Title: Active Learning Is About More Than Hands-On: A Mixed-Reality AI System to Support STEM Education

Authors: Nesra Yannier[1*], Scott E. Hudson[1], Kenneth R. Koedinger[1,2]

[*]Correspondence to: nyannier@andrew.cmu.edu, 412 680 6045

Affiliations: Human Computer Interaction Institute and Psychology Department, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213.

Abstract:

Along with substantial consensus around the power of active learning, comes some lack of precision in what its essential ingredients are. New educational technologies offer vehicles for systematically exploring benefits of alternative techniques for supporting active learning. We introduce a new genre of Intelligent Science Station technology that uses Artificial Intelligence (AI) to support children in learning science by doing science in the real world. We use this system in a randomized controlled trial that investigates whether active learning is best when it is implemented as guided deliberate practice, as constructive "hands-on" activity, or as a combination of both. Automated, reactive guidance is made possible by a specialized AI computer vision algorithm we developed to track what children are doing in the physical environment as they do experiments and discoveries with physical objects. The results support deliberate practice and indicate that having some guided discovery based on effective learning mechanism such as self-explanation, contrasting cases and personalized interactive feedback produces more robust learning compared to exploratory construction alone. Children learning through guided discovery achieve greater understanding of the scientific principles than children learning through hands-on construction alone (4 times more pre-to-post test improvement). Importantly, a combined guided discovery and hands-on construction condition leads to better learning of the very hands-on construction skills that are the sole focus of the hands-on constructive learning condition (>10 times more pre-to-post improvement). These results suggest ways to achieve powerful active learning of science and engineering that go beyond the widespread temptation to equate hands-on activity with effective learning.

Introduction

Certain forms of active learning have been shown to be more effective than traditional, more passive forms of instruction like lectures or readings (Deslauiers et al. 2011; Waldrop 2015; Freeman et al. 2014). The great promise of active learning is highlighted in (Deslauiers et al. 2011) where the results have shown increased student attendance, higher engagement, and more-than-double-learning when students were

taught using active learning instruction, compared to a traditional lecture. However, "active learning" gets used in a variety of ways indicating a lack of scientific precision (Klahr 2009; Tobias & Duffy 2009; Prince 2004). Some of these uses are at odds with each other with different researchers advocating distinctly different practices, in some cases emphasizing more open-ended "constructive" activities (Waldrop 2015; Freeman et al. 2014) and in other cases emphasizing more "deliberate" task design, formative assessment, and guidance (Deslauiers et al. 2011; Ericsson et al. 1993; Clark et al. 2011). This paper aims to contribute theoretical precision in defining and operationalizing active learning practices (Klahr & Li 2005), provide evidence for some elements and against others, and do so while leveraging an innovative AI-based educational mixed-reality technology we introduce.

Descriptions of active learning have made connections to elements of a theory of education called constructivism (Chi & Wylie 2014; Papert & Harel 1991; Kafai & Resnick 1996; Resnick 2014; Jeffrey-Clay 1998; Applefield et al. 2000; Cakir 2008; Van Joolingen 2005; Jonassen 1991; Steffe & Gale 1995). These include an emphasis on self-direction, whereby students "take charge of their education" (Freeman et al. 2014) on open-ended "challenging questions" (Deslauiers et al. 2011) and on "problem solving" (Waldrop 2015). Constructivism emphasizes open-ended hands-on exploration by creating or manipulating physical objects and development of new knowledge through sharing with others and reflection, and has become especially popular through the Maker Movement (Halverson & Sheridan 2014; Dougherty 2012). A precise scientific statement of this view is captured in the ICAP (Interactive, Constructive, Active and Passive) Framework (Chi & Wylie 2014), which suggests that learning environments that promote student construction are more effective than ones that promote activity without construction. They define "constructive behaviors as those in which learners generate or produce additional externalized outputs or products beyond what was provided in the learning materials." This notion of externalized outputs is further extended in the notion of construc*tionism* (Papert & Harel 1991; Resnick 2014; Papert 1980). Similarly, Edelson et al. make a direct connection between "open-ended exploration" and constructivist learning: "Scientific visualization tools can, therefore, provide the active, open-ended exploration that characterizes constructivist learning" (Edelson et al. 1996). Also, as Maker Spaces become more common,

many people assume that by virtue of "making," transformative learning is automatically taking place.", which is why we compare different forms of "making" to see what works best (https://designerlibrarian.wordpress.com/2015/10/19/the-recipe-for-successful-makerspace-design/).

Construct*ionism* shares construct*ivism*'s view of learning as building knowledge structures, but extends the idea to suggest that learning is especially effective when "the learner is consciously engaged in constructing a public entity" (Papert 1980). Similarly, Kolodner el al (Kolodner et al. 2003), refer to constructionism as follows: "Constructionism *(*Papert & Harel 1991; Kafai & Resnick 1996; Papert 1980) suggested that learners engage in design challenges and that they have a personally meaningful physical artifact to take home with them ...".

An alternative framing of active learning (Deslauiers et al. 2011) makes reference to the notion of deliberate practice (Ericsson et al. 1993), which recommends approaches to learning support that are potentially at odds with some constructivist recommendations. The idea of deliberate practice is to design tailored practice tasks should isolate challenging concepts and skills and come with specific feedback and repetition (Ericsson et al. 1993). The design of tailored practice tasks involves investigating what components of knowledge or skills are both critical for success and not easy to learn (Koedinger et al. 2012; Clark & Estes 1996). In other words, deliberate practice involves greater instructional structure and guidance. Evidence for the benefits of such guidance comes from a variety of sources (Waldrop 2015; Clark & Mayer 2011; Hattie 2009; Alfieri et al. 2011; Kirschner et al. 2006).

While sometimes simultaneously cited as key elements of active learning (Deslauiers et al. 2011), deliberate practice and constructivism have important differences. Deliberate practice emphasizes planned structure, particularly in the form of goal-specific task design, and immediate, explanatory feedback. Constructivism emphasizes creation of externalized outputs, the use of authentic or playful tasks, and warns against too much instruction. To illustrate some differences, consider some alternative types of scaffolding relevant to the experiment we present: Prompting students through a scientific inquiry process starting with given contrasting cases is more deliberate because the cases are chosen to address particular instructional goals. It is less constructive because, using the ICAP definition of constructive behavior,

students do not go "beyond what was provided in the learning materials" (Chi & Wylie 2014). Alternatively, prompting students to create the cases themselves or build their own structures is constructive. Asking students to explain an observed result of an experiment by selecting from a list of possible explanations is deliberate and active, but not constructive because students do not "generate or produce additional externalized outputs" (Chi & Wylie 2014). Our experiment provides a test of the ICAP prediction that the Constructive condition (Explore-Construct) should produce better learning than the Active condition (Guided-Discovery). As we will show, this prediction is contradicted.

It is worth noting substantial common ground in the learning support recommendations of deliberate practice and constructivism, particularly, focusing on engaging students in learning-by-doing and on more task-oriented or reactive guidance rather than extended up-front telling. We include these in all experimental conditions. Our goal is not to dispel the general merits of constructivism or deliberate practice, but to refine understanding of the effectiveness of particular variations within. We support a need (cf., Hmelo-Silver et al. 2007; Furtak et al. 2012) to move learning science beyond provocative binary distinctions (Kirschner et al. 2006) and beyond a third middle ground on a single guidance dimension (Furtak et al. 2012). Indeed, a combinatorial analysis of 30 instructional design dimensions suggests over 200 trillion options for supporting learning (Koedinger et al. 2013). We need more nuanced and more precisely-defined (cf., Klahr & Ji 2005; Chi & Wylie 2014) investigations within this vast space.

We see a general consensus on the importance of active learning (Deslauiers et al. 2011) and of engaging science students in inquiry activities with appropriate scaffolding (Furtak et al. 2012; Hmelo-Silver et al. 2007), but what elements of active learning and which types of scaffolding are most effective? Should we put more emphasis on exploration and construction or more emphasis on deliberately designed tasks with more explicit guidance? Furthermore, as augmented-reality and mixed-reality systems are getting more popular, can we differentiate which system features are more or less important toward enhancing learning outcomes? We investigated these questions in the context of a mixed-reality learning environment, a new genre of Intelligent Science Stations built on our mixed-reality AI system, designed to

help young children learn physical properties of stability while engaging in hands-on building/construction and/or guided scientific inquiry activities (see Figure 1).

