 minutes).

After interacting with the exhibit, students were given the same tower building task as before. This building task allowed us to measure any changes in their towers after interacting with the game. After building the tower, they were given a paper posttest consisting of the prediction and explanation items to measure their understanding of scientific principles.

Statistical inference methods

We report statistical inferences in two ways. To protect against potential deviations from parametric assumptions, we use the non-parametric Mann-Whitney test. To provide the comparison value of effect size, we use the parametric t-test.[1]

Results on learning: The intelligent science exhibit improved both scientific and engineering outcomes

As summarized in Figure 9, the Intelligent Science Exhibit group had higher correctness on the posttests of explanation and prediction and a higher improvement score for tower building. The explanation posttest served to estimate science learning outcomes (Figure 9a). On the explanation posttest,

[1]"Parametric test assumptions were met for five of six tests of normality and two of three tests of homogeneity of variance. The two exceptions were based on a ceiling effect for the Intelligent Exhibit condition on the prediction outcome producing below-threshold normality and a floor effect for the Unfacilitated Exhibit on the explanation test producing below-threshold homogeneity of variance."

we found a significant main effect of the *Intelligent Science Exhibit Condition* (M = 0.63, SD = 0.32, Mdn = 0.75) over the *Unfacilitated Exhibit Condition* (M = 0.23, SD = 0.21, Mdn = 0.25) by both a Mann-Whitney test (U = 43, p = .001) and a t-test (t(31) = 5.95, p < .001) with a large effect size (d = 1.45). Thus, the guided inquiry, facilitated by the system's AI vision and automated feedback, improves children's learning and understanding of scientific principles. We did not see any significant age or gender effects based on an ANCOVA analysis with age and gender as covariates.

The prediction posttest and the tower building improvement score served to estimate engineering outcomes. For the prediction tests (Figure 9b), we found a marginal effect of the *Intelligent Science Exhibit Condition* (M = 0.84, SD = 0.26, Mdn = 1.00) over the *Unfacilitated Exhibit Condition* (M = 0.69, SD = 0.25, Mdn = 0.75) by both a Mann-Whitney test (U = 80, p = .073) and a t-test (t(31) = 1.91, p = .066) with a substantial effect size (d = 0.62). The marginal effect on prediction items may be a consequence of extra noise due to guessing given there are only two choices (left tower or right tower—see, Figure 8). In contrast, the explanation and tower-building questions are open-ended.

For the tower building improvement score (see, Figure 9c), we found a significant main effect of the *Intelligent Science Exhibit Condition* (M = 0.875, SD = 1.26, Mdn = 1.00) over the *Unfacilitated Exhibit Condition* (M = −0.50, SD = 1.59, Mdn = −1.00) by both a Mann-Whitney test (U = 65, p = .02) and a t-test (t(31) = 2.44, p < .05) with a large effect size (d = 0.96). These results are particularly compelling given that children actually do more tower building with the *Unfacilitated Exhibit* than they do with *Intelligent Science Exhibit*. In fact, without guidance, there was no improvement in tower building from the *Unfacilitated Exhibit*.

Both the scientific and engineering outcomes were significantly higher for the *Intelligent Science Exhibit* condition than the *Unfacilitated Exhibit* condition. This result suggests that children having the Intelligent Science Exhibit support were not only having a better understanding of the physical phenomena, but they were also able to apply their knowledge better to an engineering, constructive problem-solving task.

Results on enjoyment: Guided discovery in intelligent science exhibits is just as fun

One possible concern with intelligent exhibits with guidance is whether visitors enjoy them as much as exhibits that allow for unstructured exploration. To measure their enjoyment while interacting with the exhibit, children were given a survey with 3 questions after interacting with the exhibit: "How much did you like the game?," "Would you like to play it again?," and "Would you recommend it to your friend?" They answered each question on a scale of 5 (Likert scale of 1–5 with 1 being the lowest and 5 being the highest).

Our evidence suggests that children enjoyed both conditions. Based on the three-question survey described above, we found no statistical difference between average enjoyment scores (0.87 vs. 0. 85 out of 1; $p = .7$).

Exhibit stickiness analysis/study: Does adding intelligence produce longer visits?

