Under Which Conditions are Physical vs. Virtual Representations Effective? Contrasting Conceptual and Embodied Mechanisms of Learning

Abstract

Abundant prior research has compared effects of physical and virtual manipulatives on students' conceptual learning. However, most prior research has been based on conceptual salience theory; that is, it has explained mode effects by the manipulative's capability to draw students' attention to conceptually relevant (visual or haptic) features. Yet, research based on embodied schema theory suggests that other mechanisms, which do not rely on students' explicit attention to specific features, also affect students' learning from manipulatives. This paper presents a study that contrasts predictions by different theoretical perspectives by comparing multiple versions of physical and virtual manipulatives. Specifically, we conducted a lab experiment with 119 undergraduate students who learned about three concepts related to atomic structure using one of four versions of energy diagram manipulatives. The four versions varied the representation mode (i.e., physical vs. virtual) and the actions students used to manipulate the representation (i.e., via actions that draw attention or activate embodied schemas). We assessed students' learning via reproduction and transfer posttests and interviews that measured the quality of students' explanation and the gestures they used while explaining the concepts. Our results suggest that embodied schema theory accounts for effects on the reproduction posttest, whereas conceptual salience theories account for effects on the transfer posttest. Further, when physical manipulatives offered relevant haptic cues, we found an advantage of physical manipulatives on transfer. We interpret these results based on the complexity of embodied schema and conceptual salience learning mechanisms and the complexity of the assessment tasks.

Educational Impact and Implications Statement

An experiment tested how different designs of manipulatives affect students' learning of chemistry concepts. The manipulatives were either physical or virtual and were designed so that students moved their hands differently to show information about atoms. Pretest and posttests of students' knowledge about the chemistry concepts assessed their learning gains. Results showed that no one design was consistently most effective but that different designs were effective for different learning outcomes. This finding provides guidance for matching manipulative designs with learning goals.

Keywords: Physical and virtual representations; representation modes; haptic cues; conceptual salience; embodied cognition

1. Introduction

Most STEM domains use multiple representations to illustrate key concepts. We define learning with multiple representations as situations where students encounter more than one representation (e.g., two visual representations or text plus visual representation). Compared to a single representation, multiple representations can enhance learning because students can glean information from each as well as deepen their understanding by integrating across the representations (Ainsworth, 2006; Rau, 2017; Schnotz, 2014). This article focuses on the case where one visual representation is combined with text-based representations. While much research has investigated this scenario, this article investigates a less researched question; namely, how to choose a particular set of representational features—physical vs. virtual representation modes—to maximize student learning. In doing so, we draw on additional theories that extend multiple representation theory.

Visual representations are often presented in the context of problem-solving activities that ask students to modify visual representations in ways that mimic disciplinary problem-solving practices. Such modifiable visual representations—also referred to as manipulatives—can be presented in the physical or virtual mode, or in a blended form that combines physical manipulatives with computer-based materials (e.g., Antle, Corness, & Droumeva, 2009; Johnson-Glenberg, Birchfield, Tolentino, & Koziupa, 2014; Olympiou & Zacharia, 2012). However, blended learning environments rarely combine virtual and physical versions of the same manipulative (e.g., a physical as well as a virtual model of an atom) because representations that have high informational overlap are redundant and may reduce student learning (Rau & Matthews, 2017). Rather, the goal is to choose those physical *or* virtual manipulatives that maximize student learning. To inform this choice, it is therefore critical to understand under what conditions physical or virtual representation modes are more effective (e.g., is a physical model preferable to a virtual one?).

Whereas multiple representation theory does not address how the representation mode might affect learning, a related body of prior research has compared the effects of physical versus virtual manipulatives (for reviews and meta-analyses, see Carbonneau, Marley, & Selig, 2013; de Jong, Linn, & Zacharia, 2013; Moyer-Packenham & Westenskow, 2013). However, most of this prior research has focused on conceptual learning mechanisms; for example, whether a physical or a virtual manipulative draws students' attention to a target concept (de Jong et al., 2013; Olympiou & Zacharia, 2012; Zacharia & Olympiou, 2011). Yet, embodied learning mechanisms have also been shown to have a strong impact on students' learning with manipulatives (Abrahamson & Sánchez-García, 2016; Black, Segal, Vitale, & Fadjo, 2012; Nathan & Walkington, 2017). For example, moving one's hands upwards can enhance students' learning of concepts related to growth, even if students are unaware of the connection between their movements and the concept.

While many interventions seek to integrate conceptual and embodied mechanisms (e.g., Abrahamson & Lindgren, 2014), a recent review (Rau, 2020) showed that theories of conceptual learning and theories of embodied learning often make conflicting predictions about the effectiveness of physical and virtual manipulatives. Hence, designing effective interventions that combine these mechanisms necessitates an understanding of how these mechanisms affect student learning. The goal of this article is therefore to contrast these predictions by comparing the effects of physical and virtual manipulatives on students' learning of various concepts. We conducted this research in the context of a commonly used chemistry representation. In doing so, we investigated whether conceptual or embodied mechanisms have a stronger impact on students' learning of these concepts.

By comparing predictions made by these different theoretical perspectives, our research extends prior work that has compared representation modes using only one of these theoretical perspectives. Further, our findings extend multiple representation theory by providing insights into the mechanisms through which representation modes affect students' learning. Finally, our research moves beyond prior work that has investigated *whether* combinations of multiple representations enhances learning to provide guidance as to *which* combinations of multiple representations are most productive.

2. Theoretical Background

In the following, we briefly review research on learning with multiple representations, followed by research that has compared physical and virtual manipulatives. In doing so, we summarize key theoretical perspectives that yield different predictions about the relative effectiveness of physical and virtual manipulatives.

2.1. Learning with Multiple Representations

Most STEM instruction uses multiple representations to illustrate complex concepts (Ainsworth, 2008; National Research Council, 2006). This practice builds on evidence that multiple representations enhance learning. Extensive evidence shows that adding a visual representation to a text-based representation enhances learning compared to only visual or only text (for overviews, see Mayer, 2003, 2005; Schnotz, 2005). Similarly, adding multiple visual representations to text can further enhance student learning (Rau, Aleven, & Rummel, 2015).

The present article focuses on the case where one visual representation is added to a text-based representation. For example, students learning about atomic structure are often presented with visual representations such as energy diagrams (see Figure 1a) in addition to text-based representations (see Figure 1b). Baddeley's (1992) working memory model and Mayer's Cognitive Theory of Multimedia Learning (2009) can explain why adding a visual representation to text-based representations is beneficial. Visual and text-based representations are processed via different channels. Working memory contains a visual channel that processes visual representations and a verbal channel that processes text-based representations. When students attend to information presented externally, they load it into working memory. Information presented by visual representations is loaded into the visual channel, whereas information presented by text-based representations is loaded into the verbal channel. Because these channels can process information in parallel, providing both visual and text-based representations increases students' cognitive capacity to process information relevant to the target concepts, thereby enhancing their learning. Further, integrating information from visual and verbal channels allows students to process the target concepts more deeply (Mayer, 2009; Schnotz, 2014). Indeed, prior research shows that combining visual and text-based representations can significantly enhance students' learning (Ainsworth, 2006; Rau, 2017).

---Figure 1---

Multiple representation theory describes several mechanisms through which the integration of visual and text-based representations enhances learning. According to Ainsworth's (2006) Design, Functions, and Tasks (DeFT) framework, visual representations can complement information provided by text-based representations, for example by engaging students in complementary processes (e.g., a visual representation may engage students' perceptual processes whereas a text-based representation engages them in verbal reasoning processes). Further, a visual representation can constrain the interpretation of a text-based representation and vice versa (e.g., a visual representation may contain information about an object's color that is not mentioned in the text). Finally, Ainsworth (2006) proposes that the integration of visual and text-based representations helps students construct more abstract mental models that are organized at a higher level and thereby promote transfer, compared to mental models that are based on a single representation (e.g., text alone).

Similarly, Schnotz's (2014) Integrated Theory of Picture Comprehension (ITPC) proposes that the integration of visual and text-based representations is advantageous because it allows students to combine complementary processes. Information from visual representations is loaded into working memory via analog processes that are based on a nonverbal feature analysis of the representations, yielding an analog internal representation. By contrast, information from text-based representations is processed by mapping lexical patterns to concepts in long-term memory, yielding a propositional internal representation. To construct a mental model that integrates information from text-based as well as visual representations, students need to actively relate the analog and propositional internal representations to one another. Compared to learning with only text-based or only visual representations, active processing of complementary information can enhance students' learning.

However, extant multiple representation theory provides little guidance for choosing specific types of manipulative representations. In fact, these theories do not distinguish between static images that students see (e.g., in a textbook) and interactive representations that students manipulate. Thus, they do not specifically describe processes that result from manipulating the representations via physical or virtual interaction modes. To address this gap, we turn to research that has compared physical and virtual manipulatives.

2.2. Learning with Physical and Virtual Manipulatives

Manipulatives are a type of visual representation that allow students to construct and modify representations in order to make sense of complex, abstract concepts. Research on the educational benefits of manipulatives dates back to the late 19th century when "hands-on" activities became popular (Huxley, 1897). This early work led to suggestions that students should manipulate objects to learn new information (Montessori, 1966). Kinesthetic interactions with physical objects have been argued to be motivating (Deboer, 1991; Doias, 2013; Flick, 1993). Further, physical manipulatives afford concrete experiences that can help students connect abstract concepts to realistic contexts (Clements, 1999; Goldstone & Son, 2005).

Before interactive computer technology was widely available, manipulatives were by necessity physical objects. Yet, the advent of personal computing brought up the question of whether virtual manipulatives could replace physical ones. On the one hand, some scholars argue that virtual manipulatives can deprive students from valuable hands-on experiences (Scheckler, 2003), predicting that physical manipulatives may be generally more effective than virtual ones. On the other hand, virtual manipulatives are often cheaper and can prevent cognitive load by being more flexibly designed to exclude distracting details (Sarama & Clements, 2009). They are also more easily integrated with text-based representations, which may help students connect multiple representations (Barrett, Stull, Hsu, & Hegarty, 2015; Lee & Chen, 2015). This led to the argument that virtual manipulatives may be generally more effective than physical ones.

However, there is no empirical basis for the general superiority of physical or virtual manipulatives. Many studies found no effects of representation mode (Drickey, 2000; Klahr, Triona, & Williams, 2007; Moyer-Packenham & Westenskow, 2013; Yuan, Lee, & Wang, 2010; Zacharia & Constantinou, 2008). Others found that the effects depend on the nature of the target concept (Chini, Madsen, Gire, Rebello, & Puntambekar, 2012; Gire et al., 2010; Olympiou &

Zacharia, 2012; Zacharia & Olympiou, 2011). This concept-dependency has been the focus of conceptual salience theory.

2.2.1. Conceptual Salience

Building on working memory theory, conceptual salience theory assumes that students have to explicitly attend to information in order to load it into working memory (Mayer, 2009; Miller, 1956). Representations can be designed so that they attract students' attention (Mayer & Moreno, 1998; Schnotz, 2005). Specifically, when interacting with manipulatives, students have to pay attention to the feature they manipulate. For example, if students press a button to affect change in the manipulative, they have to pay attention to the button. A representational feature is said to be *salient* if it draws students' attention to it (e.g., through interaction design). The feature is said to make a *concept salient* if it carries information relevant to a target concept. For example, if a student learns about Cartesian graphs by translating x- and y-values from a table into a graph. The action of plotting a point in the graph requires attention to the x-values and y-values on the graph's axis, and the location of these values on the axes is relevant to understanding the Cartesian graphs. Hence, the action of plotting makes the concept salient.