Figure 1. Children interacting with the Intelligent Science Station in the Guided-Discovery condition, where they make predictions, observe results, and provide explanations with interactive feedback from an AI system that can see the results of their experiments. Children's engagement can be seen in a supplementary video.

This learning environment is made possible by the mixed-reality AI vision technology we have developed which allows the learning environment to *observe and interpret* students' actions -- to accurately monitor and evaluate student predictions, experiments, and explanations and provide an intelligent guided STEM learning experience for students with varied backgrounds and support (Yannier et al. 2015; Yannier et al. 2016). We add a new intelligent layer on top of physical experimentation in the real world, integrating intelligent tutoring systems into the 3D physical environment. As a result, Intelligent Science Station technology we introduce is able to leverage the complementary benefits of a physical experience and an intelligent virtual experience. Physical experience may make science real and engaging for some students, while intelligent virtual experience provides for instructional guidance through tracking student progress and providing feedback. Thus, this research makes two critical contributions: 1) an advanced AI mixed-reality technology for providing interactive instructional guidance in the context of real world experimentation and 2) use of this technology to explore fundamental learning science questions about the active ingredients in active learning.

Our prior experimental research has evaluated the importance of the physical experience and provides an indication that such experience is powerful even without it being hands-on and constructive. We demonstrated in randomized controlled trials that children learn much more from 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 (5 times more learning gain compared to a screen-only tablet or computer version) (Yannier et al. 2015; Yannier et al. 2016). We also demonstrated that adding simple hands-on, physical control such as shaking a tablet does not improve learning, whereas physical observation and experimentation does (Yannier et al. 2015; Yannier et al. 2016). Given notions of constructivism (Applefield et al. 2000) and embodied cognition (Shapiro 2010) and the focus on hands-on exploratory learning in informal learning settings (Jeffrey-Clay 1998), it seems plausible that a more fully hands-on condition would enhance learning.

The current experiment evaluates the importance of the guidance that is made possible by the AI vision algorithm and the associated reactive pedagogical interaction to support different forms of learning. In particular, we evaluated how important is the substantial implicit and explicit guidance that has been engineered into the design of the virtual experience surrounding the physical experience. Might students learn better with less explicit guidance and with more open-ended hands-on construction?

Why is technology critical?

To provide personalized immediate feedback to individual children and do so within a scientific apparatus allowing physical experimentation, we needed to extend the intelligent tutoring technology into the real world. We do this with new computer vision and AI technology that allow the system to recognize what children are doing as they set up experiments (with guidance) and match outcomes to their predictions and explanations. This technology allows the children to be given feedback on whether or not their predictions were correct and to help them understand the reasons why. The well-established phenomenon of confirmation bias (Nickerson 1998)

suggests that children may not readily notice that their predictions have been violated unless given such immediate feedback. Furthermore, having just-in-time feedback and self-explanation prompts aligned with the physical environment through AI technology helps children have a deeper understanding of the scientific phenomena they observe. Having the technical capability to track student activities and provide such feedback is a critical feature of an active learning process that the Intelligent Science Station supports.

This technology is crucial to improve learning in different settings where there may not always be enough support provided from teachers, parents or museum staff. For example in the museum setting, current exhibits rely solely on parents, signage and staff to provide support and guidance. It is not always feasible to have knowledgeable staff and not all parents have the same knowledge and background, so children receive varied support. The problem is similar in school settings where not all teachers have a science background to provide good support and guidance to students. Intelligent Science Stations open up a new opportunity to provide personalized interactive guidance and support to children, with different backgrounds and opportunities, through an intelligent AI layer integrated on top of physical hands-on experimentation, which may not be accessible to them otherwise.

Physical Setup

The physical setup of the Intelligent Science Station includes an earthquake table, physical towers placed on the table, a Kinect depth camera facing the objects, a projector, and a display screen with the computer game. The Kinect camera and our specialized computer vision algorithm detects when an object is placed on the table giving feedback if the placed towers match the correct ones or not (determining the shape and position of the object – since it is critical for students to place the correct contrasting case towers on the table so they can isolate

target variables) and when an action happens on the table (e.g. a tower falls), ensuring that the Intelligent Science Station is in sync with what is happening in the real world. The computer vision algorithm can also detect any tower that they build (made of any material such as lego blocks, cardboard etc.), giving different challenges (height/width) as they progress. It can also tell how many seconds it took for the tower to fall down and if the tower satisfies the different challenges that are given (e.g. if the base is thinner than a given width etc). The computer game that is displayed on a display screen provides visual and audio feedback to the user (e.g., noting which tower the student predicted would fall and which actually fell) (50). Our technology and teaching method provides personalized interactive feedback to the users as they experiment and make discoveries in their physical environment (Yannier et al. 2016).

The earthquake table consists of a small motor, a switch/relay, a mechanism for converting from rotary to reciprocating linear motion and rails to support the reciprocating platform. When the switch or the relay is activated, it activates the motor, which then moves the platform back and forth.

Computer Vision Algorithm

To develop an accurate and robust computer vision algorithm required some iterative experimentation and testing. Since there were no ready-made algorithms that would serve our purposes, we developed our own Computer Vision algorithms. In the first version, we used *color segmentation and depth information* to determine where the towers are located and to detect when they fall. Depth information reliably segregates the blocks from the background and eliminates conflicts that can arise when the background and blocks are similar colors. Blob tracking (Wang et al. 2008) is then used to track each segment of the colored blocks. The size and location of

these blobs are used to interpret the live state of the blocks on the screen. Finally, falls are detected when all blobs for a tower fall below a threshold height above the table.

Our first algorithm's reliance on color information caused problems in real world settings, as the lighting of the room negatively affected algorithm performance. Therefore, in the second version, we decided to rely on more depth information than color information since the depth information does not change according to lighting and is reliable in real world settings. In this version, again a depth image bitmap is extracted from the Kinect. We do filtering on the image to remove the background past a given depth. The blob tracking processor (Wang et al. 2008) is applied producing a list of blobs. To recognize a tower, we compute a Moment of Inertia (described next) for each blob and that value is compared to pre-defined xy-moments for each tower in a Tower Database. This method helps us identify each of multiple towers in an image. The tower whose xy-moment is closest in Euclidean distance to the blobs xy-moment is returned as the recognized tower.

The Moment of Inertia is a quantity expressing a body's tendency to resist angular acceleration. It is the sum of the products of the mass of each particle in the body with the square of its distance from the axis of rotation. It can be calculated with the following formula:

$$xMoment(blob) = \frac{\sum_{p \in blob}(p.x - blob.minx)^2}{|blob|}$$
$$yMoment(blob) = \frac{\sum_{p \in blob}(p.y - blob.miny)^2}{|blob|}$$

where this equation is summed up for the pixels in the array (i.e. pixelArray[i] which contains the x,y position of the ith pixel in the blob) and then normalized by dividing by the number of pixels in the blob in the depth image (i.e. |blob|). This process is repeated for each axis, which results in unique moment of inertia values for each object, which can then be used to distinguish between different objects as they are placed on the earthquake table (Figure 2).
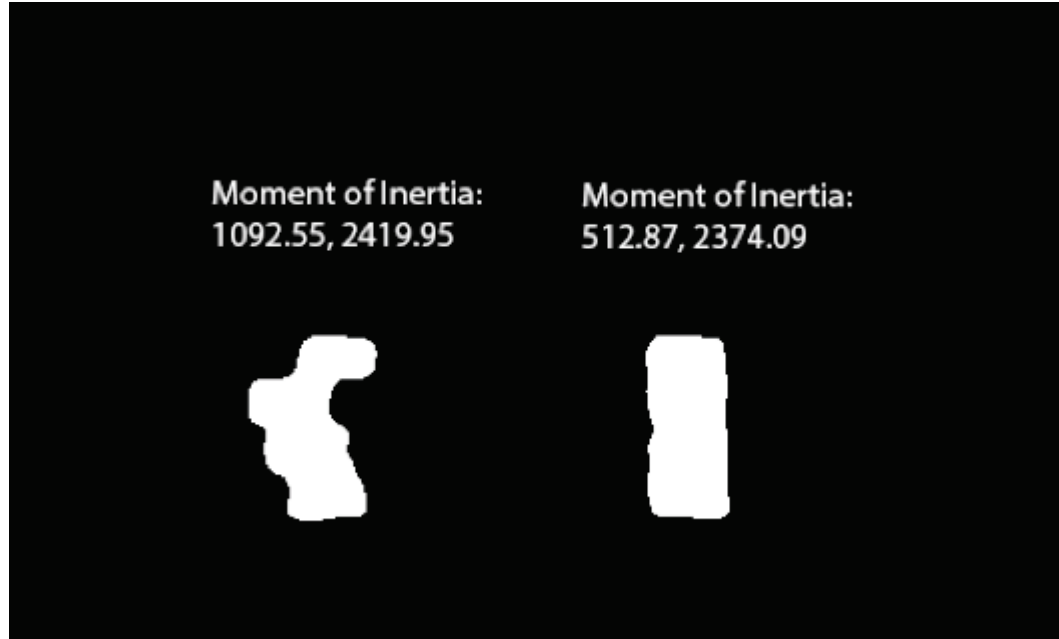
Figure 2. Physics based Moment of Inertia is calculated for each tower, which is a unique value that can help differentiate between different towers.