In addition to evaluating learning and enjoyment outcomes, we also evaluated how engaging or "sticky" the Intelligent Science Exhibit is compared to the Unfacilitated Exhibit. To do so, we observed how visitors interacted with the two exhibit versions during regular museum hours, in contrast to the previous experiment described above which was a pull-out study. We made one or the other exhibit available at the same location in the Science Center at selected times over two weeks to observe how and how much the general public would interact with them.

Methods: Exhibit logs

We tracked how much time visitors spent interacting with the Intelligent Science Exhibit and compared that to the time they spent with the Unfacilitated Exhibit (by turning on and off the intelligent features on the same exhibit/location). We looked at similar corresponding time frames (1–2 hour slots each day) in each condition over two weeks to see how much time visitors spent on average.[2] Five days of paired time periods were sampled for a total of 6.0 hours with the Unfacilitated Exhibit and 7.4 hours with the Intelligent Science Exhibit (including 2 corresponding sessions of ~2 hours and 3 corresponding sessions of ~1 hour). We report usage in terms of the proportion of available time in which the exhibit is occupied and the average length of a visit, neither of which is impacted by the difference in total time. We used the following method to estimate how long visitors spent at an exhibit from the log data. A visit ends at the last logged action before the system goes idle (no intervening logged actions) for more than 1 minute or at the end of a session (e.g., 3:41, 4:47 in the first example above). A visit starts at the first logged action after the end of a previous visit or at the start of a session (e.g., 2:45, 3:48 in the first example). A visit length is calculated as the

[2]For example, we sampled from the log data interactions on July 4 (a Tuesday) when the Intelligent Science Exhibit was available, first, from 2:45 to 3:41 pm and the Unfacilitated Exhibit was available, after, from 3:48 to 4:47pm and then on July 9 (a Sunday) when the Unfacilitated Exhibit was available, first, from 1:02 to 1:44 pm and the Intelligent Science Exhibit was available, after, from 1:55 to 2:57 pm, etc.

difference between the times of the visit start and end. An unoccupied period is defined as the time remaining outside of the start and end times of each visit.

Results on engagement: The intelligent science exhibit is stickier

The percentage of time that the exhibits were occupied (i.e., visit time divided by total time) was 56% for the Intelligent Science Exhibit (4.16 hours of 7.44 hours) versus 27% for the Unfacilitated Exhibit (1.61 hours of 6.03 hours). In the 4.16 hours that the Intelligent Science Exhibit was occupied, there were 40 different visits. This yields an average of 6.25 minutes per visit (4.16*60/40). In the 1.61 hours that the Unfacilitated Exhibit was occupied, there were 64 different visits. This yields an average of 1.51 minutes per visit (1.61*60/64), perhaps not enough time to learn much. Note that longer visits necessitate somewhat fewer visitors. The difference between the average 6.25 minutes spent on the Intelligent Science Exhibit and the average 1.51 minutes spent with the Unfacilitated Exhibit is statistically reliable ($t(102) = 5.29$, $p < .001$). In sum, not only do visitors stay longer with the Intelligent Exhibit, more than four times longer, the Intelligent Exhibit twice as likely to be in use (56% vs. 27%).

Bolstering these quantitative results on stickiness, we also observed that the kids seemed to be more persistent while interacting with the Intelligent Exhibit. If their prediction was wrong or if their tower fell over, they would try again and again. They seemed to be resilient against failure, as the gorilla said “Uh oh, your tower fell down. Please try again to make another tower!.” On the other hand, with the Unfacilitated Exhibit, such persistence and resilience were not apparent. Most visitors to the Unfacilitated Exhibit interacted for a very short time—mostly they built a tower, clicked shake to see if it stayed up, and then left. Rarely, we saw a few visitors making a second tower after the first one fell.

Feedback from parents or guardians

To get a sense for why the Intelligent Science Exhibit was more engaging, we gave a short informal survey to a small sample of 9 parents or guardians who interacted with the Intelligent Science Exhibit. Parents/guardians were identified during the day as children were interacting with the exhibit. The survey had 7 questions including: “We’re testing a new exhibit. Please let us know your thoughts.,” “Would you like to see this as a permanent exhibit? Why or why not?,” “How would you compare this exhibit to others in the museum?”