Conceptual salience theory explains the concept-dependency of findings from earlier studies by proposing that the mode that makes a concept more salient yields higher learning gains for that concept (Olympiou & Zacharia, 2012). Olympiou and Zacharia (2012) conducted an *a priori* analysis of how students interacted with physical and virtual manipulatives to understand which mode would make specific target concepts more salient. A subsequent experiment confirmed that physical manipulatives were indeed more effective if they made the concept more salient than virtual ones, whereas virtual manipulatives were more effective if they made the concept more salient. Other studies also show that interaction designs that make concepts salient enhance learning of those concepts (Schneider et al., 2016; Stull, Hegarty, Dixon, & Stieff, 2012).

While multiple representation theory does not directly address representation modes, conceptual salience theory makes sense in the context of DeFT (Ainsworth, 2006) and ITPC (Schnotz, 2014). Specifically, if a visual representation provides complementary and/or constraining information that serves to construct a more abstract mental model (Ainsworth, 2006), then this benefit should be enhanced by a representation that makes the information maximally salient. Similarly, students should be more likely to include relevant information in their analogue internal representation if the visual representation makes that information salient.

In sum, conceptual salience theory extends multiple representation theory by predicting that effects of manipulatives on learning depend on whether their interaction design matches the target concept. Learning of the concept should be enhanced by manipulatives that engage students in actions that draw attention to features that carry information relevant to that concept.

2.2.2. Haptic Encoding

A related line of research focuses on the capability of haptic cues to make concepts salient. Haptic cues are features of physical manipulatives that can be experienced by touch (Magana & Balachandran, 2017; Zaman, Vanden Abeele, Markopoulos, & Marshall, 2012). If these haptic cues are relevant to a target concept, they can draw students' attention to the concept (Shaikh et al., 2017; Skulmowski, Pradel, Kühnert, Brunnett, & Rey, 2016).

Connecting haptic cues to concepts offers at least two advantages. First, by adding touch as an additional sense, students can construct more retrieval cues for the target concept compared to virtual manipulatives that engage only the visual sense (Olympiou & Zacharia, 2012; Wang & Tseng, 2016). Consequently, it is more likely that students can recall the target concept later on.

Evidence for this claim comes from studies finding that physical manipulatives increase the memorability of data visualizations (e.g., Stusak et al., 2015).

Second, haptic cues can increase cognitive capacity. This argument builds on a modification of Baddeley's (1992) working memory model, which includes an additional haptic channel (Baddeley, 2012). Because the visual, verbal, and haptic channels can process information in parallel, haptic cues can increase students' capacity to process information about the target concept, which increases learning outcomes. Evidence for this claim comes from studies suggesting that physical manipulatives increase cognitive efficiency compared to virtual manipulatives (e.g., Antle & Wang, 2013), and from studies showing that manipulating physical objects can decrease cognitive load compared to virtual objects (e.g., Skulmowski et al., 2016).

With respect to multiple representation theory, neither DeFT (Ainsworth, 2006) nor ITPC (Schnotz, 2014) address haptic processing of external representations. Nevertheless, the general argument of DeFT that the addition of complementary and/or constraining information that can contribute to mental model construction would enhance learning seems congruent with the idea that this information could be provided by the haptic channel. Further, even though ITPC does not describe what type of internal representation would result from processing of haptic information, is seems plausible that it might result from a form of haptic nonverbal feature analysis that would yield a haptic type of analogue internal representation. Extrapolating both from DeFT and ITPC, one might therefore argue that mental model building could be enhanced by haptic information.

In sum, haptic encoding theory extends multiple representation theory by predicting an effect of representation mode; specifically, that physical manipulatives are more effective than virtual manipulatives if they offer haptic cues for explicit connections between the manipulative and the target concept.

2.2.3. Embodied Action Schemas

A mostly separate line of research has focused on the potential of objects to activate embodied action schemas. This research builds on theories of embodied cognition. While there are a large variety of embodied cognition theories, their common theoretical assumption is that body movements influence cognition (Glenberg, 2010; Glenberg, Witt, & Metcalfe, 2013; Wilson, 2002). Because cognition evolved to help humans interact with the real world by mentally simulating the effects of interactions in the world, cognition is considered to be a mental simulation of body actions in the world (Glenberg, 1997). Consequently, all higher-order thinking builds on such mental simulations of body actions. For example, understanding of growth functions builds on a mental simulation of growth in the real world. Therefore, the schema of an abstract concept (e.g., growth functions) is in essence embodied. A prominent example of embodied schemas is the use of metaphors in language (Lakoff & Johnson, 1980). For example, languages often metaphorically associate upward movements with improvement or growth (e.g., "the future looks up" or "she has high socioeconomic status"). Such metaphors are grounded in real-world experiences (e.g., upward movement or being on higher ground).

Embodied schema theory conceptualizes learning as the formation of a mental simulation of a target concept, which is grounded in embodied action schemas (Abrahamson & Lindgren, 2014; Clark, 2013). Consequently, embodied schema theory predicts that instruction is more effective if it allows for bodily experiences of the concept (Johnson-Glenberg et al., 2014). For instance, studies show that bodily experiences of a concept enhance learning of the concept by activating sensorimotor brain systems associated with that experience (Kontra, Lyons, Fischer, &

Beilock, 2015). Embodied experiences can be invoked by moving the body in ways that activate an embodied schema; that is, schemas that align with a mental simulation or metaphor that matches the concept (e.g., moving upwards while learning about growth concepts) (Black et al., 2012; Johnson-Glenberg et al., 2014; Nathan et al., 2014).

Support for embodied schema theory comes from studies showing that learning of a concept is enhanced if students move their bodies in ways that are consistent with an embodied action schema that aligns with the concept (Hayes & Kraemer, 2017; Nathan & Walkington, 2017). For instance, Nathan and Walkington (2017) found that students who used their bodies to form obtuse and acute triangles showed advantages in understanding concepts related to these triangles. Further, interaction designs that require students to move their bodies in ways that align with a target concept enhances learning of the concept (Abrahamson & Lindgren, 2014; Antle et al., 2009; Bamberger & diSessa, 2003). While many embodied learning interventions ask students to make explicit connections between their sensorimotor experiences and the concepts they are learning (e.g., Abrahamson & Lindgren, 2014), embodied schema theories do *not* assume that students' awareness of the connection between their body actions and the concepts is a precondition for learning. For instance, students in Nathan and Walkington's (2017) study were unaware of the connection between their actions and the concepts. This suggests that actions that activate embodied schemas affect learning via nonverbal and potentially implicit mechanisms. Hence, an important contrast to conceptual salience and haptic encoding theories is that embodied schema theory does not assume that students need to be aware of how their embodied experiences relate to the target concepts.

Further evidence for embodied schema theory comes from research that used gestures to assess students' internal representations (Cienki, 2005; Hostetter & Boncoddo, 2017; Nathan, 2008). Such representational gestures reflect the sensorimotor experiences that form students' internal representations (Goldin-Meadow & Beilock, 2010). Thus, representational gestures can reveal the sensorimotor basis of embodied schemas (Hostetter & Boncoddo, 2017). Representational gestures can be iconic (i.e., standing for a specific entity, such as an upward gesture indicating a specific location high on a blackboard) or metaphorical (i.e., referring to abstract referents, such as an upward gesture indicating improvement) (McNeill, 1992; Stephens, 1983). Research has indeed revealed several links between representational gestures and students' thinking about concepts. First, representational gestures can help students organize their thoughts to prepare them for explaining a concept (Kita, 2000) and to retrieve knowledge about the concept (So, Shum, & Wong, 2015), especially when this knowledge is difficult to retrieve (Alibali, Yeo, Hostetter, & Kita, 2017). Second, representational gestures can offer insights into student thinking that go beyond their verbal explanations, such as the physical experiences that were the basis for a student's experiences (Cook & Tanenhaus, 2008). Hence, representational gestures can reveal which embodied schemas underlie students' thought processes.

With respect to manipulatives, embodied schema theories have been used to argue that physical manipulatives are more effective than virtual ones because they involve bodily actions (Bakker, Antle, & Van Den Hoven, 2012; Melcer et al., 2017; Pan, 2013; Skulmowski et al., 2016; Yannier, Koedinger, & Hudson, 2015). However, this argument contrasts with empirical research on embodied action schemas. Manipulatives designed to engage students in actions that are consistent with the target concept (e.g., by inducing students to move their hands in certain ways when manipulating the representation) enhance learning, whereas manipulatives that are manipulated via movements unrelated to the target concept do not enhance learning (Black et al., 2012; Segal, Tversky, & Black, 2014). Thus, not all bodily actions activate helpful embodied schemas. A body action may invoke a schema that aligns with the target concept, which can enhance learning; or it may invoke a schema that conflicts with the concept, which can impede learning. Thus, the effects of body action depend on whether the action activates an embodied schema that aligns with the target concept. Because physical and virtual manipulatives require different body actions to affect change in the representation, they may invoke different embodied action schemas, which may enhance or impede learning. An important contrast to conceptual salience and haptic encoding theory is that these effects should be present even when students are not aware of how their body actions relate to the target concepts.

With respect to multiple representation theory, neither DeFT (Ainsworth, 2006) nor ITPC (Schnotz, 2014) describe processes related to how internal representations are shaped by body movements or any process that does not require students' explicit attention. Hence, if predictions by embodied schema theory have bearing on students' learning with physical and virtual representations, this would indicate that multiple representation theory needs to integrate embodied learning processes to describe how representation mode affects learning.

In sum, embodied schema theory describes a separate set of processes from those described by multiple representation theory and predicts that effects of manipulatives on learning depend on whether their interaction design is suitable for a target concept. Learning of the concept is enhanced by manipulatives that engage students in actions that activate embodied schemas consistent with the target concept.

2.2.4. Overview of Predictions

While multiple representation theory explains how the integration of text-based and visual representations can enhance learning, it does not address how the mode of the visual representation affects this process. We reviewed three theoretical perspectives that can address this gap. However, the three theoretical perspectives make different, sometimes opposing predictions. Specifically, haptic encoding theory predicts an effect of representation mode; that is, an advantage of physical representations—provided they offer haptic cues relevant to the target concept. By contrast, conceptual salience theory and embodied schema theory predict effects of interaction designs irrespective of representation mode. For example, a manipulative that engages students in actions that draw their attention to specific features that are conceptually relevant (conceptual salience theory) or that activate embodied schemas consistent to the target concept (embodied schema theory) should be more effective—regardless of representation mode.

In addition, the predictions of conceptual salience theory and embodied schema theory often differ in *direction*. Embodied schema theory predicts effects based on interactions with a representation that implicitly invoke embodied schemas. By contrast, conceptual salience theory predicts effects based on interactions with representations that make representational features salient by drawing students' explicit attention to them. We note that although many interventions based on embodied schema theory ask students to explicitly relate embodied schemas to conceptually salient aspects of the representations, we focus on implicit embodied schema mechanisms so as to disambiguate conceptual salience and embodied schema mechanisms. When examining embodied schema theory, it becomes clear that interactions that implicitly invoke embodied schemas make the concept more intuitively accessible, therefore often reducing students' need to pay explicit attention to them (Rau, 2020).