From a technical perspective, the challenge is in creating tangible interfaces that are sophisticated enough to not only provide children with room for exploration, but also to provide them with interactive feedback that adapts to changes in the physical environment. Such feedback is critical for effective learning (Corbett & Anderson 2001). Without technological support, it is often difficult in real-world tangible interaction to impose pedagogical structure and especially track students' actions. Such structure and logging is comparatively easy in purely virtual settings, but not intuitive in real-world physical setting. We use the Kinect camera and a specialized AI vision algorithm to overcome this challenge.

Using Kinect to blend the physical and virtual environments also expands the paradigm of tangibility beyond specially instrumented objects. Many tangible systems require computation within the physical objects and are not affordable enough for widespread use. Systems such as

11

MirageTable (Benko et al. 2012) and DuploTrack (Gupta et al. 2012) have demonstrated the potential of merging real and virtual worlds into a single spatial experience. With the introduction of inexpensive depth cameras such as the Microsoft Kinect, there is an opportunity for new, scalable paradigms for interaction with everyday physical objects.

The base platform we have created can be extended to many content areas (e.g. our second Intelligent Science Station is about Cars and Ramps to teach Forces and Motion concepts). We use the same base platform including the depth camera, display screen, tablet input screen etc. and the physical objects/apparatus can be placed on top of the modular base platform. The AI vision algorithm can be adapted to work with different content areas, tracking different types of experiments, actions and physical objects on the base platform.

Active Learning via Deliberate Practice: The Guided-Discovery Condition

We used the Intelligent Science Station technology to implement variations of active learning consistent with the alternative theories reviewed above. We start with a description of the version that implements deliberate practice -- see Figure 3. We refer to this version as "guided discovery". Given that it involves multiple forms of implicit and explicit guidance (see Table 1), it has similarity with recommendations of direct or explicit instruction (Klahr & Li 2005). It is distinct from such recommendations in not providing any up-front telling (i.e., no up-front verbal descriptions of the scientific principles) nor any worked examples of how to build a stable tower.

Table 1. Different types of guidance and scaffolding used in each condition.

Forms of scaffolding	*Guided-Discovery Condition*	*Explore-Construct Condition*	*Combined Condition*
Challenge tasks	Yes (predict & explain)	Yes (construction)	Yes (varies)
Feedback on task success	Yes (step-level for predict & explain)	Yes (task-level for tower stability)	Yes (varies)
Provided contrasting cases	Yes	No	Most*
Elicited prediction	Yes	No	Most
Elicited open-ended explanation (no feedback)	Yes	No	Most
Elicited menu-based explanation	Yes	No	Most
Elicited construction	No	Yes	Some*
Up-front telling of the scientific principles	No	No	No

* "Most" indicates that about 2/3 of the experiences have this feature where as "Some" means about 1/3 do.

The pedagogy implemented in the inquiry cycle in the Guided Discovery condition implements a version of deliberate practice by drawing on specific evidence-based techniques in the learning science literature and intelligent tutoring systems, including predict-explain-observe-explain (White & Gunstone 1992), contrasting cases (Chase et al. 2010), self-explanation (Chi et al. 1989; Aleven & Koedinger 2000), and real-time interactive feedback (Corbett & Anderson, 2001). It is aligned with research that recommends students make decisions about comparisons and get feedback on those decisions to best develop critical thinking skills (Holmes et al. 2015). We utilize a character to guide the users through the scientific process, based on recommendations for using pedagogical agents (Moreno et al. 2001; Moreno 2005; Lester et al. 1997).
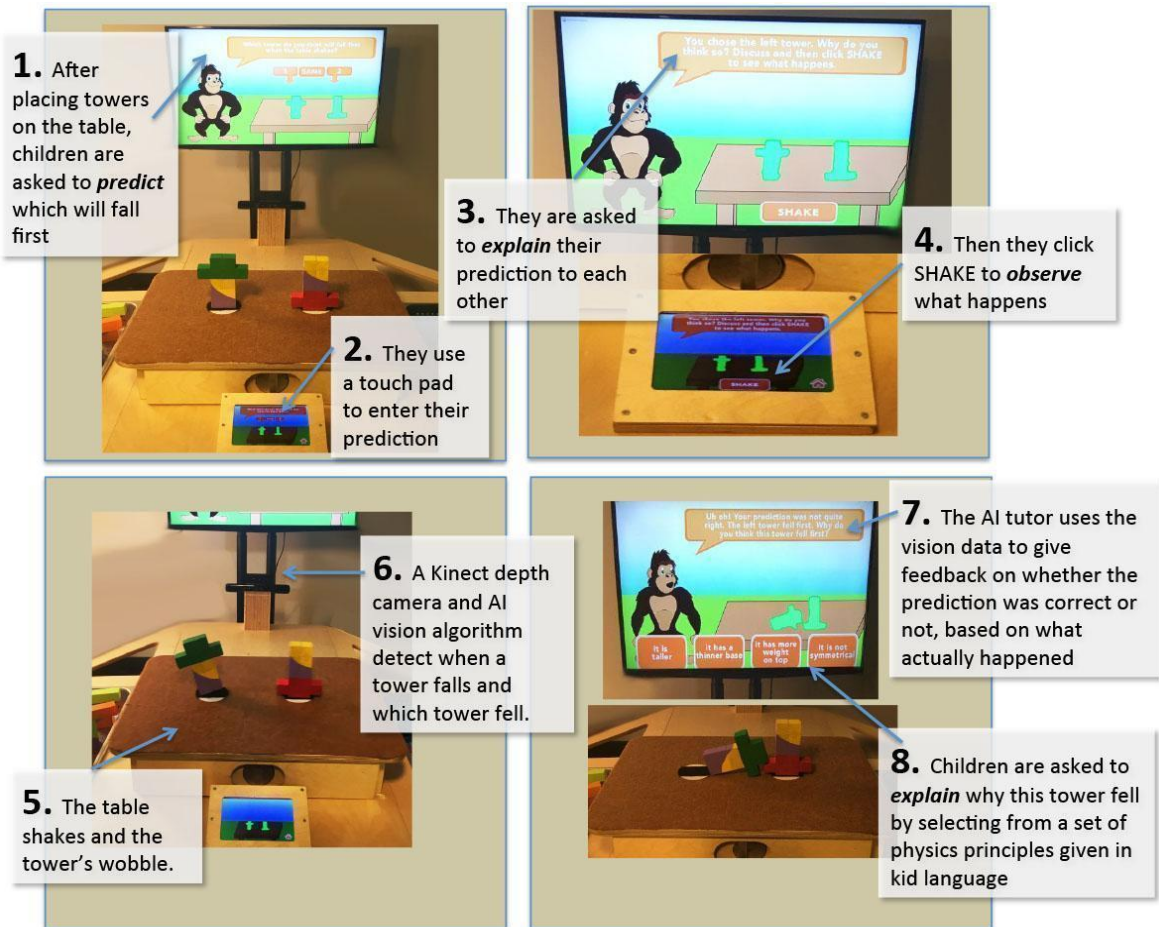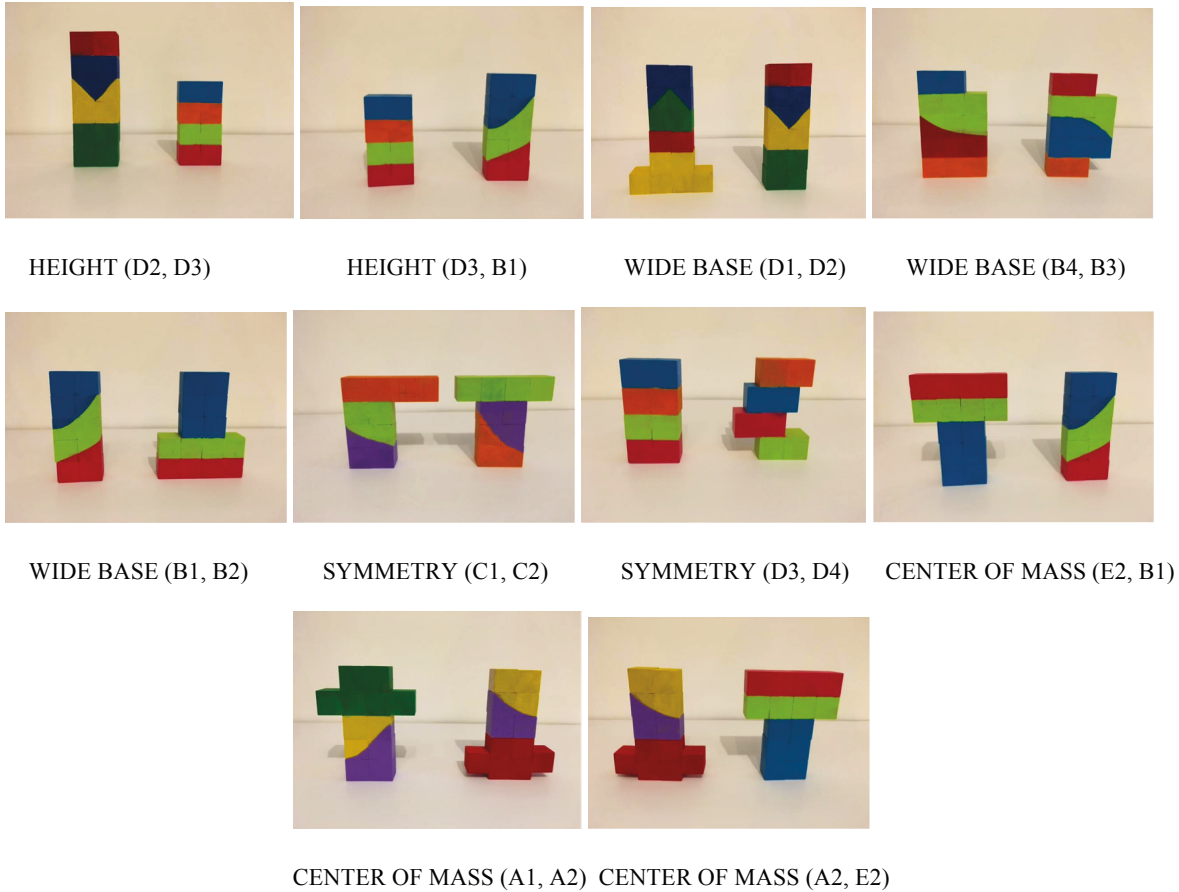
Figure 3. The mixed-reality system and Intelligent Science Station uses a specialized AI computer vision technology to track physical objects and children's actions as they experiment and make discoveries in the real world, facilitating guided inquiry process through a predict (step 1 & 2), explain (step 3), observe (step 4), explain (step 8) cycle *(22)*. In this Guided-Discovery condition, children learn physics principles by discovering them as the gorilla character provides feedback on whether the right towers are placed (step 1), whether the prediction and observation match (step 7), and whether the selected explanation (step 8) is correct (feedback not shown).