The parent/guardian survey indicated that the interactive character of the exhibit was appealing and helped hold the attention of the kids. One parent wrote “I especially liked the making predictions part of it, and then having to come up with an explanation of the results.” She elaborated that “It employs inquiry learning, which is the heart of how kids learn. But it is also a play model, so it does not seem like a learning activity.” When asked to compare the Intelligent Science Exhibits with other exhibits, parents/guardians indicated it is “more interactive with the prompts and monitor,” “It’s more directed and instructional than most other exhibits” and “Compared to similar exhibits, this one has 2-way communication. I think it’s better.” Parents/guardians also commented on how the guidance was pitched at the right level, for example, “For my kids age (5), it is a top exhibit. Minimal reading (which is good for older visitors), so it is great to engage early elementary-age kids, and still allows them to do science.” Parents/guardians’ comments indicated that the guidance provided in the system was a key reason for kids staying engaged while they were interacting with the exhibit.

General discussion

We found that adding guided inquiry scaffolding facilitated by having an intelligent layer on top of physical experimentation in traditional hands-on exhibits helps children formulate better, more scientific theories of the physical phenomena they experience. Our results demonstrate children receiving interactive guidance and feedback while predicting and explaining are better able to learn to *apply* science in engineering tasks, especially when combined with exploration and construction activities. Importantly, free exploration and construction on its own, as in many current exhibits in museums, led to substantially less learning and, for some measures, none at all. Adding the guided-discovery support and intelligent layer that the mixed-reality AI system provides, not only fosters better learning of scientific principles of stability, but also improves the application of those principles in a hands-on, constructive engineering task.

As indicated in the introduction, the Intelligent Science Exhibit utilizes learning support strategies based on prior research such as contrasting cases (Chase et al., 2010), predict-observe-explain prompting (White & Gunstone, 1992), menu-based self-explanation prompting (Aleven & Koedinger, 2002; Chi et al., 1989), and interactive feedback (Corbett & Anderson, 2001). These strategies do not tell the science to students but promote relevant thinking so that students construct accurate scientific conceptions on their own. The contrasting case towers help children focus on scientifically relevant features and the self-explanation prompts get children to attend more closely to these

Figure 10. Child predicts that one of the contrasting case towers has more weight on the side so it will fall more quickly compared to the other one.

features and to put them into words. Figure 10 shows a contrasting case regarding symmetry where the children have predicted the left tower will fall first. When the gorilla asks why this tower will fall first, a child answers: “because this one [pointing to the left tower as shown] has more weight on the side and that one [pointing to the right tower] doesn’t.”

Sometimes these articulations of scientifically relevant features are started by one child and picked up by another. For example, faced with a case contrasting lower versus higher center of mass, one child explains “There is no blocks under it [pointing at the tower with more weight on top],” while another girl adds: “Yeah … And that will keep it up [pointing to the lower supporting part of the tower].” So, the contrasting case towers help focus children’s attention on the relevant scientific features that are different between the towers causing one to fall more quickly than the other.

Feedback from the system—“Uh oh! Your prediction was not quite right”—helps children focus their attention. It highlights an opportunity to learn. The prompted self-explanation supports this opportunity to learn, when the gorilla follows up “The right tower fell first. Why do you think this tower fell first?” The Gorilla provides alternative possible explanations: “Because it’s taller? Because it has a thinner base? Because it has more weight on top than bottom? Because it is not symmetrical or same on both sides?.” This menu-based self-explanation is an interesting variation on the “time for telling” idea (Schwartz

& Bransford, 1998). Similar to that approach, children's incorrect prediction puts them in a more receptive mode than when up-front telling is provided. But importantly, the Gorilla is not simply telling. Children still need to figure out which explanation works in this situation. A child's self-correction in this context is indicative of active sense making: "Because it's taller? Oh, it has a thinner base!." The gorilla's feedback on the clicked-on explanation, "Good job! It fell first because it had a thinner base than the other tower.," helps the child confirm their insight, as indicated by a "we did it" fist pump.

The sequencing of scientific inquiry before the engineering activity of building stable towers seems to afford transfer of learning. Concepts children pick up during inquiry are incorporated into their building efforts. For example, in "Test My Tower" mode one pair used the principles they had learned from the Guided-discovery mode, saying "Let's use this (pointing at the long thin blocks). It has a big base. The bigger the base, the better it's gonna be!"