An example from Rau (2020) illustrates these different predictions. Imagine a student who learns about constant functions by plotting points in an interactive coordinate graph. Suppose the target concept is that in a constant function, all x-values have the same y-value. If the student can plot points by moving the hand up to the y-value and then plot the points from left to right at multiple x-values, her hand moves horizontally. The horizontal movement is thought to induce an embodied metaphor related to equality (Lakoff & Johnson, 1980), which aligns with the target concept. Hence, embodied schema theory predicts an advantage for any graph (physical or virtual) where students plot points horizontally. Yet, the horizontal movement implies that the student does not have to pay explicit attention to the y-values on the y-axis. Hence, the horizontal movement makes the target concept less salient.

However, if the graph was designed so that the student has to pick up markers from the bottom of the graph to plot the points, this would induce a vertical movement that requires the student to pay attention to the y-value of each point. Hence, conceptual salience theory predicts an advantage for any graph (physical or virtual) where students plot points using a vertical movement. However, the vertical movement is thought to induce an embodied schema of increase (Lakoff & Johnson, 1980), which conflicts with the target concept.

In sum, the different theoretical perspectives make different predictions about how physical and virtual manipulatives affect student learning. While conceptual salience and haptic encoding theory seem compatible with multiple representation theory, embodied schema theory describes processes that extend multiple representation theory. Hence, to advance multiple representation theory, we need to examine to what extent each of these three theoretical perspectives describe mechanisms that affect learning with physical and virtual representations.

3. Research Questions

In addressing gaps of multiple representation theory, our working assumption is that conceptual salience, haptic encoding, and embodied schema theories all describe valid mechanisms, given that prior research provides evidence for all three theories. In line with Rau

(2020), we propose that conceptual salience, haptic encoding, and embodied schema mechanisms likely co-occur when students learn with manipulatives. Therefore, we see little use in trying to prove or disprove any one of these theories. However, an important open question is which of the theories describes the dominant mechanism when they conflict with one another. For example, if a physical manipulative offers haptic cues for a target concept, whereas one virtual manipulative makes the concept salient and another virtual manipulative activates an embodied schema that aligns with the concept, which mechanism has the largest impact on students' learning of the concept? This question is of practical importance when an instructor has to choose a manipulative to help students learn a specific target concept. Further, addressing this question will address a gap in prior research, which has investigated these mechanisms separately, hence leaving us ignorant about boundary conditions of the three theories and interactions among them. Therefore, our goal is to investigate:

Research question 1: If physical and virtual manipulatives differ in terms of whether they make a concept salient, allow for haptic encoding of the concept, and activate targeted embodied action schemas, which is more effective at enhancing students' learning of the concept?

To address this question, we created different versions of physical and virtual energy diagrams. Our own informal review of undergraduate chemistry curricula (Brown, LeMay, Bursten, Murphy, & Woodward, 2011; Hornback, 1998; Loudon, 2009; Moore & Stanitski, 2015) showed that energy diagrams are widely used in undergraduate chemistry education to illustrate concepts about atomic structure. As detailed below, the different versions allow us to contrast predictions by the three theoretical perspectives and to rule out that physical or virtual manipulatives are generally more effective. A between-subjects experiment tests these predictions for reproduction and transfer of knowledge about three chemistry concepts.

Further, we sought to understand how physical and virtual manipulatives affect students' ability to explain the concepts in their own words. In essence, all three theories suggest that increased learning gains should translate into increased ability to explain how the manipulatives show the concepts. Therefore, we investigated:

Research question 2: Do the manipulatives affect students' explanations of the target concepts?

In addition, embodied schema theory suggests that gestures give insights into the sensorimotor experiences that form the basis for students' explanations of concepts. Therefore, students who worked with manipulatives that invoked embodied schemas for a given concept should make more gestures aligned with that concept. Thus, we investigated:

Research question 3: Do the manipulatives affect students' representational gestures while they explain the target concepts?

4. Methods

4.1. Participants

Undergraduate students were recruited from a large university in the Midwest United States via email announcements in educational psychology courses as well as flyers and posters across campus and in the community. The population at this university is predominantly Caucasian (65.64%) with 2.79% African-American, 8.3% Asian, 1.2% Native American, 5.2% Hispanic, 14.28% international, and 2.63% students with unknown ethnicity. 51.92% of the enrolled students are female. 86.25% of the students enrolled at the university are full-time students.

Students received $30 for their participation. Screening questions about students' history of chemistry classes served to ensure students did not have prior knowledge about the targeted concepts. To this end, students who had taken chemistry courses at the undergraduate level were excluded. The target sample size was determined with an a priori power analysis with an anticipated effect size of $f^2 = .27$ based on prior research (Rau, 2015; Rau, Bowman, & Moore, 2017), four conditions, and correlation among the covariate and the repeated post-test measures of $\rho = .49$. Assuming a Type I probability of $\alpha = .05$, the power analysis indicated that a minimum of $N = 120$ participants were required for a statistical power of $\beta = .80$. Thus, recruitment was stopped after 120 participants were reached. However, one participant was excluded from the analyses because his pretest score revealed substantial prior knowledge about the targeted concepts. Hence, the final sample was $N = 119$. Students' average age was 22.06 years (SD = 2.859), 80 students identified as female, 33 as male, and 6 preferred not to say. Taken together, $n = 30$ participants were in the V_C condition (20 female, 9 male, 1 unknown), $n = 30$ in the V_E condition (22 female, 5 male, 1 unknown), $n = 30$ in the P_C condition (19 female, 10 male, 1 unknown) and $n = 29$ in the P_E condition (19 female, 9 male, 1 unknown) (conditions described below).

4.2. Instructional Materials

All students worked through a sequence of instructional problems that were presented on a computer. The instructional problems were designed to support iterative representation-reflection practices of creating a representation to reflect on a concept as well as using representations to depict one's reflections (see Figure 2). To this end, each problem asked students to create an energy diagram. Then, it prompted students to reflect on how the energy diagram shows the target concepts. Students responded to the prompts by completing fill-in-the-gap sentences. If students selected an incorrect answer, they received feedback that targeted common misconceptions. All students completed eight of these problems in a row. Each problem covered each of the three target concepts and students had to construct an energy diagram for each of them. As illustrated in Figure 2, the sequence of the three target concepts was random within problems but the same for all students. Across the problems, students answered the same number of questions about each concept.

---Figure 2---

In the following, we briefly describe each target concept.

4.2.1. Concept A: Electrons Randomly Fill Equal-Energy Orbitals

Many properties of an element can be explained by the structure of its atoms. In brief, atoms are composed of a nucleus that contains protons and neutrons, as well as electrons. Energy diagrams (Figure 3) show only an atom's electrons. A small arrow stands for an electron. Electrons are organized in subatomic regions called orbitals, which are shown as horizontal lines. Orbitals are categorized by their energy level, with lower-energy orbitals shown at the bottom of the energy diagram and higher-energy orbitals shown towards the top. For example, a 1s orbital has a lower energy level than a 2s orbital. Assuming an atom is in its ground state (i.e., it has not been excited, for example, by light), its electrons are more likely to be in lower-energy orbitals than in higher-energy orbitals. However, some orbitals have the same energy level (e.g., the $2p_x$, $2p_y$, and $2p_z$ orbitals in Figure 3). Hence, equal-energy orbitals have the same likelihood to be filled (e.g., an electron is as likely to be in the $2p_x$ orbital as it is to be in the $2p_y$ orbital). A common misconception among students is that they fail to recognize that electrons fill equal-energy orbitals randomly and instead assume that electrons fill equal-energy orbitals from left to right (e.g., the $2p_x$ orbital before the $2p_y$ orbital before the $2p_z$ orbital). In sum, Concept A is that electrons randomly fill equal-energy orbitals.

---Figure 3---

4.2.2. Concept B: Up and Down Spins have Equal Energy

Magnetic properties of an element can be explained by electron spins. An electron can either have an up spin or a down spin. Energy diagrams indicate electron spin with the direction of the small arrows that depict the electrons (see Figure 3). Electrons in the same orbital have opposite spin states. However, the spin states are random (assuming there is no external magnetic field). That is, if an orbital has only one electron, the electron is equally likely to have an up spin or a down spin. If an orbital has two electrons, the first electron is equally likely to have an up or down spin—but the second electron has to have the opposite spin direction than the first one. A common misconception is that the first electron in an orbital always has an up spin. In sum, Concept B is that up and down spins are equally likely because they have equal energy.

4.2.3. Concept C: Spin States are Rotational Movements

A further important concept is related to the meaning of electron spin state. Electrons rotate around their own axis. Because electrons have a negative charge, their rotation creates a small electromagnetic field. If the electron rotates counterclockwise around its axis, it creates an electromagnetic field with an upward magnetic moment. If it rotates clockwise, its electromagnetic field has a downward moment. The spin state describes the direction of the electron's magnetic field. A common misconception is that the spin state describes an electron's directional movement towards or away from the nucleus or its movement clockwise or counterclockwise around the nucleus, instead of a rotation around its own axis. In sum, Concept C is that spin states describe rotational movements of electrons around their own axis.

4.3. Experimental Design

To test the predictions by conceptual salience, haptic encoding, and embodied schema theories, we applied the theories to the three target concepts and created four manipulatives that allow us to test the theories' predictions, described in the following. Table 1 summarizes how the four manipulatives vary by representation mode and interaction design.

---Table 1---

4.3.1. Concept A

For Concept A (that electrons randomly fill equal-energy orbitals), conceptual salience theory predicts that a manipulative enhances learning if it engages students in an action that directs their attention to the height of the orbital because equal height denotes equal energy. Haptic encoding theory predicts that a physical manipulative enhances learning if it provides haptic cues for the concept by offering a physical experience of equal height corresponding to equal energy. Embodied schema theory predicts that a manipulative enhances learning if it engages students in an action that activates an embodied schema related to equality.

To test these predictions, we created two physical and two virtual energy diagrams. The two physical energy diagrams (Figure 3a,b) provide haptic experiences of height because they allow students to feel the effort associated with lifting their hand. By contrast, the virtual energy diagrams (Figure 3c,d) do not provide haptic cues for the concept.

The physical-conceptual (P_C) energy diagram (Figure 3a) was designed based on conceptual salience theory. To show electrons in an atom's orbitals, students took cards that showed the electrons from the bottom of the diagram and lifted them up to put them into the orbitals. This vertical action makes the concept salient because planning the motor action involved in the vertical movement requires that students pay attention to the height of the orbital each time they put a card into an orbital. Likewise, virtual-conceptual (V_C) energy diagram (Figure 3c) was designed to align with conceptual salience theory. Students had to click a button at the bottom of the diagram to "take" an electron before they could click into the orbitals to place the electron. This induces a vertical movement, which requires attention to the height of the orbital when adding an electron, hence making the concept salient. Note that from an embodied schema perspective, these vertical actions induce an embodied metaphor of growth or increase, which conflicts with the target concept.