Children, usually working in pairs, are asked to place two particular prebuilt towers ("contrasting cases" (Chase et al. 2010)) on an earthquake table and then predict which tower will fall first when the table shakes (steps 1 & 2 in Figure 3) (White & Gunstone 1992). The contrasting cases (Figure 4) are built to help them isolate different target variables to teach different scientific principles (explained below). The computer vision algorithm tracks if they placed the correct tower or not (using physics-based object detection) to ensure that the correct towers are placed that will help the students isolate the targeted variables. After they make a prediction, a gorilla character asks them to explain to each other why they think the tower they chose will fall first (step 3) (Chi et al. 1989; Aleven & Koedinger 2000). When they

14

click the "Shake" button (step 4), the table moves side to side and the towers begin to wobble (step 5). All throughout, the mixed-reality technology is using a depth sensing camera and specialized AI computer vision algorithms to track children's actions (e.g. what type of object they have placed on the table, if it matches the given contrasting cases or not, what position and state the object is in, if the outcome of the experiment matches their prediction or not). At this point, the camera detects when one of the towers has fallen and stops moving the earthquake table (step 6). The information about which tower fell is passed to the system's pedagogical agent and the on-screen gorilla reports the result, for example: "Oh oh, your prediction was not quite right. The left tower fell first." (step 7) (Corbett & Anderson 2001). The gorilla once again prompts children to engage in scientific explanation: "Why do you think this tower fell first?" (Chi et al. 1989). This prompt to explain differs from the previous one in two important ways: First, the outcome is now known ("The left tower fell first") and visible (fallen tower lays on the table). Second, a menu of possible explanations is provided (step 8), in child-friendly language, of relevant physical principles that determine stability. They are spoken by the gorilla and appear on the screen for children to choose from: "It is taller", "It has a thinner base", "It has more weight on the top", "It is not symmetrical or same on both sides". After making a choice from the explanation menu, the gorilla character confirms whether the explanation was correct or wrong with a visualization that illustrates the corresponding physics principle.

This scenario is repeated for different contrasting cases. Below (Figure 4) are the contrasting case (prebuilt) towers that have been designed to teach the physics principles of height (taller tower falls first), wide base (a tower with a wider base stays up longer), symmetry (a symmetrical tower stays up longer) and center of mass concepts (a tower with more weight on top falls first).

HEIGHT (D2, D3)	HEIGHT (D3, B1)	WIDE BASE (D1, D2)	WIDE BASE (B4, B3)
WIDE BASE (B1, B2)	SYMMETRY (C1, C2)	SYMMETRY (D3, D4)	CENTER OF MASS (E2, B1)

CENTER OF MASS (A1, A2) CENTER OF MASS (A2, E2)

Figure 4. Contrasting case towers given in the game to teach different principles (height, wide base, symmetry, center of mass).

Active Learning via Hands-on Constructionism: The Explore-Construct Condition

While the Guided-Discovery condition implements deliberate practice recommendations to guide children in scientific inquiry toward learning science content, the Explore-Construction condition implements constructivist recommendations (Chi & Wylie 2014; Papert & Harel 1991; Kafai & Resnick 1996; Resnick 2014; Jeffrey-Clay 1998; Applefield et al. 2000; Cakir 2008; Van Joolingen 2005; Jonassen 1991; Steffe & Gale 1995) through a more authentic construction/building activity. In this activity children *construct* "externalized outputs" *(*Chi & Wylie 2014*)* in the form of towers, and *explore* their stability in challenge tasks that increase in challenge as children succeed so as to implicitly guide toward increasing use of the principles. This Explore-Construct mode of instruction asks children to build towers using wooden, Lego or magnetic blocks. 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." When they have built their tower and click the SHAKE button on the tablet, it triggers the motor in the physical earthquake table, and the table starts shaking. If their tower falls down, the depth camera and 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 5) and children try again. 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 dancing and says: "Good job! Your tower stayed up! Press CONTINUE to make another tower". Once their tower stays up, a new challenge is added by asking them to build a tower that is taller than their previous tower. The Explore-Construct condition has multiple such challenges intended to provide increasing motivation to employ more of the principles of stability. As implemented in our experiment, students in this condition spend the same amount of time as those in the Guided-Discovery condition.

Figure 5. In the Explore-Construct condition, the gorilla character asks the users to make a tower that will stay up when the table shakes. The AI computer vision algorithm tracks any tower they build, and gives feedback if the tower stayed up or not and for how long. If the tower stays up for 5 seconds, children were given a challenge to try to make a taller tower than the one they did before.

Active Learning via Combined Guided-Discovery and Hands-on Constructionism: The Combined Condition

To test the hypothesis that there are complementary benefits of constructivism and guided deliberate practice, we developed a third condition in which children alternate between Guided-Discovery and Explore-Construct game modes described above, matched to the overall timing of the other conditions.