The overall learning outcome and engagement results suggest that the *Intelligent Exhibit* evoked both playful and thoughtful interactions, whereas the *Unfacilitated Exhibit* evoked play but less thought. When children built and tested towers with the Unfacilitated Exhibit, they were not incorporating principles of balance or spontaneously designing experiments to understand the underlying principles. Instead, they focused on building from intuition, testing, and then tweaking, making changes to their towers with no apparent systematic plan. As shown in Figure 11, their towers tended not to incorporate underlying principles. In rare cases where they built multiple towers at the same time, they did not express comparison goals or explicitly reflect on comparative outcomes.

Figure 11. Children's constructed towers in the Unfacilitated Exhibit tended to not incorporate the underlying scientific principles of balance, being all and thin (left image), asymmetrical (right), and with not more weight near the bottom than the top (both).

While the Intelligent Science Exhibit provides many forms of guidance and scaffolding, it is not over-scaffolding. First, it provides for some fading of guidance as students transition from the Guided Discovery mode to the minimally guided Test My Tower mode. Second, our post-assessments involve no guidance and provide evidence that student learning from the Intelligent Science Exhibit transfers to a context without guidance, as students may experience in the real world or in future science learning activities. Furthermore, our Guided Discovery mode is different from direct instruction, in that children are guided through an inquiry task where they are discovering the reasons behind physical phenomena with interactive feedback rather than being given the answers upfront.

These results add value to popular views and theories of learning that are apparent in museums, maker spaces and other learning environments, that have a strong focus on open-ended hands-on exploratory learning (Jeffery-Clay, 1998). That value derives, in part, from articulating and testing more precise and operational scientific descriptions of key elements of active and constructive learning that can be most reliably replicated (cf., Klahr et al., 2009). Ideas for what elements constitute effective constructive learning vary (e.g., Chi & Wylie, 2014; Kafai & Resnick, 1996; Resnick, 2014). One element some authors emphasize is the production of a public or externalized product, for example, "producing additional externalized outputs or products beyond what was provided" (Chi & Wylie, 2014) and "the learner is consciously engaged in constructing a public entity" (Papert, 1980). A consequent prediction in our context is that the tower building activities in the *Unfacilitated Exhibit* should provide greater learning benefits than the prediction and selection-based explanations in the *Intelligent Science Exhibit*. The former produces an "additional externalized output" whereas the latter does not. Our findings are inconsistent with this prediction.

The *Unfacilitated Exhibit* was not better on any of our multiple learning measures. The *Intelligent Science Exhibit* condition performed significantly better than the *Unfacilitated Exhibit* in overall principle test results and, surprisingly, in transferring to hands-on activities in the Tower-Building tests that are directly analogous to the constructive activities in the *Unfacilitated Exhibit condition*. This result is especially interesting, as the *Unfacilitated Exhibit* condition is similar to how many museum exhibits and maker spaces are designed. The emphasis on construction and open-ended exploration, which is good when combined with well-designed instruction (cf., Yannier et al., 2021), can get misinterpreted as a recommendation for construction and open-ended exploration with no explicit guidance/instruction.

Our experiment also adds value to a cluster of theoretical notions suggesting that one learns what one practices. Related theoretical concepts and recommendations include transfer-appropriate processing (Bransford et al., 2000), "match the job task" (Clark & Mayer, 2016) or "active learning" such that students become better physicists if they "practice physicist like reasoning" (Deslauriers et al., 2011). The straightforward interpretation of these recommendations correctly predicts that children interacting with the *Intelligent Science Exhibit*, practicing explanation, should have better outcomes on the explanation measures, compared to those in the *Unfacilitated Exhibit* condition, who do not practice explanation. However, a straightforward application *incorrectly* predicts that better outcomes on tower building should result from the *Unfacilitated Museum Exhibit* condition, which practices more tower building whereas the *Intelligent Science Exhibit* condition does not. We find that the scientific thinking that is enhanced through guided inquiry practice transfers to the engineering task of tower construction better than the thinking elicited by tower construction practice. Thus, the theoretical value added by this work is both a caution against using task similarity alone in predicting transfer and a recommendation to also evaluate the similarity in the underlying thinking that instructional tasks evoke and that assessment tasks require. Such evaluation can be done, for example, using methods like cognitive task analysis (cf., Koedinger & McLaughlin, 2010; Lovett, 1998; Velmahos et al., 2004).