Physical-embodied (P_E) energy diagram (Figure 3b) was designed based on embodied schema theory. Students had to hold the cards next to the orbitals, then take a card and move it horizontally to the orbital. This horizontal action induces an embodied metaphor of equality, which is in line with the concept of orbitals having equal energy levels. Similarly, virtual-embodied (V_E) energy diagram (Figure 3d) was designed based on embodied schema theory. To show electrons in orbitals, students clicked on the orbital to add electrons. When adding electrons to equal-energy orbitals, students moved the mouse horizontally, which induces an embodied metaphor of equality that aligns with the concept. Note that according to conceptual salience theory, the horizontal actions do not draw attention to the height of the orbital because they do not require a height change and therefore fail to make the concept salient.

4.3.2. Concept B

For Concept B (up and down spins are equally likely because they have equal energy), conceptual salience theory predicts that a manipulative enhances learning if it engages students in an action that requires explicit attention to spin states being random. Haptic encoding theory does not apply to this concept because the spin state is encoded only visually but has no haptic correspondence in energy diagrams. Embodied schema theory predicts that a manipulative enhances learning if the actions students use to show an up spin induce an embodied schema in line with the idea that up and down spins require the same amount of energy.

To test these predictions, we designed the actions used to show spin states in the four manipulatives in the following way. For P_C, which aligns with conceptual salience theory, students picked up cards that showed the electrons from a stack of cards. The electron spin that appears first is determined by the direction the card faces in the stack. We sorted the stack of cards so that all cards showed an up spin. Students had to flip the card to show a down spin. To illustrate that the up spin is just as likely as a down spin, students had to purposefully flip the arrows so that they appear random. This action requires explicit attention to the randomness of the spin state. Hence, these fixed actions make the concept salient. Likewise, V_C was programmed so that the first click that added an electron to the orbitals always created an up arrow. Students had to click again to turn the up arrow into a down arrow. Hence, students had to perform an action to show that the spin state is random, which requires explicit attention and therefore makes the concept salient. Note that from an embodied schema perspective, it takes two actions to show a down spin but one action to show an up spin. Hence, these fixed actions mean it takes the student more energy to show a down spin than up spin, which invokes an embodied schema that conflicts with the concept.

For P_E, which aligns with embodied schema theory, the stack of cards had a random order, so that any card is equally likely show an up or down spin. The randomness means that it takes the same number of actions to show an up spin as it does to show a down spin and that it takes students the same amount of energy to show both spin states. Hence, random actions induce an embodied schema that aligns with the concept. Similarly, V_E was programmed so that the first click in an orbital created an arrow with a random spin state. The second click flipped the arrow. Again, this means it takes the same number of actions and hence the same amount of energy to show an up or down spin, which induces an embodied schema that aligns with the concept. Note that from a conceptual salience perspective, the fact that the spin states are already random means that showing randomness does not require an action and therefore students do not have to pay attention to randomness. Hence, random spin states make the concept less salient.

4.3.3. Concept C

Concept C (spin states describe rotational movements of electrons around their own axis) served as a control because it is not explicitly shown by energy diagrams. Therefore, neither of the theories made predictions about which representation mode or interaction design would enhance learning of the concept. However, if one representation mode is generally more effective than the other, we would expect to see differences on this concept.

4.3.4. Assignment to Conditions

Students were randomly assigned to one of the four manipulatives just described (see Table 2). For students in the virtual conditions, the energy diagram was integrated into the instructional problems as shown in Figure 2. Students who worked with virtual energy diagrams received corrective feedback from the computer if they made a mistake in creating an energy diagram. For example, if they added an electron with an incorrect spin state, they received information about the nature of their mistake and a prompt to think about how to resolve it.

Students who worked with physical energy diagrams solved all steps except for those that involved the energy diagram on the computer. The physical energy diagram was placed next to the computer, as shown in Figure 4. The experimenter provided feedback on students' physical energy diagrams in the same way as the computer did for the virtual energy diagrams. That is, if the student made a specific mistake, the experimenter would provide information about the nature of the student's mistake and prompt the student to think about how to resolve the mistake.

Students in all conditions had to create a correct energy diagram before they could move on to the next step in the given problem.

---Figure 4---

4.4. Measures

4.4.1. Learning gains

We assessed students' learning of the target concepts with a pretest, immediate posttest, and delayed posttest. Three equivalent test forms were created (i.e., isomorphic items assessed the same concept but used different examples, such as different atoms), and the test forms were counterbalanced across the three test times. For each concept, we assessed reproduction (i.e., the ability to recall the information that was presented about the concept in the instructional problems) and transfer (i.e., the ability to apply this information to novel problems that were structurally different from the instructional problems). The test had 45 items altogether, with 15 items per concept (six items for reproduction, seven items for transfer). A random order of test items was created prior to the experiment, but the order was identical for all students. Example test items are provided in the supplementary material.

4.4.2. Quality of explanations

To assess the quality of students' explanations of the target concepts, we conducted interviews after the immediate posttest and after the delayed posttest. Students were first asked to explain in their own words: "How does an energy diagram represent an atom?" This question was intended to be open-ended to gain insight into what students remembered about the instructional material. The second question focused on Concept A: "How does an energy diagram show the energy of electrons in an atom?" The third question focused on Concept C:

"What do the up and down arrows represent in the energy diagram? The fourth question focused on Concept B: "What determines whether an arrow points up or down in the energy diagram?"

We then transcribed and coded students' answers using a three-step procedure. First, we segmented the transcripts by sentence. Second, we coded whether the segment regarded Concept A, B, or C. Third, students received one point each time they gave a unique and correct explanation of the given concept. For Concept A, they had to mention that electrons at the same height of the energy diagram have equal energy levels. For Concept B, they had to state that the spin state of the first electron in an orbital is random in the absence of the magnetic field. For Concept C, they had to say that spin state indicates an electron's rotational movement around its own axis.

4.4.3. Gestures

To code representational gestures students produced during the interview, we relied on definitions by Stephens (1983) and McNeill (1992) who describe representational gestures as those that refer to objects, spatial relationships, actions, sensations, or abstract concepts. We did not further separate iconic and metaphorical gestures because these were difficult to distinguish in the context of our study given that the energy diagram representations themselves make use of metaphors. For example, orbitals higher in the diagram have higher energy. Hence, if students make an upward gesture, it is impossible to know whether they refer to the actual location in the diagram or to a metaphor of upward motion being associated with increase.

We coded gestures for their meaning, using the student's speech to disambiguate how to interpret the gesture. For example, an up and down movement of the hands would be interpreted as indicating equality if it was accompanied with speech related to equality (e.g., "equally likely", see Figure 5), but not if it was accompanied by speech related to electron spin states (e.g., "electrons can have an up or down spin). We were particularly interested in gestures relating to *equality*, because it corresponds to the embodied action schema that aligns with students' understanding of Concepts A and B. Metaphors related to equality are in line with an internal representation of electrons equal-energy orbitals having an equal amount of energy (Concept A), and of up and down spin states being equally likely (Concept B). To receive the code *equality*, the student's speech had to relate to something being equal (e.g., "their energy level is equal" or "equally likely"), and the hands had to correspond to comparing two quantities or showing equal amounts (e.g., moving one hand from side to side or moving both hands up and down quickly) (see Figure 5). While other gestures occurred, we do not report on them here because the conceptual salience, haptic encoding, and embodied schema theories did not make predictions about them. Interrater reliability between two independent coders on 10% of the data was kappa = .785.

We computed the dependent variable as the number of gestures divided by the number of utterances across the two interviews separately for Concepts A and B. Utterances were defined based on turns of talk (i.e., the interviewer asking one of the three interview questions and the student responding). This yielded the average number of times students produced an equality gesture when describing each of the concepts.

---Figure 5---

4.4.4. Time on task

We assessed the time student spent solving the instructional problems using log data. The time spent on a problem was determined as the duration between the time when the problem had fully loaded and the time when the student clicked the 'done' button to move to the next problem in the set.

4.5. Procedure

The experiment involved two sessions in a research lab, scheduled three to six days apart. In session 1 (about 1.5 hours), students first filled out a consent form. Then, they completed the pretest and received instruction according to their experimental condition. The instructional phase lasted about 1 hour. Next, they took the immediate posttest and participated in an interview. During the interviews, the energy diagram was not present.

In session 2 (about 1 hour), students took the delayed posttest, followed by an interview about how energy diagrams show atoms. At the end of session 2, students filled out a demographics survey and received payment for their participation.

4.6. Analyses

Before addressing the research questions, we conducted several prior checks. First, to check if students showed learning gains on the target concepts, we used a repeated measures ANOVA model with test-time (i.e., pretest, immediate posttest, and delayed posttest), concept (i.e., Concepts A, B, and C), and test scale (i.e., reproduction and transfer scales) as repeated factors, and scores on the tests as dependent measures. Second, to check for pretest differences, we used a repeated measures ANOVA with mode and action as independent factors, concept and test scale as repeated factors, and pretest scores as the dependent measure. Finally, to check whether condition affected time on task, we used an ANOVA with mode and action as independent factors and time on task as dependent measure.

To test which manipulative is most effective for students' learning of the target concepts (research question 1), we used a repeated measures ANCOVA with mode and action as independent factors, and pretest scores as covariate. As repeated factors, we included test-time (i.e., immediate and delayed posttest), concept (i.e., Concepts A, B, and C), and test scale (i.e., reproduction and transfer scales). Posttest scores were the dependent measures.

To test whether the manipulatives affect students' explanations of the target concepts (research question 2), we used a repeated measures ANCOVA with mode and action as independent factors, pretest scores as the covariate, and concept as repeated factor. The quality score of students' explanations was the dependent variable.

To test whether manipulatives affect students' representational gestures while explaining the concepts (research question 3), we used the same repeated measures ANCOVA. As covariate, we included students' total number of gestures across the two interviews because students' amount of gesturing varied between participants and correlated with the dependent measures. Dependent measures were the proportion of gesture codes across the two interviews.

For all ANCOVA models, we used Bonferroni-corrected effect slices, which compare the levels of one factor for all levels of the interacting factor to understand the nature of statistically significant interaction effects and to test the predicted effects. If the factor had more than two levels, we additionally computed Bonferroni-corrected pairwise comparisons. All *p*-values for post-hoc tests reported below were adjusted by the Bonferroni correction.

5. Results

In the following analyses, we report partial η^2 and *d* for effect sizes. According to Cohen (1988), an effect size partial η^2 of .01 or d of .2 corresponds to a small effect, partial η^2 of .06 or d of .5 to a medium effect, and partial η^2 of .14 or d of .8 to a large effect. Table 2 shows the means and standard deviations for all conditions and measures.

---Table 2---

5.1. Prior Checks

First, we checked for pretest-to-posttest learning gains on each concept and scale. We found a statistically significant main effect of test-time, $F(1, 236) = 172.288$, $p < .001$, p. $\eta^2 = .594$. As summarized in Table 3, effect slices and pairwise comparisons showed statistically significant learning gains for all concepts and scales.

---Table 3---

Second, we checked for differences between conditions on the pretest (Fs < 1). As detailed in Table 4, effect slices and pairwise comparisons showed no statistically significant effects on any of the pretest measures.