This Combined condition was designed to take the same total time as the other two conditions, by incorporating about ⅔ of the Guided-Discovery activities and ⅓ of the Explore-Construct activities. We decided on this ratio because we thought that students would need at least two Guided-Discovery experiences for each of the three harder principles and one of the easiest principles (height). Based on our pilot studies, we estimated that these seven experiences amounted to about ⅔ of the total time and left enough time for about two Explore-Construct building experiences.

Methods

We conducted the study within a controlled setting with 75 children in first and second grade from a local school district that is categorized as Title 1 and has 34% economically disadvantaged students. Children interacted with the system in pairs in all conditions. We developed assessments of both scientific and engineering outcomes, used both before and after learning interactions (Yannier et al. 2015; Yannier et al. 2016). For scientific outcomes, we evaluated whether children correctly used the four principles of stability and balance (height, base-width, symmetry and more weight on top versus the bottom) when explaining their predictions of given tower contrasts and when explaining the towers they built. In order to accomplish this, we used paper pre and post-tests as prepared based on the NRC Framework & Asset Science Curriculum (National Research Council 2012). For engineering outcomes, we evaluated the quality of the towers children built before and after learning interactions based on which principles were exhibited (or violated) irrespective of whether children expressed those principles. We also assessed students' ability to predict which tower would fall first in given pairs of towers. Both scientific and engineering outcomes are important because we will not differentiate constructivism from deliberate practice if we merely find that the instructional experiences produce analogous learning outcomes - Explore-Construct is more directly relevant to the engineering outcomes whereas Guided-Discovery is more directly relevant to the scientific outcomes. To differentiate, we need to see transfer to dissimilar

outcomes -- either Explore-Construct to scientific outcomes, supporting constructivism, or Guided-Discovery to engineering outcomes, supporting deliberate practice.

Children in the Explore-Construct condition practiced tower building for the whole time of the experiment whereas Combined condition did two tower building tasks. The timing between different conditions was matched, so that there was little difference between the time on task for children in different conditions. Based on the video and log data, average time during learning (excluding pre and post assessment) for Combined Condition is approximately 15 minutes, for Guided-Discovery Condition is 15 minutes and for Explore-Construct Condition is 16 minutes.

Experimental Procedure

The students were pulled out from their classroom in pairs (except for one group of three). They were randomly assigned to one of the three conditions so as to distribute assignments across the three conditions evenly as the pairs arrived. Before interacting with the game, students were first given a tower pre-test, consisting of two tower tasks. First, the experimenter showed them a tower that was prebuilt (Figure 6) and told them: "I built this tower, but it is not very stable and it would fall down if I shake the table. Can you make a tower that is more stable using the same blocks" and handed them a bag of blocks that they can use to build a tower together with their partner. After they were done building their tower, the experimenter asked them to explain how they built it and if they had any strategies in mind. After they explained their towers, they were then given another bag of blocks. This time they were asked to build a tower using all the blocks in the bag, using a specific red block on the bottom. The experimenter told them that there were two rules this time: first, they had to use all the blocks in the bag and they had to use the red block on the bottom and nothing else can touch the table. Again, after they built the tower in collaboration with their partner, they were asked to explain how they built their tower.
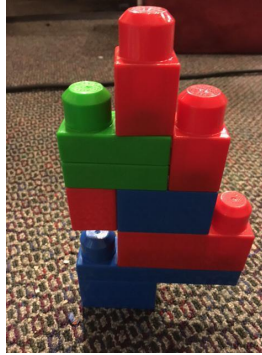
Figure 6. Tower building assessment example: Participants were given a prebuilt tower and asked to make a tower that is more stable than this tower using the given blocks.

Then they were asked to complete a paper pretest to measure what they already knew about the stability. and balance principles in the game. Students then interacted with their randomly-assigned game, either Guided-Discovery, Explore-Constuct, or the Combined condition. The condition content was designed through piloting to involve the same time and some maximum time count-off was employed in the case some children take much longer than expected.

For the Guided-Discovery condition, participants interacted with 10 contrasting cases (in the sequence given in Figure 23). For the Combined condition, participants interacted with 5 contrasting cases (in the following sequence: D2&D3, D1&D2, B4&B3, D3&D4, A2&E2), then they did one exploration activity in the game and then they interacted with 2 more contrasting cases and 2 more exploration activities (C1&C2 and A1&A2). This sequence was determined based on piloting before the experiment to match the timing of the Guided-Discovery condition and to make sure children get exposed to enough contrasting cases before the first exploration. Thus the overall time on task was the same as Guided-Discovery condition in the Combined condition, however the time spend on guided-discovery activities was less than the Guided-Discovery condition (7 guided-discovery activities in Combined versus 10 guided-discovery activities in Guided-Discovery). On the other hand for the Explore-Construct condition, they were only given exploration tasks where they were asked to build a tower that would stay up when the table shakes. If they were able to build a tower that stayed up, then they were asked to build a tower that is taller than
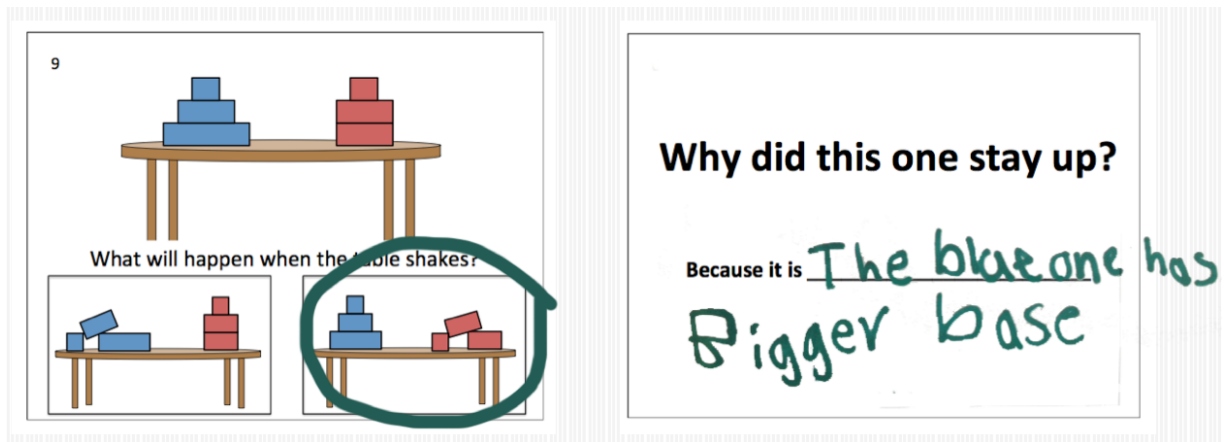
their previous tower that would still stay up. Again, the time that children interacted with the game was calculated to match the other conditions.

After interacting with their game, students were given a matched paper post-test. After the paper post-test, the students were given the same tower building task as before game play. As in prior studies the contrast between the pre- and post- tower building assessment was used to measure student improvement in tower building, and in particular, how well they incorporate the principles of balance. Finally, the students were asked to fill out a survey to see how much they enjoyed the game.

Measures

The paper pre and post-tests are prepared based on the NRC Framework & Asset Science Curriculum *(*National Research Council 2012*)*. We used the same tests as we used in the previous experiments (Yannier et al. 2015; Yannier et al. 2016). The tests consist of two types of items: prediction items and explanation items. For the prediction items, the students were given a picture of a table with two towers and were asked to predict which will fall when the table shakes. In the explanation items, they were asked to explain why they chose their answer (Figure 7). Also, children were given a survey at the end of the game to measure their enjoyment.

The survey consisted of three questions (See Figure 8). The first question was: "How much did you like the game?" They could choose one of: "I didn't like it at all", "I didn't like it", "It was OK", "I liked it", "I liked it very much". The second question was: "Would you like to play it again?". They could choose "Yes", "No" or "Maybe" by choosing one of the smiley faces from a scale of 1-5. Finally, the third question was: "Would you recommend it to a friend?". Again, they could choose "Yes", "No" or "Maybe" by choosing one of the smiley faces from a scale of 1-5.

Figure 7. Prediction (left) and explanation (right) items used in the paper pre/post-tests.

Figure 8. Survey questions.