In addition to learning benefits, we also found the Intelligent Science Exhibit kept visitors engaged for longer periods of time. Visitors spent four times more time at the Intelligent Science Exhibit voluntarily compared to the Unfacilitated Exhibit. Qualitatively, we observed more persistence among visitors interacting with the Intelligent Science Exhibit, trying again and again to get their towers to stay up when the earthquake happened whereas in the Unfacilitated Exhibit such persistence and resilience were not apparent. This extra time and persistence suggest potential further learning benefits when Intelligent Science Exhibits are deployed in a museum. The large learning gains we saw in the experimental study where time was controlled may be multiplied in a museum setting where visitors stick with the Intelligent Science Exhibit much longer than a matched Unfacilitated Exhibit.

Intelligent Science Exhibits automate effective forms of scaffolding and guidance, in a way that affords both reliable experimental variation and wider dissemination, but without sacrificing the benefits of hands-on physical experimentation in the real world. This interactive support is especially important in museums where it's not always feasible to have knowledgeable staff around the exhibits and where parents/guardians do not always have relevant knowledge. We demonstrated how guidance and personalized support provides for better learning outcomes. Because the AI technology in Intelligent Science Exhibits automates this support, it can be provided to more children of different backgrounds and at a wider scale. While the study in this paper was conducted when there was no facilitator or museum staff present, the AI technology can also be a model when there are facilitators or parents around, exemplifying how they can guide the children for better learning outcomes.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the NSF [1612744,1701107]

References

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, *26*(2), 147–179. https://doi.org/10.1207/s15516709cog2602_1

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. *Handbook of Research on Learning and Instruction*, 522–560. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1207/s15516709cog2602_1

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *103*(1), 1. https://doi.org/10.1037/a0021017

Alimisis, D. (2013). Educational robotics: Open questions and new challenges. *Themes in Science and Technology Education*, *6*(1), 63–71.

Anderson, D. R., Bryant, J., Wilder, A., Santomero, A., Williams, M., & Crawley, A. M. (2000). Researching blue's clues: Viewing behavior and impact. *Media Psychology*, *2*(2), 179–194. https://doi.org/10.1207/S1532785XMEP0202_4

Bamberger, Y., & Tal, T. (2007). Learning in a personal context: Levels of choice in a free choice learning environment in science and natural history museums. *Science Education*, *91*(1), 75–95. https://doi.org/10.1002/sce.20174

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, *13*(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1

Benko, H., Jota, R., & Wilson, A. (2012, May). MirageTable: Freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 199–208). https://dl.acm.org/doi/pdf/10.1145/2207676.2207704?casa_token=8IKRibF_RzgAAAAA:H-snuBR1vRx9HfLLpRQEqMrYA03xFXSkpvLtfFJ-Mw0axb9_YW-iQ_BMX4Pnd1LDiqSfMa8xpfs

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (Vol. 11). Washington, DC: National academy press .

Cassell, J. (2000). *Embodied conversational agents*. MIT press.

Chase, C. C., Shemwell, J. T., & Schwartz, D. L. (2010). Explaining across contrasting cases for deep understanding in science: An example using interactive simulations. *Proceedings of the 2010 International Conference of the Learning Sciences*. https://repository.isls.org/bitstream/1/2665/1/153-160.pdf

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1

Chi, M., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Christel, M. G., Stevens, S. M., Maher, B. S., Brice, S., Champer, M., Jayapalan, L., and Zhang, X. (2012, July). RumbleBlocks: Teaching science concepts to young children through a unity game. In *2012 17th International Conference on Computer Games (CGAMES)* (pp. 162–166). IEEE. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6314570&casa_token=8Ux94beEcXAAAAAA:eLqLYvj20lOPwzrfUTkhGEIL126PsiB6ALI4tJEZOwcOcEMOY2bK5TH

Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons.

Cody, A. (2010). *Sharing our visions: What does great education look like to you?* Education Week.