---Table 4---

Finally, we tested whether condition affected time on task. Mode had a statistically significant effect on instructional time, such that students working with physical representations took significantly longer than students working with virtual representations, $F(1, 115) = 14.447$, $p < .001$, p. $\eta^2 = .112$. There was no effect of action, $F(1, 115) = 1.848$, $p = .177$, nor a statistically significant interaction between mode and action, $F(1, 115) = 1.637$, $p = .203$. To ensure our conclusions are independent of time on task, we conducted all following analyses with and without time on task as a covariate.

5.2. Effects on Learning Gains

For the analyses corresponding to the research questions, Table 5 shows a comparison of the results to the predictions of each theory, Table 6 lists the statistics for all analyses with and without time on task as a covariate, and Figure 6 provides a summary of the results.

---Table 5---

---Table 6---

---Figure 6---

First, we investigated which manipulative is most effective for students' learning of the target concepts (research question 1). Results showed no statistically significant main effects of action or mode, but significant between-within interactions between mode and concept, between action and scale, between action, mode, test-time, concept, and scale. The interaction between mode, concept, and scale was significant only when time on task was included as a covariate.

To gain insights into the nature of these interaction effects, we used Bonferroni-corrected effect slices to compare the effects of action and mode for each concept and scale.

5.2.1. Reproduction

Recall that for *Concept A*, conceptual salience theory predicted an advantage of vertical over horizontal actions (i.e., P_C / V_C > P_E / V_E), haptic encoding theory predicted an advantage of physical over virtual manipulatives (i.e., P_C / P_E > V_C / V_E), and embodied schema theory predicted an advantage of horizontal over vertical actions (i.e., P_E / V_E > P_C / V_C).

For *reproduction of knowledge about Concept A*, we found a statistically significant advantage of horizontal (i.e., P_E / V_E) over vertical actions (i.e., P_C / V_C). This effect is in line with embodied schema theory. The difference between physical and virtual manipulatives was not statistically significant.

Recall that for *Concept B*, conceptual salience theory predicted an advantage of vertical over horizontal actions (i.e., P_C / V_C > P_E / V_E), and embodied schema theory predicted an advantage of horizontal over vertical actions (i.e., P_E / V_E > P_C / V_C). Haptic encoding theory did not make predictions for this concept.

For *reproduction of knowledge about Concept B*, we found a statistically significant advantage of horizontal (i.e., P_E / V_E) over vertical actions (i.e., P_C / V_C). This effect is in line with embodied schema theory. The difference between physical and virtual manipulatives was not statistically significant.

Recall that *Concept C* served as a control because neither theory predicted an effect. We found no difference between horizontal and vertical actions, and no difference between physical and virtual manipulatives. These null results are in line with the tested theories.

5.2.2. Transfer

For *transfer of knowledge about Concept A*, we found a statistically significant advantage of physical (i.e., P_C / P_E) over virtual manipulatives (i.e., V_C / V_E). This effect is in line with haptic encoding theory. The difference between horizontal and vertical actions was not statistically significant.

For *transfer of knowledge about Concept B*, we found a statistically significant advantage of vertical actions (i.e., P_C / V_C) over horizontal actions (i.e., P_E / V_E). This effect is in line with conceptual salience theory. The difference between physical and virtual manipulatives was not statistically significant.

For *transfer of knowledge about Concept C*, we found no difference between horizontal and vertical actions, and no difference between physical and virtual manipulatives. These null results are in line with the tested theories.

5.3. Effects on Quality of Explanations

Next, we investigated whether the manipulatives affect students' explanations of the target concepts (research question 2). Recall that conceptual salience predicted that students who worked with manipulatives that make the target concept salient (i.e., P_C / V_C) provide higher-quality explanations than students who worked with manipulatives that did not make the concept salient (i.e., P_E / V_E). This prediction applied to Concepts A and B. Similarly, haptic encoding predicted that students who worked with physical manipulatives (i.e., P_C / P_E) provide higher-quality explanations than students who worked with virtual manipulatives (i.e., V_C / V_E). This prediction applied to Concept A only. In contrast, embodied schema theory predicted that students who worked with manipulatives that activate target embodied schemas (i.e., P_E / V_E) provide higher-quality explanations than students who worked with manipulatives that activated conflicting embodied schemas (i.e., P_C / V_C). This prediction applied to Concepts A and B.

Results showed no main effects of mode or action. An interaction between mode and concept was statistically significant only when time on task was included as a covariate.

For *Concept A*, Bonferroni-corrected effect slices showed that students gave higher-quality explanations if they had worked with physical manipulatives (i.e., P_C / P_E) than if they had worked with virtual manipulatives (i.e., V_C / V_E), which was statistically significant only when time on task was not included as a covariate. This effect is in line with haptic encoding theory.

No other effect slices were statistically significant.

5.4. Effects on Gestures

Finally, we tested whether the manipulatives affect students' representational gestures while they explain the target concepts (research question 3). Recall that embodied schema theory predicted that students who worked with manipulatives that invoked targeted embodied schemas (i.e., P_E / V_E) would make more equality gestures than students who worked with manipulatives that did not invoke the targeted embodied schemas (i.e., P_C / V_C). This prediction applied to Concepts A and B.

Results showed a statistically significant main effect of action, as well as a significant interaction between action and concept.

For *Concept A*, Bonferroni-corrected effect slices showed no differences between horizontal and vertical actions. For *Concept B*, the effect slices showed that students made more equality gestures if they had worked with manipulatives that invoked the targeted embodied schema through random actions (i.e., P_E / V_E) than if they had worked with manipulatives that involved fixed actions (i.e., P_C / V_C). This effect is in line with embodied schema theory.

No other effect slices were statistically significant.

6. Discussion

Our study addresses a gap in multiple representation theory, which does not specify how representation modes affect student learning. To address this gap, we turned to conceptual salience, haptic encoding, and embodied schema theories, which describe how representation mode affects learning with visual representations. While conceptual salience theory and haptic encoding theory appear to be compatible with multiple representation theory, embodied schema theory describes mechanisms that extend multiple representation theory. Our review of prior research shows that even though many studies have compared physical and virtual manipulatives, most of them have focused on conceptual salience mechanisms. Yet, haptic encoding and embodied schema theories yield hypotheses about how manipulatives affect learning that differ and often conflict with conceptual salience theory. The goal of this article was hence to address this issue by contrasting hypotheses based on conceptual salience theory, haptic encoding theory, and embodied schema theory. The underlying rationale was that in light of prior evidence for each of these theories, they likely describe mechanisms that co-occur while students interact with manipulatives. However, prior research leaves us ignorant as to the nature of such co-occurrence. Therefore, our goal was not to pit these theories against one another, but rather, to investigate which mechanism prevails when students use manipulatives to learn abstract concepts. Addressing this question can yield novel theoretical insights into boundary conditions and interactions among the theories. Further, it yields new hypotheses that—pending further investigation—could help teachers and instructional designers choose manipulatives that best enhance learning of specific concepts. To our knowledge, this experiment is the first to integrate conceptual, haptic, and embodied theories by systematically comparing effects of representation modes and actions used to manipulate the representations.

6.1. Summary of Findings

Our results indicate that there is no one theory that prevails. Instead, different theories apply to different outcome measures, which suggests boundary conditions for the different theories. On *learning gains*, we found disparate results for reproduction and transfer of knowledge. For *reproduction of knowledge*, our findings were in line with embodied schema theory. Specifically, we found that manipulatives enhanced learning if they were designed to engage students in actions that induce embodied metaphors that align with the target concept, irrespective of representation mode. For *transfer of knowledge*, our findings were in line with haptic encoding theory for the concept where the manipulative offered a haptic experience of the concept. Specifically, in this case, the physical manipulatives yielded higher learning gains than the virtual manipulatives. For the concept where haptic cues were not available, our findings were in line with conceptual salience theory. Specifically, in this case, we found that manipulatives enhanced learning if they were designed to draw students' attention to conceptually relevant aspects of the manipulative, irrespective of representation mode. Finally, none of the theories had predicted effects for a control concept, and our results did not show effects of representation mode or action.

These results on learning gains were consistent across analyses with and without time on task as a covariate. This indicates that these effects were not the result of students working with physical manipulatives spending more time with the instructional materials.

The interviews showed very limited support for the tested theories. With respect to the quality of students' *explanations of the concepts*, we only found a statistically significant advantage of physical over virtual manipulatives for one of the concepts, in line with haptic encoding theory. However, the effect was statistically significant only when time on task was not included as a covariate, suggesting that the advantage of physical manipulatives may be due to students spending more time with the instructional materials. We found no statistically significant effects that aligned with the predictions by conceptual salience and embodied schema theory. Again, there were also no statistically significant effects for the control concept, which is in line with the tested theories.

With respect to *gestures*, we tested predictions offered by embodied schema theory, which were partially supported by our results. Specifically, for one of the target concepts, we found that students made more gestures in line with the concept if they worked with manipulatives that were designed to induce embodied metaphors, irrespective of representation mode. This result was robust across analyses with or without time on task as covariate, indicating that the effect was not influenced by students working with physical manipulatives spending more time with the instructional materials.

6.2. Interpretation of Findings

6.2.1. Comparison of mechanisms

Our results on reproduction of knowledge suggest that actions that invoked embodied schemas related to the target concept enhance students' ability to recall basic information about the concept. Hence, the embodied schema mechanism seemed to outweigh the other mechanisms when the goal was to recall basic information about a concept. By contrast, our results on transfer of knowledge suggest that physical manipulatives enhanced students' ability to apply what they learned to novel tasks if they offer haptic cues for the target concept (as in the case of Concept A). Hence, if a physical manipulative offered haptic cues for a concept, it seemed that the haptic encoding mechanism outweighed the other mechanisms. If haptic cues were not available (as in the case of Concept B), actions that made the target concept salient seemed to enhance students' ability to apply what they learned to novel situations. Hence, if the goal was to apply knowledge to novel situations and no haptic cues were available, the conceptual salience mechanism seemed to prevail.

---Figure 7---

6.2.2. Complexity of mechanism and learning goals

We interpret these findings in relation to the complexity of the underlying mechanism and the learning goal, as illustrated in Figure 7. Let us first consider the complexity of the learning goal. Because reproduction of knowledge does not involve elaboration or alteration of the learned association between the learning material (e.g., the manipulative) and the concept but merely requires recall of the association, reproduction tasks rely on simple knowledge structures that represent one-to-one mappings between stimulus and response (Koedinger, Corbett, & Perfetti, 2012).

By contrast, transfer of knowledge requires elaboration and goes beyond simple one-to-one mappings between the learning material and the concept. That is, transfer tasks have no one clear solution but allow for several answer options (e.g., different approaches to explaining one's reasoning) and therefore access complex knowledge structures (Koedinger et al., 2012).

Elaboration itself is an explicit process because it requires that students consciously reason about the novel connections they seek to establish. Hence, it seems reasonable to assume that an awareness of the connections between manipulatives and target concepts is a prerequisite for the ability to transfer knowledge to novel situations. If students are not aware of the connection between the manipulative and the concept, becoming aware of the connection would itself constitute transfer of implicit knowledge to the explicit realm – an additional cognitive step compared to students who are already aware of the connection.