We also gave the participants some hands-on building tasks as pre and post tower tests, to see how much they could transfer the knowledge they gained to practical hands-on activities in the real world. Different than the previous experiments (Yannier et al. 2015; Yannier et al. 2016), this time we gave children two

different tower tests, to have more hands-on activities as pre and post tests and have a better understanding of how their learning from the different modes of the game translate to real-world hands-on activities.

Another measure of scientific outcomes included an analysis of student explanations to the experimenter's questions, "Can you explain how you built the tower? Did you have any strategies in mind?" after they were asked to build a tower that would stay up. Videos of the pre and post tower tests were coded and analyzed.

To measure pre- to post-test 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 post-test, -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. 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 9).

	S	CoM	H	Σ		S	CoM	H	Σ
	0	-1	-1	**-2**		1	1	1	**3**
	0	1	1	**2**		1	0	0	**1**
	-1	-1	-1	**-3**		0	1	1	**2**

Figure 9. Coding scheme for Tower pre/post tests change.

Results

Transfer Benefits of Deliberate Practice Through Guided Discovery

Our findings are summarized in Figure 10. An ANOVA testing the effect of the three conditions on the engineering outcome (improvement in tower building; see panel (a) in Figure 10) revealed a statistically significant effect of condition ($F_{(2, 72)}=4.24$, $p<0.02$). Most surprisingly, despite the exclusive focus on tower building by the Explore-Construct condition, children in this condition showed the least improvement on tower building ($M = 0.17$ mean improvement). The Combined condition produced nearly a standard deviation effect ($d=.92$; $M = 2.31$) over the Explore-Construct condition ($F_{(1,48)}= 9.38$, $p<0.01$) on tower building, even though children in this condition were practicing less tower building. Thus, some scientific inquiry guidance enhances engineering outcomes better than less-guided constructive exploration, especially when some construction is intermixed with the inquiry guidance. Learning of

construction/engineering skills, as measured by tower building improvement, was more than 10 times larger for the Combined condition (2.31) than the Explore-Construct condition (0.17). In addition, an ANOVA test for the Combined and Guided-Discovery conditions, showed that there was a trend in favor of the Combined condition – Combined condition transferred marginally better to hands-on experimentation compared to the Guided-discovery condition ($F(1,50)= 2.01$, $p=0.1$, $d=0.43$).

The prediction measure (panel (b) in Figure 10) shows a similar, though less differentiated pattern. An ANCOVA with post-test prediction score as the outcome variable and pre-test prediction score as covariate revealed a marginal main effect of the Combined condition over the Explore-Construct condition ($F(1,48)=3.6$, $p=0.06$) with a moderate ($d=0.38$) effect size.

While better tower building and prediction can be achieved through implicit knowledge of principles of stability revealed through behavior, the explanation assessments evaluate children's explicit scientific knowledge of these principles. Here we find clear and consistent evidence of the benefit of guided inquiry over constructive exploration (see panel (c) in Figure 10). An ANCOVA with post-test prediction explanation score as the outcome variable (left bars 1c) and pre-test prediction explanation score as a covariate revealed that the Combined ($F(1,48)= 7.72$, $p<0.01$, $d=0.66$) and Guided-Discovery ($F(1,47)= 9.74$, $p<0.01$, $d=0.75$) conditions learned to provide better explanations of predictions than the Explore-Construct condition. Learning and understanding of scientific principles, as measured by explanation tests, was more than 4 times greater for the Guided-Discovery condition ($M=0.29$ mean improvement) and the Combined condition ($M=0.22$ mean improvement) than the Explore-Construct condition ($M=0.06$).

Explanations of tower building (right bars in 1c) revealed the same pattern. Children in the Combined ($F(1,48)= 4.04$, $p=0.05$, $d=0.45$) and Guided-Discovery ($F(1,47)= 3.93$, $p=0.05$, $d=0.56$) conditions learned to provide better explanations of towers they build than those in the Explore-Construct condition.
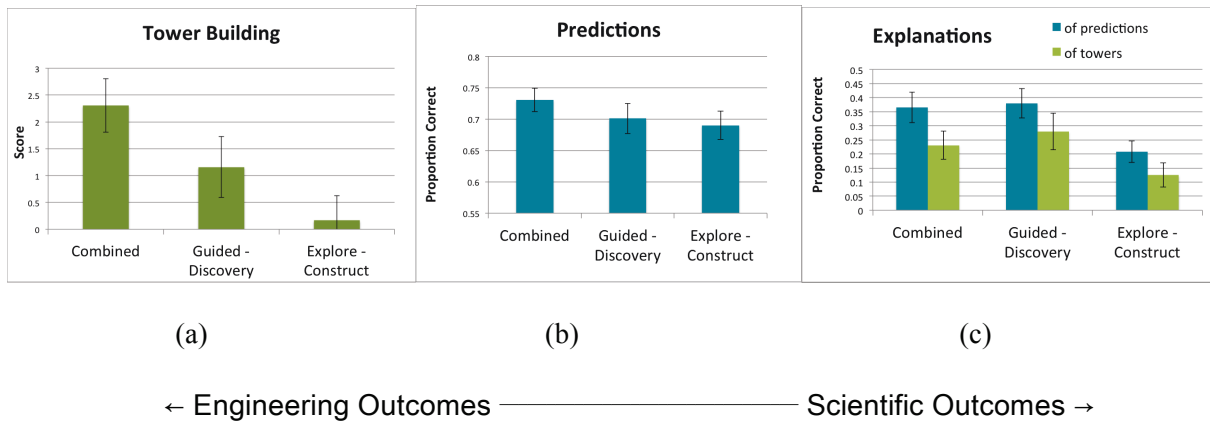
← Engineering Outcomes ——————— Scientific Outcomes →

Figure 10. Students benefited from scientific inquiry guidance. They learned less from exploration and construction across all outcome measures, even a measure of tower building that mirrored the experience in the exploration condition. a) Combined is significantly better than Explore-Construct on a Tower Building assessment of engineering outcomes. b) Combined is marginally better than Explore-Construct on predicting experimental contrasts. c) Combined and Guided are significantly better than Explore-Construct on using scientific principles to explain both predictions (left bars) and towers (right bars).

In general, we found that guided inquiry enhances both scientific and engineering outcomes. The engineering outcomes are especially enhanced when some constructive building experience is combined with scientific inquiry guidance. Exploration through a minimally-guided constructive experience yielded relatively poor learning in general. This finding is particularly striking when one considers that children's activity in the Explore-Construct condition mirrors the tower building assessment. Even though children do less building in the Combined condition, they learn to be better builders as a consequence of the guided inquiry experience integrated with construction. The Guided-Discovery condition, in which there is no building, appears to produce as much or more transfer to building as the less-guided Explore-Construct condition.

Deliberate Practice through Guided Discovery is no Less Fun than Construction

A purported benefit of constructivism is that it better engages students, for example, through "playful experimentation" or "tinkering with materials" (Resnick 2014*).* Our evidence suggests that children enjoyed the conditions involving (this particular form of) deliberate practice as much as the Explore-Construct condition. Based on a three-question survey, we found no statistically significant difference

26

between average enjoyment scores: 94% for the Combined condition, 90% for the Guided-Discovery condition and 91% for the Explore-Construct condition. In addition to the high ratings, many children said things we rarely hear in school, including "This is the best day of my life!", "I'm gonna be a builder when I grow up, because a lot of these didn't fall!", and "Can you make one for my birthday? I can trade my toys."

Discussion

We found that the guided discovery facilitated by the AI agent in our Intelligent Science Station helped children formulate better, more scientific theories of the physical phenomena they experience. Our results further indicate children receiving guidance during inquiry are better able to learn to *apply* science in engineering tasks, particularly when guided discovery is interleaved with construction.

Integrating across the results, an interesting pattern emerges. The Guided-Discovery and Combined conditions yield better explanation of predictions and constructed towers. Such better explanation indicates better scientific understanding of the underlying physics principles. Our results indicate the power of guided discovery and potential weakness of a sole focus on less-guided exploration and construction. We also find evidence of benefits of combining some exploration and construction along with guided discovery. In particular, the results of the tower building and prediction tasks suggest that children may better learn to use scientific explanations (e.g., "more weight on the bottom") to facilitate engineering (i.e., building towers) when some exploratory construction is added to guided discovery.