Corbett, A. T., & Anderson, J. R. (2001, March). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 245–252). ACM. https://dl.acm.org/doi/pdf/10.1145/365024.365111?casa_token=-2o62h5-R4cAAAAA:Xq_DtlNqks31iny1o1GkoQS9HGMmjzcmZeBCnoE51SSJOsDq6ssnuVSCXC2g7gcOmtXtmOFnIlo

Crowley, K., Callanan, M. A., Jipson, J. L., Galco, J., Topping, K., & Shrager, J. (2001). Shared scientific thinking in everyday parent-child activity. *Science Education*, *85*(6), 712–732. https://doi.org/10.1002/sce.1035

De Jong, T., & Lazonder, A. W. (2014). 15 The guided discovery learning principle in multimedia learning. *The Cambridge Handbook of Multimedia Learning*, 371.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5–8. https://doi.org/10.3102/0013189X032001005

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, *332*(603), 862–864. https://doi.org/10.1126/science.1201783

Eberbach, C. (2009). *The effect of parents' conversational style and disciplinary knowledge on children's observation of biological phenomena* [Doctoral dissertation]. University of Pittsburgh.

Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine*, *15*(11), 988–994. https://doi.org/10.1111/j.1553-2712.2008.00227.x

Eysink, T. H., Gersen, L., & Gijlers, H. (2015). Inquiry learning for gifted children. *High Ability Studies*, *26*(1), 63–74. https://doi.org/10.1080/13598139.2015.1038379

Falloon, G. (2010). Using avatars and virtual environments in learning: What do they have to offer? *British Journal of Educational Technology*, *41*(1), 108–122. https://doi.org/10.1111/j.1467-8535.2009.00991.x

Fender, J. G., & Crowley, K. (2007). How parent explanation changes what children learn from everyday scientific thinking. *Journal of Applied Developmental Psychology*, *28*(3), 189–210. https://doi.org/10.1016/j.appdev.2007.02.007

Franse, R. K., van Schijndel, T. J., Plankman, T. I., & Raijmakers, M. E. (2021). Families' experiments and conversations at an open-ended exhibit in a science museum: Individual characteristics and the influence of minimal guidance strategies. *Science Education*, *105*(4), 707–742. https://doi.org/10.1002/sce.21620

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, *82*(3), 300–329. https://doi.org/10.3102/0034654312457206

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62*(1), 32. https://doi.org/10.1037/h0048826

Gick, M. L., & Paterson, K. (1992). Do contrasting examples facilitate schema acquisition and analogical transfer? *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *46*(4), 539. https://doi.org/10.1037/h0084333

Gormally, C., Brickman, P., Hallar, B., & Armstrong, N. (2009). Effects of inquiry-based learning on students' science literacy skills and confidence. *International Journal for the Scholarship of Teaching and Learning*, *3*(2), n2. https://doi.org/10.20429/ijsotl.2009.030216

Gupta, A., Fox, D., Curless, B., & Cohen, M. (2012, October). DuploTrack: A real-time system for authoring and guiding duplo block assembly. In *Proceedings of the 25th annual ACM symposium on User interface software and technology* (pp. 389–402). https://dl.acm.org/doi/pdf/10.1145/2380116.2380167?casa_token=ofFg-yXTcM4AAAAA:Krhr359TWcOhL0_244RmE4Pv6jRrCU_xgwSgreUE_d7novTrSCNT9p4nyzBMcZHzuztBQnkbyLo

Hammady, R., & Ma, M. (2019). Designing spatial ui as a solution of the narrow fov of microsoft hololens: Prototype of virtual museum guide. In *Augmented reality and virtual reality* (pp. 217–231) Springer. https://pure.solent.ac.uk/ws/files/10722373/Designing_Spatial_UI_as_a_Solution_of_the_Narrow_FOV_of_Microsoft_HoloLens_Prototype_of_Virtual_Museum_Guide_compressed.pdf

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and. *Educational Psychologist*, *42*(2), 99–107. https://doi.org/10.1080/00461520701263368

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning.* Center for Curriculum Redesign.

Hovious, A. (2015), *The recipe for successful makerspace design.* https://designerlibrarian.wordpress.com/2015/10/19/the-recipe-for-successful-makerspace-design/

Jeffery-Clay, K. R. (1998). Constructivism in museums: How museums create meaningful learning environments. *Journal of Museum Education*, *23*(1), 3–7. https://doi.org/10.1080/10598650.1998.11510362

Kafai, Y. B., & Resnick, M. (1996). *Constructionism in practice: Designing, thinking, and learning in a digital world*. Routledge.