Let us now consider the complexity of the learning mechanisms. As detailed above (see Nathan & Walkington, 2017), the embodied schema mechanism does not assume that students are aware of the relation between their movements and the target concepts, which is why the embodied schema mechanism has been considered to be implicit. Unless students are explicitly asked to make connections between the concepts and their body movements (which was not the case in our intervention), the embodied schema mechanism therefore involves simple associations of a sensorimotor experience with the target concept. By contrast, both the haptic encoding and the conceptual salience mechanism rely on explicit learning processes in the sense that they assume that students pay attention to connections between the manipulative and the target concepts (Manches et al., 2010; Olympiou & Zacharia, 2012). That is, both mechanisms involve complex reasoning about how representational cues relate to the concept. Arguably, the haptic encoding mechanism is more complex than the conceptual salience mechanism because conceptual salience relies only on visual cues and their connection to the target concept, whereas haptic encoding involves cues of multiple modalities. Thus, in the case of haptic encoding, the addition of the haptic mode to the visual mode allows for a larger number of cues that the student can connect to the concept, hence yielding a richer set of elaborations about the connections.

Further, the addition of the haptic mode may increase the cognitive capacity of the student to establish complex connections because the student can make use of an additional processing channel (Baddeley, 2012).

6.2.3. Alignment between mechanisms and learning goals

Finally, let us consider the alignment between the mechanisms and learning goals suggested by our results. Our results suggest that simple mechanisms enhanced attainment of the simple learning goals we examined in our study, whereas complex mechanisms enhanced attainment of the complex learning goals in our study. Specifically, learning simple associations of body movements with the target concepts in our study may have helped students acquire knowledge traces that increased the likelihood that they could reproduce the association, possibly because the cue to recall the association triggered a sensorimotor simulation of the concept. By contrast, learning complex connections between visual and/or haptic cues with the target concepts in our study enhanced students' ability to apply knowledge about the concept to novel situations in our transfer tasks. This may explain why we found an advantage of energy diagram manipulatives that were designed to activate embodied schemas for our reproduction of knowledge tasks, but an advantage of the physical manipulatives that offered haptic cues or manipulatives that were designed to make concepts salient for our transfer of knowledge tasks. Further, because the haptic encoding mechanism is more complex than the conceptual salience mechanism, it might have yielded more complex connections between the energy diagram manipulatives and the target concepts in our study—when haptic cues were available. This may explain why we found an advantage of the physical energy diagram manipulatives over manipulatives that made the target concept salient for Concept A.

It is important to note that our results do not imply that embodied learning experiences are always best suited for simple learning goals. Many realistic embodied interventions ask students to explicitly connect their sensorimotor experiences to the target concepts, for example through explanations. Such interventions therefore use the embodied experience as grounding for complex conceptual reasoning. Similarly, conceptual salience can be combined with designs aimed at enhancing recall. Therefore, we do not mean to imply that learning experiences designed to make concepts salient are only effective for complex learning goals. Yet, in the context of energy diagram manipulatives that were designed to enhance embodied or conceptual mechanisms in isolation from one another, we found that embodied schema mechanisms aligned with simple learning goals and conceptual salience mechanisms aligned with complex learning goals.

The alignment between the complexity of mechanisms and the complexity of the learning goals we examined in our study might also explain the weak results on the quality of students' explanations and gestures. In terms of complexity, we consider our interview measures to be between reproduction and transfer as assessed by the posttests: The ability to recall the concept is a prerequisite to explaining it, and explaining it is more complex than reproducing it on a multiple-choice test because there are many correct ways to explain it. However, our explanation tasks did not require new connections to novel situations as the transfer test did. Hence, this measure may not have differentiated between simple and complex processes. Consequently, on measures of moderate complexity, the explicit mechanisms (i.e., haptic encoding and conceptual salience) and the implicit mechanism (i.e., embodied schemas) may have cancelled each other out. Nevertheless, the interview measures offered some evidence that the different mechanisms co-occurred because haptic encoding seemed to some degree enhance students' ability to explain target concepts in their own words and because gestures seemed to some extent reflect the sensorimotor experiences involved in learning the concepts.

Across the measures, the null effects on control Concept C indicated that we found no evidence for alternative predictions regarding the general superiority of physical or virtual manipulatives. This rules out, for example, that physical manipulatives were generally more effective because they may have been more motivating or that virtual manipulatives were generally more effective because they integrated multiple representations on one screen. Consequently, we can rule out that general mode effects might have distorted the effects we found on Concepts A and B. This is in line with prior research, which shows that representation modes per se have negligible effects but that their effects depend on the learning goals. We extend this prior research by showing which learning mechanisms can explain which manipulatives are suitable for which learning goals.

6.2.4. Towards an integrative theory of learning with manipulatives

Altogether, our results provide a first step towards a theory of learning with manipulatives that integrates conceptual, haptic, and embodied mechanisms. Because our experiment varied learning outcome measures within subjects—that is, we found differential effects depending on learning outcome measures for the same students—our results suggest that these mechanisms indeed co-occurred while students learned with manipulatives. Consequently, none of the tested theories proved right or wrong. Instead, as illustrated in Figure 7, they had different scopes of application. We propose that the effectiveness of a manipulative was maximized if it affected a learning mechanism that matched the given learning goal in terms of complexity. Conversely, the manipulative’s effectiveness diminished as the difference in complexity between the affected mechanism and the learning goal increased. Hence, if a manipulative exclusively affects simple, associative mechanisms (i.e., activation of embodied schemas via sensorimotor actions without involving explicit connections to concepts), it may be most effective for attaining goals that involve learning of simple knowledge structures that are assessed by reproduction tasks. By contrast, if a manipulative exclusively affects complex, reflective mechanisms (i.e., connections between manipulatives and visual and/or haptic cues), it may be most effective for goals that involve learning of complex knowledge structures. Multi-modal connections to concepts are more complex than visual connections to concepts. Therefore, when considered in isolation from other mechanisms, haptic encoding mechanisms may be better suited for the attainment of complex learning goals than conceptual salience mechanisms. Thus, the match between mechanism and learning goals may constitute a boundary condition for embodied schema, conceptual salience, and haptic encoding theories. Finally, for learning goals of moderate complexity, effects of manipulatives that affect both simple and complex mechanisms at the same time may cancel out. Thus, interactions between co-occurring mechanisms may explain null effects in comparisons of manipulatives. In sum, our research yields a tentative and preliminary integrative model that can explain null results for learning outcome measures of moderate complexity, as well as conflicting findings by prior research about the relative effectiveness of different manipulatives on simple and complex learning outcome measures.

Albeit being tentative, our preliminary integrative model highlights how multiple representation theory might be expanded to capture effects of representation modes on learning. Our findings show that embodied schema mechanisms do affect students' learning, thereby demonstrating that a mechanism that is not captured by extant multiple representation theory affects learning. Hence, our findings suggest that embodied schema theory is a useful lens to examine why specific combinations of representations enhance student learning, especially when considering effects of representation modes. Further, our results demonstrate how conceptual and haptic processes, which seem compatible with multiple representation theory, impact students' learning. Finally, our findings suggest that multiple representation theory might have to address how the complexity of a learning goal affects which mechanisms are at play when students integrate information from text-based and visual representations.

At a practical level, our preliminary theoretical model yields hypotheses that can be tested in broader educational contexts. While our findings have to be interpreted with caution as they were obtained with a specific set of representations and tasks, future research should consider using different manipulatives to enhance reproduction versus transfer. It is possible that manipulatives that activate embodied schemas congruent with the target concept are particularly suitable to enhance reproduction of knowledge. By contrast, physical manipulatives that offer haptic cues for the target concepts may be particularly suitable to enhance transfer. If haptic cues are unavailable (e.g., because the concept has no haptic correspondence or because physical manipulatives are impractical in the given context), then manipulatives that engage students in actions that make the target concept salient may be a suitable alternative to enhance transfer. We consider our study to be a first step towards offering guidance for instructional practice by offering these insights as tentative hypotheses that need to be examined by future research.

6.3. Limitations and Future Directions

First and foremost, we consider our integrative theoretical model to be tentative and preliminary because it is based on a single experiment that focused on a small number of concepts related to atomic structure, on specific chemistry manipulatives used for specific tasks, and a particular population of undergraduate students who had had no formal exposure to the target concepts. Consequently, the sample was not representative of the chemistry undergraduate population. Especially in the light of results that suggest that different theories apply to different learning goals, it is imperative that future research broadens the set of concepts, manipulatives, and tasks. Hence, future research should test if our findings generalize to concepts, manipulatives, and populations beyond the focus of our study.

Second, future research could solidify our preliminary theoretical model by collecting data about the proposed learning processes. Conceptual salience, haptic encoding, and embodied schema theories make assumptions about the mechanisms that lead to effects of manipulatives on learning outcomes. Because the goal of our experiment was to test predictions, we accepted the theories' assumptions about underlying mechanisms as results from prior research. Future research should assess learning mechanisms to verify that these assumptions are indeed true in the context of learning with manipulatives. Further, assessing the learning mechanisms would provide further evidence that simple mechanisms affect learning of simple knowledge structures, whereas complex mechanisms affect learning of complex knowledge structures.

Third, while we purposefully selected concepts for which the tested theories made conflicting predictions, we did not test all possible conflicts between these theories. For example, we did not include a case where conceptual salience theory and embodied schema theory align but conflict with haptic encoding theory.

Fourth, with a duration of about 1 hour, the instructional phase in our study was relatively short. Over longer learning periods, it is possible that sequence effects emerge, such that one mechanism prevails at first and another mechanism later. In particular, the disparate findings on reproduction and transfer might play out as sequence effects in a longer intervention. Instruction typically starts with simple concepts and moves towards complex concepts. If manipulatives that invoke embodied schemas that align with the target concepts are effective for simple knowledge structures, these manipulatives might be particularly effective early in a longer learning intervention. By contrast, if manipulatives that offer haptic cues or make concepts salient are effective for complex knowledge structures, these manipulatives might be particularly effective later in a learning intervention. On the other hand, an argument could be made for the opposite sequence. Students may first need to acquire deep understanding of a concept by engaging in complex learning processes, before they should focus on recall, because unless students have a solid foundational understanding of the concept, they may learn to recall incorrect or incomplete information. Future research should test such sequencing effects, so as to provide new insights into embodied grounding of concepts (Nathan, 2008) and concrete-to-abstract sequences that have not yet accounted for effects of embodied schema mechanisms (Goldstone & Son, 2005).

Fifth, while we consider a delay between the immediate and delayed posttest of three to six days appropriate for an intervention of only 1 hour, it may not be representative of delays between intervention and testing in many realistic educational settings. In future research, experiments done as part of real classroom interventions could test whether our findings generalize to longer delays.

Sixth, we note that most realistic interventions are more complex than the ones we considered in our study. Specifically, manipulatives are usually designed to invoke multiple mechanisms at once, or they are often embedded in activities that invoke additional learning mechanisms. The purpose of our study was to examine conceptual, haptic, and embodied mechanisms in isolation to gain insights into how each of them affects learning. Future research can build on our findings by testing manipulative designs that engage students in different combinations of multiple learning mechanisms.