Importantly, exploration and construction did not lead to better tower construction, even though children were practicing more construction/building in this condition. This result is *practically important* because the Explore-Construct condition is similar to how many museum exhibits, tangible interfaces and Maker Spaces in formal and informal settings are designed to support tinkering/exploration and to what some proponents of constructivism advocate (Chi & Wylie 2014; Jeffrey-Clay 1998): "Museums have

requisites for constructivist learning. They offer free choice environments ...Visitors … need to find their own meaning from the exhibit ... [and] guide their own learning".

This result is also *scientifically important* as it adds important caveats or boundary conditions on existing theories of learning and instruction (Chi &Wylie 2014; Bransford et al. 2000). According to the ICAP Framework (Chi & Wylie 2014), the Explore-Construct condition should work better than the Guided-Discovery because it is Constructive whereas Guided-Discovery is Active. According to ICAP, Constructive is better than Active, but we found just the opposite: The Active condition (Guided-Discovery) produces better outcomes than the Constructive condition (Explore-Construct).

This result also helps refine learning theories, like transfer appropriate processing (Bransford et al. 2000), which suggest, in a nutshell, you learn what you practice. Recommendations deriving from such theories suggest to "match the job task" (Clark et al. 2011). However, these theory-derived recommendations incorrectly predict that better outcomes on tower building should result from the Explore-Construct condition, in which students practice tower building. Task similarity does not ensure transfer. A more refined interpretation of transfer appropriate *processing* puts an emphasis on the hidden processing that tasks evoke. As we elaborate below, the guided-discovery task evokes processing (i.e., principle application) that is more transfer appropriate to tower building than is the hands-on tower building task.

Adding the guided-discovery support and intelligent layer that the mixed-reality AI system provides, not only fosters better learning of the balance/stability physics principles, but also improves the application of those principles in a hands-on, constructive problem-solving task. This result can be understood by considering differences in the way children are thinking while building. In the Explore-Construct condition, given limited relevant prior knowledge, children mostly engage in unsystematic tweaking, responding to a fallen tower by trying another configuration of blocks without a plan (e.g., "maybe I can go like this" or "you just stack them").

In contrast in the Combined condition, children bring to bear the scientific principles they learned during guided discovery and form plans to build or rebuild towers based on the principles. This condition

facilitates the interleaving of focused practice between "learn the theory" and "apply the theory". Children learn to formulate accurate scientific explanations and then, in turn, this emerging explanatory theory is activated, tested, and strengthened when put to use in exploratory construction. One child expressed well how she used the theory she learned in guided discovery (i.e., that weight on the bottom adds stability) in tower construction: "That's why I put that at the bottom… cos I learned today that if it's like that at the bottom it won't fall!"

What about the notion of active learning more generally? The need for more scientific precision with respect to the notion of active learning is indicated by the wide variation in ways it is described in different publications. One report emphasizes that students "actively grapple with questions" and that students "are framing the questions themselves" (Waldrop 2015), another emphasizes "deliberate practice", "constructivism", "formative assessment", and that students "practice physicist like reasoning" (Deslauiers et al. 2011), and a third focuses on "problem solving" (Freeman et al. 2014). Which of these are the active ingredients of active learning?

Consider how these different attempts at defining and operationalizing active learning map to our conditions. The notion that "students gain a much deeper understanding of science when they actively grapple with questions than when they passively listen to answers" (Waldrop 2015) applies to the Explore-Construct condition, where children actively grapple with the question: Can you make a tower that stays up for 5 seconds when the table shakes? However, it also applies to Guided-Discovery where children actively grapple with questions, such as "Why do you think this tower fell first?" So, while critical, this feature of active learning does not predict the observed differences. The notion that students "take charge of their own education" in an active-learning class (Waldrop 2015) and "... learners generate or produce additional externalized outputs" (Chi & Wylie 2014) seem best enacted in the Explore-Construct condition. Similarly, the emphasis on "problem solving" (Freeman et al. 2014) seems to apply at least as well to the Explore-Construct condition. These elements of active learning do not capture well what is distinctly advantageous about Guided-Discovery.

Other reported elements of active learning are more differentiating, in particular, prior reports (Deslauiers et al. 2011; Ericsson et al. 1993) emphasize the notion of deliberate practice, which involves immediate feedback, knowledge of results, and repetition (as described above). These features of active learning match the Guided-Discovery condition where students are repeatedly performing similar predict-explain-observe-explain tasks and get immediate informative feedback on their predictions and explanations.

Another feature of deliberate practice is "a steady accumulation of knowledge about the best methods to attain a high level of performance" and "associated practice activities" (Ericsson et al. 1993). Our prior cognitive task analysis of this early physics domain identified four principles of balance (height, base-width, symmetry and more weight on top versus the bottom) that are critical to building stable structures but are not easy for children to learn (Christel et al. 2012). Much like tennis players benefit from practicing their overhand serve in isolation from game playing (or even tossing the ball to the right height in isolation from serving), our guided-inquiry practice is isolated from the full game of building towers. The key point regarding the "accumulation of knowledge" is to distinguish which part tasks are worth isolated practice and which are not, at different points in the learning process. Isolated practice of one's forehand in tennis or placing one block on top of another are also part tasks of the whole game, but are not productive targets for isolated practice if they are relatively easy to learn.

During focused practice, informative feedback is also critical. In the case of Intelligent Science Stations, we implemented AI computer vision algorithms to provide such immediate feedback on children's real world experimentation efforts. Going beyond the direct recommendations of deliberate practice, the Guided-Discovery condition employs contrasting cases, predict-explain-observe-explain, and self-explanation and the Combined condition employs an alternation between isolated guided practice and constructive exploration.

Our results are aligned with Hattie and Donoghue's learning strategies model that suggests that higher level thinking requires a sufficient corpus of lower level knowledge to be effective (Hattie & Donoghue 2016). Similarly, our results show that learning lower-level scientific principles through guided

inquiry facilitates student learning of higher-level engineering applications of those principles (as measured by the engineering outcomes). In other words, students develop deeper understanding and can transfer better to real-world building if they learn some fundamentals (via Guided-Discovery) in service of learning higher level thinking (e.g. problem solving via Explore-Construct). To be sure, it is possible to engage in engineering applications without the lower-level scientific principles as children do in the Explore-Construct condition. However, engineering is more effective when using these principles as evidenced by better performance on building in the Combined condition. We suspect the sequencing in the Combined condition, with Guided-Discovery first, is important. Consistent with Hattie and Donoghue, initial practice on the fundamental principles, which the guided discovery provides, may aid thinking and learning during the building challenges. When children fail during the building challenges in the Combined condition, they bring to bear the scientific principles they have learned to iterate and build better structures -- a kind of productive failure (cf., Kapur 2010). The Explore-Construct condition, in contrast, produces more "unproductive failure" because children tend to respond to failure with random tweaking when they do not have more principle knowledge to employ. Without the prior preparation through guided discovery of the scientific principles, it is too complex (too much cognitive load) and/or time consuming (too big a discovery search space) for children to discover these principles through building.

As a result, we find that less guided exploration and constructivism without inquiry guidance and focused practice may not be a productive form of active learning. An integration of key features of deliberate practice are productive: They include a) immediate, interactive feedback during b) isolated practice on c) part tasks known to be relevant and difficult. As implemented in the Combined condition, alternating between part-task practice and "whole task" constructive exploration is generally better than constructive exploration on its own, even when the target outcome is whole task construction.

The Combined condition, which emphasizes guided-discovery but interleaves some exploratory hands-on exploration, also appears to produce better results on tower building than the Guided Discovery condition. This difference may be explained by the fact that the Combined condition provides practice on

the use of principles in the context of building. In other words, the Combined condition provides an opportunity for students to activate explanatory theory in action, whereas the Guided Discovery condition does not. Children in the Guided Discovery condition clearly learn the principles, as evidenced especially by the explanation assessments, but they do not get practice in applying them to building. Despite children in the Combined condition getting less experience in guided discovery tasks than those in the Guided Discovery condition (78% vs. 100% of the time), they appear to learn the principles just as well, as evidenced by similar explanation and prediction scores. To be clear, this study was conducted with children ages 5-9 (K-3rd grade), so the results of the study apply to this age group. Further research may be needed to investigate if these results can be applied to other age groups as well."