Kearney, M. (2004). Classroom use of multimedia-supported predict–observe–explain tasks in a social constructivist learning environment. *Research in Science Education*, *34*(4), 427–453. https://doi.org/10.1007/s11165-004-8795-y

Kim, N. J., Belland, B. R., & Walker, A. E. (2018). Effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education: Bayesian meta-analysis. *Educational Psychology Review*, *30*(2), 397–429. https://doi.org/10.1007/s10648-017-9419-1

Kim, K. Y., & Crowley, K. (2010). Negotiating the goal of museum inquiry: How families engineer and experiment. In *Instructional explanations in the disciplines* (pp. 51–65) Springer. http://www.upclose.pitt.edu/articles/kim%20and%20crowley.pdf

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Richard e. Clark,"Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential and inquiry-based teaching." *Educational Psychologist*, *41*(2), 2. https://doi.org/10.1207/s15326985ep4102_1

Klahr, D., Ambady, N., Eccles, J., Gomez, L., Klahr, D., Linn, M., Miller, K., & Mix, K. (2009). "To every thing there is a season, and a time to every purpose under the heavens": What about direct instruction? *The American Psychologist*, *64*(6), 538–550. https://doi.org/10.1037/a0014813

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Journal of Phsicological Science*, *2*(1). https://journals.sagepub.com/doi/pdf/10.1111/j.0956-7976.2004.00737.x?casa_token=Q8z9V1N15W4AAAAA%3A-ZQ5esW6xXNwA_FuEgfVmDgglhptkKhC4NsU6yetQnBK2UTE8SJ7MdWmhzf_OW9sAt0OOwsXY5k&

Klahr, D., Triona, L. M., & 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

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, *342*(6161), 935–937. https://doi.org/10.1126/science.1238056

Koedinger, K., & McLaughlin, E. (2010). Seeing language learning inside the math: Cognitive analysis yields transfer. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32 No.32). https://escholarship.org/content/qt2n8407gq/qt2n8407gq.pdf

Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., Puntambekar, S., & Ryan, M. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design (tm) into practice. *The Journal of the Learning Sciences*, *12*(4), 495–547. https://doi.org/10.1207/S15327809JLS1204_2

Laput, G., & Harrison, C. (2019, May). SurfaceSight: A New Spin on Touch, User, and Object Sensing for IoT Experiences. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). https://dl.acm.org/doi/pdf/10.1145/3290605.3300559?casa_token=Fixn5xqgbG4AAAAA:PkAWl-jrHkb9WVIjS23RqQQeNXpD7GfWBsVK3DmiSKyTa_ZW1VXvg-rEFAnzmABv6aMvo5pzkvM

Lovett, M. C. (1998). Cognitive task analysis in service of intelligent tutoring system design: A case study in statistics. In *International Conference on Intelligent Tutoring Systems* (pp. 234–243). Springer, Berlin, Heidelberg.

Ma, J., Shahar, E., Eliceiri, K., Mehta, G., & Yu, K. (2019, June). Shedding light: Integrating bioimaging technologies into the design of an interactive museum exhibit. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (pp. 837–848). https://dl.acm.org/doi/pdf/10.1145/3322276.3322299

Manches, A., & O'malley, C. (2012). Tangibles for learning: A representational analysis of physical manipulation. *Personal and Ubiquitous Computing*, *16*(4), 405–419. https://doi.org/10.1007/s00779-011-0406-0

Manches, A., & Price, S. (2011, June). Designing learning representations around physical manipulation: Hands and objects. In *Proceedings of the 10th International Conference on Interaction Design and Children* (pp. 81–89). https://dl.acm.org/doi/pdf/10.1145/1999030.1999040?casa_token=CVcCyKUTab8AAAAA:s2eC8iqIyv1N1342LNMQHc1KfIErBVWfbotefwpSlGYSrBvwu_eusraFgUQivsTt2wC_8TRyIGw

Marshall, P. (2007, February). Do tangible interfaces enhance learning? In *Proceedings of the 1st international conference on Tangible and embedded interaction* (pp. 163–170).

Marshall, P., Rogers, Y., & Hornecker, E. (2007). Are tangible interfaces really any better than other kinds of interfaces? In: CHI'07 workshop on Tangible User Interfaces in Context & Theory, 28 Apr 2007, San Jose, California, USA. http://oro.open.ac.uk/19535/1/TUIworkshop-Marshall.pdf

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

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

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc.

Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://www.sciencedirect.com/science/article/pii/S1747938X15000068

Puchner, L., Rapoport, R., & Gaskins, S. (2001). Learning in children's museums: Is it really happening? *Curator: The Museum Journal*, *44*(3), 237–259. https://doi.org/10.1111/j.2151-6952.2001.tb01164.x

Resnick, M. (2014). Give P's a chance: Projects, peers, passion, play. *Constructionism and creativity: Proceedings of the Third International Constructionism Conference*. Austrian Computer Society, Vienna.

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, *101*(3), 529. https://doi.org/10.1037/a0014224

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475–5223. https://doi.org/10.1207/s1532690xci1604_4

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759. https://doi.org/10.1037/a0025140

Schwartz, D. L., Tsang, J. M., & Blair, K. P. (2016). *The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them*. WW Norton & Company.

Sottilare, R. A., Baker, R. S., Graesser, A. C., & Lester, J. C. (2018). Special issue on the Generalized Intelligent Framework for Tutoring (GIFT): Creating a stable and flexible platform for innovations in AIED research. *International Journal of Artificial Intelligence in Education*, *28*(2), 139–151. https://doi.org/10.1007/s40593-017-0149-9

Timms, M. J. (2016). Letting artificial intelligence in education out of the box: Educational cobots and smart classrooms. *International Journal of Artificial Intelligence in Education*, *26*(2), 701–712. https://doi.org/10.1007/s40593-016-0095-y

Velmahos, G. C., Toutouzas, K. G., Sillin, L. F., Chan, L., Clark, R. E., Theodorou, D., & Maupin, F. (2004). Cognitive task analysis for teaching technical skills in an inanimate surgical skills laboratory. *The American Journal of Surgery*, *18*(1), 114–119. https://doi.org/10.1016/j.amjsurg.2002.12.005

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, *53*(4), 5–23. https://doi.org/10.1007/BF02504682

White, R., & Gunstone, R. (1992). Prediction-observation-explanation. *Probing Understanding*, 4, 44-64.

Yannier, N. (2016). *Bridging physical and virtual learning: A mixed-reality system for early science* [Doctoral dissertation]. Carnegie Mellon University.

Yannier, N., Hudson, S. E., & Koedinger, K. R. (2020). Active learning is about more than hands-on: A mixed-reality AI system to support STEM education. *International Journal of Artificial Intelligence in Education*, 30(1), 74-96. https://par.nsf.gov/servlets/purl/10151045

Yannier, N., Hudson, S. E., Koedinger, K. R., Hirsh-Pasek, K., Golinkoff, R. M., Munakata, Y., Doebel, S., Schwartz, D. L., Deslauriers, L., McCarty, L., Callaghan, K., Theobald, E. J., Freeman, S., Cooper, K. M., & Brownell, S. E. (2021). Active learning: "Hands-on" meets "minds-on." *Science*, *374*(6563), 26–30. https://doi.org/10.1126/science.abj9957

Yannier, N., Hudson, S. E., Wiese, E. S., & Koedinger, K. R. (2016). Adding physical objects to an interactive game improves learning and enjoyment: Evidence from EarthShake. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *23*(4), 26. https://doi.org/10.1145/2934668

Yannier, N., Koedinger, K. R., & Hudson, S. E. (2013, July). Tangible collaborative learning with a mixed-reality game: Earthshake. In *International Conference on Artificial Intelligence in Education* (pp. 131–140). Springer, Berlin, Heidelberg.

Yannier, N., Koedinger, K. R., & Hudson, S. E. (2015). Learning from mixed-reality games: Is shaking a tablet as effective as physical observation? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1045–1054). ACM. Seoul, Korea https://dl.acm.org/doi/pdf/10.1145/2702123.2702397

Yoon, S. A., Elinich, K., Wang, J., Steinmeier, C., & Tucker, S. (2012). Using augmented reality and knowledge-building scaffolds to improve learning in a science museum. *International Journal of Computer-Supported Collaborative Learning*, *7*(4), 519–541. https://doi.org/10.1007/s11412-012-9156-x

Zhang, Y., Yang, C., Hudson, S. E., Harrison, C., & Sample, A. (2018, April). Wall++ Room-scale interactive and context-aware sensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). https://dl.acm.org/doi/pdf/10.1145/3173574.3173847?casa_token=jET4NfwY01oAAAAA:g6ya7YSDDYGc2ZC4jrHj30V4GBzSBIZAmw-pGZ3PBt9n9yk0Ykoz2MTxRe_iPH_m7D3y2wQgz00