Finally, our research illustrates that drawing on theories that expand multiple representation theory can help us choose the most effective type of representation for specific concepts. Specifically, our work suggests that multiple representation theory would benefit from incorporating embodied mechanisms of learning. Expanding multiple representation theory in this way may reveal additional questions about how multiple representations enhance learning. We focused on a specific representational feature—namely, representation modes—and identified theories that made predictions about how physical engagement with learning materials may have different effects than virtual engagement. Yet, there are many other types of representations to consider; for example, domain-specific vs. domain-general representations. When examining other characteristics of representations, future research may benefit from looking beyond multiple representation theory to identify hypotheses about productive combinations. Incorporating new mechanisms into multiple representation theory would not only help address open questions about which specific constellations of representations are effective but may also give rise to new questions about how multiple representations affect learning via these mechanisms.

7. Conclusion

Our experiment juxtaposed multiple mechanisms that might explain effects of physical and virtual representation modes on students' conceptual learning but which have been considered in mostly separate literatures. Our findings suggests a tentative integrative model that maps embodied, conceptual, and haptic mechanisms to learning goals based on their complexity. Specifically, we found that energy diagram manipulatives that invoked embodied schemas through students' body movements enhanced learning of simple knowledge about atomic structure, assessed by reproduction tests. For example, engaging students in horizontal movements helped them recall knowledge about equality. By contrast, physical manipulatives that offered haptic cues enhanced learning of complex knowledge about atomic structure, assessed by transfer tests. For example, a physical experience of lifting objects to a certain height helped students transfer knowledge related to height. When haptic cues were not available, manipulatives that made the target concept salient enhanced transfer. For example, even when a physical experience was not available, engaging students in actions that required attention to moving the mouse to a certain height enhanced their ability to transfer knowledge related to height. At a practical level, our integrative model offers tentative hypotheses for how instructional designers and instructors might choose manipulatives based on instructional goals. Our research offers directions for future research to examine whether different manipulatives are generally suitable to enhance reproduction and transfer of knowledge. Taken together, our findings expand multiple representation theory by suggesting that incorporating conceptual salience, haptic encoding, and embodied schema mechanisms may yield insights into which combinations of multiple representations are most effective.

8. References

Abrahamson, D., & Lindgren, R. (2014). Embodiment and Embodied Design. In R. K. Sawyer (Ed.), *The Cambridge handbook of the Learning Sciences* (2 ed., pp. 358-376). New York, NY: Cambridge University Press.

Abrahamson, D., & Sánchez-García, R. (2016). Learning is moving in new ways: The ecological dynamics of mathematics education. *Journal of the Learning Sciences, 25*(2), 203-239. doi:10.1080/10508406.2016.1143370

Ainsworth, S. (2008). The educational value of multiple-representations when learning complex scientific concepts. In J. K. Gilbert, M. Reiner, & A. Nakama (Eds.), *Visualization: Theory and practice in science education* (pp. 191-208). Netherlands: Springer.

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences, 21*(2), 247-286.

Alibali, M. W., Yeo, A., Hostetter, A. B., & Kita, S. (2017). Representational gestures help speakers package information for speaking. In B. Church & S. A. Kelly, M. W. (Eds.), *Why Gesture? How the hands function in speaking, thinking and communicating* (pp. 15-37). Amsterdam / Philadelphia: Joyhn Benjamins Publishing Company.

Antle, A. N., Corness, G., & Droumeva, M. (2009). What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers, 21*(1), 66-75. doi:10.1016/j.intcom.2008.10.005

Antle, A. N., & Wang, S. (2013). Comparing motor-cognitive strategies for spatial problem solving with tangible and multi-touch interfaces. In S. Jorda & N. Pares (Eds.), *Proceedings of the 7th International Conference on tangible, embedded and embodied interaction* (pp. 65-72). New York, NY: ACM.

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29. doi:http://dx.doi.org/10.1146/annurev-psych-120710-100422

Bakker, S., Antle, A. N., & Van Den Hoven, E. (2012). Embodied metaphors in tangible interaction design. *Personal and Ubiquitous Computing, 16*(4), 433-449. doi:10.1007/s00779-011-0410-4

Bamberger, J., & diSessa, A. (2003). Music as embodied mathematics: A study of a mutually informing affinity. *International Journal of Computers for Mathematical Learning, 8*(2), 123-160.

Barrett, T. J., Stull, A. T., Hsu, T. M., & Hegarty, M. (2015). Constrained interactivity for relating multiple representations in science: When virtual is better than real. *Computers and Education, 81*, 69-81. doi:10.1016/j.compedu.2014.09.009

Bieda, K. N., & Nathan, M. J. (2009). Representational disfluency in algebra: evidence from student gestures and speech. *ZDM, 41*(5), 637-650.

Black, J. B., Segal, A., Vitale, J., & Fadjo, C. L. (2012). Embodied cognition and learning environment design. In D. H. Jonassen & S. M. Land (Eds.), *Theoretical foundations of learning environments* (pp. 198-223). New York: Routledge Taylor & Francis Group.

Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, MA: Belknap.

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380–400.

Chini, J. J., Madsen, A., Gire, E., Rebello, N. S., & Puntambekar, S. (2012). Exploration of factors that affect the comparative effectiveness of physical and virtual manipulatives in an undergraduate laboratory. *Physical Review Special Topics-Physics Education Research, 8*(1), 010113. doi:DOI: http://dx.doi.org/10.1103/PhysRevSTPER.8.010113

Cienki, A. (2005). Image schemas and gestures. In B. Hampe (Ed.), *From Perception to Meaning: Image Schemas in Cognitive Linguistics* (pp. 421-441). Berlin, Germany: Mouton de Gruyter.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204. doi:http://dx.doi.org/10.1017/S0140525X12000477

Clausner, T. C. (2005). Image schemas: From linguistic analysis to neural grounding. In B. Hampe (Ed.), *From Perception to Meaning: Image Schemas in Cognitive Linguistics* (pp. 93-110). Berlin, Germany: Mouton de Gruyter.

Clements, D. H. (1999). 'Concrete'Manipulatives, Concrete Ideas. *Contemporary Issues in Early Childhood, 1*(1), 45-60. doi:http://dx.doi.org/10.2304/ciec.2000.1.1.7

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, S. W., & Tanenhaus, M. K. (2008). Speakers Communicate Their Perceptual-Motor Experience to Listeners Nonverbally *Proceedings of the Cognitive Science Society* (Vol. 30, pp. 957-962).

Cuendet, S., Bumbacher, E., & Dillenbourg, P. (2012). Tangible vs. virtual representations: when tangibles benefit the training of spatial skills *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (pp. 99-108): ACM.

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science, 340*(6130), 305-308. doi:10.1126/science.1230579

Deboer, G. (1991). *A history of ideas in science education*. New York: Teachers College Press.

Dienes, Z. P. (1961). *The Dienes M.A.B. multibase arithmetic blocks*. London, UK: National Foundation for Educational Research.

Doias, E. D. (2013). *The effect of manipulatives on achievement scores in the middle school mathematics class.* (Doctor of Education), Lindenwood University, Saint Charles, Missouri.

Donald, M. (1991). *Origins of the modern mind*. Cambridge, MA: Harvard University Press.

Drickey, N. A. (2000). *A comparison of virtual and physical manipulatives in teaching visualization and spatial reasoning to middle school mathematics students.* Utah State University, Logan, Utah.

Fan, J. E. (2015). Drawing to learn: How producing graphical representations enhances scientific thinking. *Translational Issues in Psychological Science, 1*(2), 170–181. doi:http://dx.doi.org/10.1037/tps0000037

Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S., . . . LeMaster, R. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical Review Special Topics—Physics Education Research, 1*, 1-8.

Flick, L. B. (1993). The meanings of hands-on science. *Journal of Science Teacher Education, 4*(1), 1-8.

Gibson, J. J. (1997). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67-82). Hillsdale, NJ: Erlbaum.

Gilbert, J. K. (2005). Visualization: A metacognitive skill in science and science education. In J. K. Gilbert (Ed.), *Visualization: Theory and practice in science education* (pp. 9-27). Dordrecht, Netherlands: Springer.

Gire, E., Carmichael, A., Chini, J. J., Rouinfar, A., Rebello, S., Smith, G., & Puntambekar, S. (2010). The effects of physical and virtual manipulatives on students' conceptual learning about pulleys. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *9th International Conference of the Learning Sciences* (Vol. 1, pp. 937-943): International Society of the Learning Sciences.

Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and Brain Sciences, 20*(1), 41-50.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*, 586–596. doi:10.1002/wcs.55

Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment 25 years of cognitive psychology. *Perspectives on Psychological Science, 8*(5), 573-585.

Goldin-Meadow, S., & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Perspectives on Psychological Science, 5*(6), 664-674. doi:10.1177/1745691610388764

Goldstone, R., Schyns, P. G., & Medin, D. L. (1997). Learning to bridge between perception and cognition. *Psychology of learning and motivation, 36*, 1-14. doi:10.1016/S0079-7421(08)60279-0

Goldstone, R., & Son, J. (2005). The Transfer of Scientific Principles Using Concrete and Idealized Simulations. *Journal of the Learning Sciences, 14*(1), 69-110. doi:10.1207/s15327809jls1401_4

Han, I. (2013). Embodiment: a new perspective for evaluating physicality in learning. *Journal of Educational Computing Research, 49*(1), 41-59.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica D: Nonlinear Phenomena, 42*(1), 335-346. doi:10.1016/0167-2789(90)90087-6

Hayes, J. C., & Kraemer, D. J. (2017). Grounded understanding of abstract concepts: The case of STEM learning. *Cognitive research: principles and implications, 2*(1). doi:10.1186/s41235-016-0046-z

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15*(3), 495-514.

Hostetter, A. B., & Boncoddo, R. (2017). Gestures highlight perceptual-motor representations in thinking. In B. Church & S. A. Kelly, M. W. (Eds.), *Why Gesture? How the hands function in speaking, thinking and communicating* (pp. 155-174). Amsterdam / Philadelphia: Joyhn Benjamins Publishing Company.

Huxley, T. H. (1897). Scientific Education: Notes of an After-Dinner Speech. In T. H. Huxley (Ed.), *Collected Essays: Science and education* (Vol. 3, pp. 111-133). New York: Appleton.

Jaakkola, T., Nurmi, S., & Lehtinen, E. (2010). Conceptual change in learning electricity: Using virtual and concrete external representations simultaneously. In L. Verschaffel, E. De Corte, T. de Jong, & J. Elen (Eds.), *Use of Representations in Reasoning and Problem Solving: Analysis and Improvement* (pp. 133-152). New York, NY: Routledge.

Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology, 106*(1), 86-104. doi:10.1037/a0034008

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162-185). Cambridge, UK: Cambridge University Press.

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. doi:10.1002/tea.20152

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. doi:10.1111/j.1551-6709.2012.01245.x

Kontra, C., Lyons, D. J., Fischer, S. M., & Beilock, S. L. (2015). Physical experience enhances science learning. *Psychological Science, 26*(6), 737-749.

Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *The Journal of the Learning Sciences, 9*(2), 105-143. doi:10.1207/s15327809jls0902_1

Kozma, R., & Russell, J. (2005). Multimedia learning of chemistry. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 409-428). New York, NY: Cambridge University Press.

Lakoff, G. J., & Johnson, M. (1980). *Metaphors we live by*. Chicago-London: University of Chicago Press.

Latour, B. (1986). Visualization and cognition: Thinking with eyes and hands. *Knowledge and society: Studies in the Sociology of Culture Past and Present, 6*, 1-40.