In summary, the Combined and Guided-Discovery conditions build on multiple elements of prior research of effective strategies for integrating discovery and guidance, including contrasting cases (Chase et al. 2010), predict-observe-explain (White & Gunstone 1992), and menu-based self-explanation (Chi et al. 1989; Aleven & Koedinger 2000) and for practice on complex tasks informed by prior instruction on fundamentals (Hattie & Dononghue 2016). Intelligent Science Stations and the AI agent we introduce, automate these effective ingredients by adapting to children's scientific and engineering activities in the physical world, 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. Our random assignment experiment rigorously demonstrates that children learning with AI adaptive inquiry guidance in a mixed-reality setting achieve greater understanding of the scientific principles and over ten times more transfer to real world construction/engineering than children learning through hands-on construction alone. These results suggest that mixed-reality/augmented-reality systems may be more effective when forms of interactive guidance are added to facilitate active learning and understanding of fundamental scientific principles, rather than having a sole focus on hands-on exploration. To make such interactive guidance possible, may require innovation in AI tutoring support, like the vision algorithms used here, but the efforts seem to pay off. More generally, these results suggest that hands-on experiences are more effective

learning experiences with more potential to improve STEM education, when integrated with guided inquiry and deliberate practice activities.

References and Notes:

1. Aleven, V., Koedinger, K. R. (2000) The need for tutorial dialog to support self-explanation. Building dialogue systems for tutorial applications, papers of the 2000 AAAI Fall Symposium.

2. 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.

3. Applefield, J. M., Huber, R., & Moallem, M. (2000) Constructivism in theory and practice: Toward a better understanding. The High School Journal, 84(2), 35-53.

4. 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). ACM.

5. Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000) How people learn.

6. Cakir, M. (2008) Constructivist Approaches to Learning in Science and Their Implications for Science Pedagogy: A Literature Review. International Journal of Environmental & Science Education, 3(4), 193-206.

7. 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.

8. 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.

9. Chi, M., & Wylie, R. (2014) The ICAP framework: Linking cognitive engagement to active learning outcomes. Educational Psychologist, 49, 219–243.

10. Christel, M. G., Stevens, S. M., Maher, B. S., Brice, S., Champer, M., Jayapalan, L., ... & Zhang, X. (2012) RumbleBlocks: Teaching science concepts to young children through a Unity game. In Computer Games (CGAMES), 17th International Conference on (pp. 162-166). IEEE.

11. Clark, R. E., & Estes, F. (1996) Cognitive task analysis for training. International Journal of Educational Research, 25(5), 403-417.

12. Clark, Ruth C., and Richard E. Mayer. (2011) E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning. John Wiley & Sons.

13. Corbett, A. T., and Anderson, J. R. (2001) Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. Proceedings of the SIGCHI conference on Human factors in computing systems. ACM.

14. Deslauriers, L., Schelew, E., & Wieman, C. (2011) Improved learning in a large-enrollment physics class. Science 332.603, 862-864.

15. Dougherty, D. (2012) The maker movement. Innovations: Technology, Governance, Globalization, 7(3), 11-14.

16. Edelson, D., Pea, R., & Gomez, L. (1996). Constructivism in the collaboratory. Constructivist learning environments: Case studies in instructional design, 151.

17. Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993) The role of deliberate practice in the acquisition of expert performance. Psychological review, 100(3), 363.

18. Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014) Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences, 111, 8410–8415.

19. 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.

20. Gibson, J. J., & Gibson, E. J. (1955) Perceptual learning: Differentiation or enrichment. Psychological Review, 62, 32–41. doi:10.1037/h0048826.

21. 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). ACM.

22. Halverson, E. R., & Sheridan, K. (2014) The maker movement in education. Harvard Educational Review, 84(4), 495-504.

23. Hattie, J. A. C. (2009) Visible learning: A synthesis of 800+ meta-analyses on achievement. Routledge, Oxford, UK.

24. Hattie, J. A., & Donoghue, G. M. (2016) Learning strategies: a synthesis and conceptual model. npj Science of Learning, 1, 16013.

25. 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.

26. Holmes, N. G., Wieman, C. E., & Bonn, D. A. (2015). Teaching critical thinking. Proceedings of the National Academy of Sciences, 112(36), 11199-11204.

27. Jeffery-Clay, K. R. (1998) Constructivism in museums: How museums create meaningful learning environments. Journal of Museum Education, 23(1), 3-7.

28. Jonassen, D. H. (1991) Objectivism versus constructivism: Do we need a new philosophical paradigm?. Educational technology research and development 39.3, 5-14.

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

30. Kapur, M. (2010) Productive failure in mathematical problem solving. Instructional Science, 38(6), 523-550.

31. Kirschner, P. A., Sweller, J., & Clark, R. E. (2006) 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, 75-86.

32. Klahr, D., & Li, J. (2005) Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. Journal of Science Education and Technology, 14(2), 217-238.

33. Klahr, D. (2009) To every thing there is a season, and a time to every purpose under the heavens: What about direct instruction?.

34. Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012) The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. Cognitive science, 36(5), 757-798.

35. Koedinger, K. R., Booth, J. L., & Klahr, D. (2013) Instructional complexity and the science to constrain it. Science, 342(6161), 935-937.

36. Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., ... & 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.

37. Morris, C. D., Bransford, J. D., & Franks, J. J. (1977) Levels of processing versus transfer appropriate processing. Journal of Memory and Language, 16(5), 519.

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

39. Nickerson, R. S. (1998) Confirmation bias: A ubiquitous phenomenon in many guises. Review of general psychology, 2(2), 175.

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

41. Papert, S., & Harel, I. (1991) Situating constructionism. Constructionism, 36(2), 1-11.

42. Prince, M. Does active learning work? (2004) A review of the research. Journal of engineering education, 93(3), 223-231.

43. 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.

44. Shapiro, L. (2010) Embodied cognition. Routledge.

45. Steffe, L. P., & Gale, J. E. (1995) Constructivism in education. Hillsdale, NJ: Lawrence Erlbaum.

46. Tobias, S. & Duffy, T. M. (2009) Constructivist Instruction: Success Or Failure? New York: Routledge.

47. Van Joolingen, W. R., et al. (2005) Co-Lab: research and development of an online learning environment for collaborative scientific discovery learning. Computers in human behavior 21.4, 671-688.

48. Waldrop, M. M. (2015) The science of teaching science. Nature 523.7560, 272.

49. White, R., & Gunstone, R. (1992) Prediction-observation-explanation. Probing understanding, 4.

50. 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.

51. 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.

52. 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.

53. https://designerlibrarian.wordpress.com/2015/10/19/the-recipe-for-successful-makerspace-design/

Figure Captions

Figure 1. Children interacting with the Intelligent Science Station in the Guided-Discovery condition, where they make predictions, observe results, and provide explanations with interactive feedback from an AI system that can see the results of their experiments. Children's engagement can be seen in a supplementary video.

Figure 2. Physics based Moment of Inertia is calculated for each tower, which is a unique value that can help differentiate between different towers.

Figure 3. The mixed-reality system and Intelligent Science Station uses a specialized AI computer vision technology to track physical objects and children's actions as they experiment and make discoveries in the real world, facilitating guided inquiry process through a predict (step 1 & 2), explain (step 3), observe (step 4), explain (step 8) cycle *(22)*. In this Guided-Discovery condition, children learn physics principles by discovering them as the gorilla character provides feedback on whether the right towers are placed (step 1), whether the prediction and observation match (step 7), and whether the selected explanation (step 8) is correct (feedback not shown).

Figure 4. Contrasting case towers given in the game to teach different principles (height, wide base, symmetry, center of mass).

Figure 5. In the Explore-Construct condition, the gorilla character asks the users to make a tower that will stay up when the table shakes. The AI computer vision algorithm tracks any tower they build, and gives feedback if the tower stayed up or not and for how long. If the tower stays up for 5 seconds, children were given a challenge to try to make a taller tower than the one they did before.

Figure 6. Tower building assessment example: Participants were given a prebuilt tower and asked to make a tower that is more stable than this tower using the given blocks.

Figure 7. Prediction (left) and explanation (right) items used in the paper pre/post-tests.

Figure 8. Survey questions.

Figure 9. Coding scheme for Tower pre/post tests change.

Figure 10. Students benefited from scientific inquiry guidance. They learned less from exploration and construction across all outcome measures, even a measure of tower building that mirrored the experience in the exploration condition. a) Combined is significantly better than Explore-Construct on a Tower Building assessment of engineering outcomes. b) Combined is marginally better than Explore-Construct on predicting experimental

contrasts. c) Combined and Guided are significantly better than Explore-Construct on using scientific principles to explain both predictions (left bars) and towers (right bars).