Lee, C. Y., & Chen, M. J. (2015). Effects of worked examples using manipulatives on fifth graders' learning performance and attitude toward mathematics. *Journal of Educational Technology & Society, 18*(1), 264-275.

Lira, M. E., & Stieff, M. (2018). Using Gesture Analysis to Assess Students' Developing Representational Competence. In K. L. Daniel (Ed.), *Towards a Framework for Representational Competence in Science Education* (pp. 205-228). Dordrecht, Netherlands: Springer.

Magana, A. J., & Balachandran, S. (2017). Students' development of representational competence through the sense of touch. *Journal of Science Education and Technology, 26*(3), 332-346.

Manches, A., O'Malley, C., & Benford, S. (2010). The role of physical representations in solving number problems: A comparison of young children's use of physical and virtual materials. *Computers & Education, 54*(3), 622-640.

Mayer, R. E. (2009). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 31-48). New York, NY: Cambridge University Press.

Mayer, R. E., & Moreno, R. (1998). *A cognitive theory of multimedia learning: Implications for design principles*. Paper presented at the CHI-98 Workshop on Hyped-Media to Hyper-Media.

McElhaney, K. W., Chang, H. Y., Chiu, J. L., & Linn, M. C. (2015). Evidence for effective uses of dynamic visualisations in science curriculum materials. *Studies in Science Education, 51*(1), 49-85. doi:10.1080/03057267.2014.984506

McNeill, D. (1992). Guide to Gesture Classification, Transcription, and Distribution. In D. McNeill (Ed.), *Hand and mind. What gestures reveal about thought* (pp. 75-104). Chicago, IL: University of Chicago Press.

Melcer, E. F., Hollis, V., & Isbister, K. (2017). Tangibles vs. Mouse in Educational Programming Games: Influences on Enjoyment and Self-Beliefs. In G. Mark & S. Fussel (Eds.), *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1901-1908). New York, NY: ACM.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Montessori, M. (1966). *Secret of childhood*. New York, NY: Ballantine Books.

Moyer-Packenham, P. S., & Westenskow, A. (2013). Effects of virtual manipulatives on student achievement and mathematics learning. *International Journal of Virtual and Personal Learning Environments, 4*(3), 35-50.

Nathan, M. J. (2008). An embodied cognition perspective on symbols, grounding, and instructional gesture *Symbols and embodiment: Debates on meaning and cognition* (pp. 375-396).

Nathan, M. J., & Walkington, C. (2017). Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *Cognitive research: principles and implications, 2*(9), 1-20. doi:10.1186/s41235-016-0040-5

Nathan, M. J., Walkington, C., Boncoddo, R., Pier, E. L., Williams, C. C., & Alibali, M. W. (2014). Actions speak louder with words: The roles of action and pedagogical language for grounding mathematical reasoning. *Learning and Instruction, 33*, 182-193. doi:10.1016/j.learninstruc.2014.07.001

National Research Council. (2006). *Learning to Think Spatially*. Washington, D.C.: National Academies Press.

Olympiou, G., & Zacharia, Z. C. (2012). Blending physical and virtual manipulatives: An effort to improve students' conceptual understanding through science laboratory experimentation. *Science Education, 96*(1), 21-47. doi:10.1002/sce.20463

Pan, E. A. (2013). *The use of physical and virtual manipulatives in an undergraduatemechanical engineering (dynamics) course.* (Doctor of Philosophy), University of Virginia, Charlottesville, Virginia.

Pyatt, K., & Sims, R. (2012). Virtual and physical experimentation in inquiry-based science labs: Attitudes, performance and access. *Journal of Science Education and Technology, 2*(1), 133-147.

Rau, M. A. (2020). Comparing Multiple Theories about Learning with Physical and Virtual Representations: Conflicting or Complementary Effects? *Educational Psychology Review, 32*(2), 297-325. doi:10.1007/s10648-020-09517-1

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review, 29*(4), 717–761. doi:10.1007/s10648-016-9365-3

Rau, M. A. (2015). Enhancing undergraduate chemistry learning by helping students make connections among multiple graphical representations. *Chemistry Education Research and Practice, 16*, 654-669. doi:10.1039/C5RP00065C

Rau, M. A., & Matthews, P. G. (2017). How to make 'more' better? Principles for effective use of multiple representations to enhance student learning. ZDM - Mathematics Education, 49(4), 491-496. doi:10.1007/s11858-017-0846-8

Rau, M. A., Bowman, H. E., & Moore, J. W. (2017). Intelligent technology-support for collaborative connection-making among multiple visual representations in chemistry. *Computers & Education, 109*(C), 38-55. doi:http://dx.doi.org/10.1016/j.compedu.2017.02.006

Sarama, J., & Clements, D. H. (2009). “Concrete” computer manipulatives in mathematics education. *Child Development Perspectives, 3*(3), 145-150.

Scheckler, R. K. (2003). Virtual labs: A substitute for traditional labs? . *International Journal of Developmental Biology, 47*, 231–236.

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016). Using mobile eye-trackers to unpack the perceptual benefits of a tangible user interface for collaborative learning. *ACM Transactions on Computer-Human Interaction (TOCHI), 23*(6), 39.

Schnotz, W. (2014). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2 ed., pp. 72-103). New York, NY: Cambridge University Press.

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49-69). New York, NY: Cambridge University Press.

Segal, A., Tversky, B., & Black, J. (2014). Conceptually congruent actions can promote thought. *Journal of Applied Research in Memory and Cognition, 3*(3), 124-130.

Shaikh, U. A., Magana, A. J., Neri, L., Escobar-Castillejos, D., Noguez, J., & Benes, B. (2017). Undergraduate students' conceptual interpretation and perceptions of haptic-enabled learning experiences. *International Journal of Educational Technology in Higher Education, 14*(1), 1-21.

Skulmowski, A., Pradel, S., Kühnert, T., Brunnett, G., & Rey, G. D. (2016). Embodied learning using a tangible user interface: The effects of haptic perception and selective pointing on a spatial learning task. *Computers & Education, 92*, 64-75. doi:http://dx.doi.org/10.1016/j.compedu.2015.10.011

So, W. C., Shum, P. L. C., & Wong, M. K. Y. (2015). Gesture is more effective than spatial language in encoding spatial information. *The Quarterly Journal of Experimental Psychology, 68*(12), 2384-2401. doi:10.1080/17470218.2015.1015431

Stephens, D. (1983). *Hemispheric language dominance and gesture hand preference.* University of Chicago.

Stull, A. T., Gainer, M., Padalkar, S., & Hegarty, M. (2016). Promoting representational competence with molecular models in organic chemistry. *Journal of Chemical Education, 93*, 994−1001. doi:10.1021/acs.jchemed.6b00194

Stull, A. T., & Hegarty, M. (2016). Model manipulation and learning: Fostering representational competence with virtual and concrete models. *Journal of Educational Psychology, 108*(4), 509-527. doi:http://dx.doi.org/10.1037/edu0000077

Stull, A. T., Hegarty, M., Dixon, B., & Stieff, M. (2012). Representational translation with concrete models in organic chemistry. *Cognition and Instruction, 30*(4), 404-434. doi:10.1080/07370008.2012.719956

Stusak, S., Schwarz, J., & Butz, A. (2015). Evaluating the memorability of physical visualizations. In B. Begole, J. Kim, K. Inkpen, & W. Woo (Eds.), *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3247-3250). New York, NY: ACM.

van der Meij, J., & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction, 16*(3), 199-212. doi:10.1016/j.learninstruc.2006.03.007

Wang, T. L., & Tseng, Y. K. (2016). The Comparative Effectiveness of Physical, Virtual, and Virtual-Physical Manipulatives on Third-Grade Students' Science Achievement and Conceptual Understanding of Evaporation and Condensation. *International Journal of Science and Mathematics Education*, 1-17.

Wang, T. L., & Tseng, Y. K. (2018). The comparative effectiveness of physical, virtual, and virtual-physical manipulatives on third-grade students' science achievement and conceptual understanding of evaporation and condensation. *International Journal of Science and Mathematics Education, 16*(2), 203-219.

Wertsch, J. V., & Kazak, S. (2011). Saying more than you know in instructional settings. In T. Koschmann (Ed.), *Theories of learning and studies of instructional practice* (pp. 153-166). New York: Springer.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625-636.

Wolfe, P. (2001). *Brain matters: Translating research into classroom practice*. Alexandria, VA: Association for Supervision and Curriculum Development.

Yannier, N., Koedinger, K. R., & Hudson, S. E. (2015). Learning from Mixed-Reality Games: Is Shaking a Tablet as Effective as Physical Observation? *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1045-1054): ACM.

Yuan, Y., Lee, C. Y., & Wang, C. H. (2010). A comparison study of polyominoes explorations in a physical and virtual manipulative environment. *Journal of Computer Assisted Learning, 26*(4), 307-316.

Zacharia, Z. C., & Constantinou, C. P. (2008). Comparing the influence of physical and virtual manipulatives in the context of the Physics by Inquiry curriculum: The case of undergraduate students' conceptual understanding of heat and temperature. *American Journal of physics, 76*(4), 425-430. doi:http://dx.doi.org/10.1119/1.2885059

Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction, 21*(3), 317-331.

Zaman, B., Vanden Abeele, V., Markopoulos, P., & Marshall, P. (2012). Editorial: the evolving field of tangible interaction for children: the challenge of empirical validation. *Personal and Ubiquitous Computing, 16*(4), 367-378.

Supplementary Material

TEST

What is the difference between the $2p_x$ and $2p_y$ orbitals?

- ○ The $2p_x$ orbital is filled before the $2p_y$ orbital.
- ○ The orbitals differ in their spatial orientation.
- ○ There are no differences between these orbitals.

TEST

The first electron to fill the $2p_y$ orbital always ----.

- ○ has a 100% chance of spinning up
- ○ has a 50% chance of spinning up
- ○ has a 100% chance of spinning down

TEST

Which statement correctly describes what the direction of an arrow in energy diagrams means?

- ○ If an arrow points up, the electron is attracted to the external magnetic field.
- ○ If an arrow points down, the electron has a tendency to move towards the nucleus.
- ○ If an arrow points up, the electron's magnetic moment points up.

Figure A1. Sample items from the reproduction test

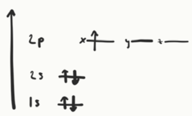
TEST

A classmate says that all atoms in groups 3A, 4A, 6A, and 7A have more than one correct electron configuration.

Do you agree?

Explain your reasoning.

The drawing shows an energy diagram for boron, which is in group 3A.



What second energy diagram would you draw to illustrate your point?

Submit

TEST

Consider ground-state nitrogen, which has an electron configuration of $1s^2\ 2s^2\ 2p_x^1\ 2p_y^1\ 2p_z^1$.

Why is the electron in the $2p_z$ orbital equally likely to have an up-spin or a down-spin, assuming there's no external magnetic field?

Now, assume the nitrogen atom exists in an external magnetic field. How does the external magnetic field affect the spin state of the electron in the $2p_z$ orbital?

Submit

TEST

One of your classmates says that energy diagrams are unfortunately designed because they suggest that electrons move towards or away from the nucleus, which is inaccurate.

Do you agree with your classmate?

Explain your reasoning.

How could the energy diagram be redesigned to satisfy your classmate?

Submit

Figure A2. Sample items from the transfer test